

Facettenbasierte Indexierung multipler Artefakte

Ein Framework für vage Anfragen in der
Produktentwicklung

von Nadine Weber



UNIVERSITY OF
BAMBERG
PRESS

Schriften aus der Fakultät Wirtschaftsinformatik
und Angewandte Informatik der Otto-Friedrich-
Universität Bamberg

Schriften aus der Fakultät Wirtschaftsinformatik und Angewandte Informatik

Band 11



University of Bamberg Press 2011

Facettenbasierte Indexierung multipler Artefakte

Ein Framework für vage Anfragen in der
Produktentwicklung

von Nadine Weber



University of Bamberg Press 2011

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliographie; detaillierte bibliographische
Informationen sind im Internet über <http://dnb.ddb.de/> abrufbar

Diese Arbeit hat der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-Friedrich-Universität als Dissertation vorgelegen

1. Gutachter: Prof. Dr. Andreas Henrich

2. Gutachter: Prof. Dr. Otto K. Ferstl

Tag der mündlichen Prüfung: 11. November 2011

Dieses Werk ist als freie Onlineversion über den Hochschulschriften-Server (OPUS; <http://www.opus-bayern.de/uni-bamberg/>) der Universitätsbibliothek Bamberg erreichbar. Kopien und Ausdrücke dürfen nur zum privaten und sonstigen eigenen Gebrauch angefertigt werden.

Herstellung und Druck: docupoint, Barleben

Umschlaggestaltung: Dezernat Kommunikation und Alumni der Otto-Friedrich-Universität Bamberg

© University of Bamberg Press Bamberg 2011

<http://www.uni-bamberg.de/ubp/>

ISSN: 1867-7401

ISBN: 978-3-86309-058-6 (Druckausgabe)

eISBN: 978-3-86309-059-3 (Online-Ausgabe)

URN: urn:nbn:de:bvb:473-opus-3945

Danksagung

*Those who are not confused today
have not understood the problem.
(Jack Welsh)*

Diese Promotion präsentiert die Forschungsergebnisse meiner Tätigkeit im Forschungsverein FORFLOW, die ich als wissenschaftliche Mitarbeiterin am Lehrstuhl für Medieninformatik der Otto-Friedrich-Universität Bamberg ausübte.

Aus diesem Grund möchte ich mich ganz besonders bei meinem Doktorvater Herrn *Professor Dr. Andreas Henrich* bedanken, der mir nicht nur die Möglichkeit gab, dieses Promotionsvorhaben umzusetzen. Vielmehr möchte ich mich dafür bedanken, dass er immer ein offenes Ohr für meine Fragen hatte und selbst dann Verständnis zeigte, wenn es in schwierigen Zeiten nicht so vorwärts ging, wie geplant.

Des Weiteren möchte ich mich bei allen FORFLOW-Kollegen bedanken, ohne die es diese Arbeit nicht gegeben hätte. Besonders erwähnen möchte ich zum einen die Industriepartner meines Teilprojektes Herrn *Ralf Mendgen*, Herrn *Christof Schaff* und Herrn *Christoph Gesterkamp* von der Robert Bosch GmbH, Herrn *Walter Glück*, Herrn *Norbert Friedrich* und Herrn *Stefan Rudolph* von der Leoni Bordnetz-Systeme GmbH & Co. KG sowie Herrn *Dr. Erich Meier* von der method park Software AG, die mir interessante Einblicke in den Bereich der Produktentwicklung gaben.

Zum anderen möchte ich meinem ehemaligen Lehrstuhlkollegen und FORFLOW-Mitstreiter *Raiko Eckstein* für die tolle Zusammenarbeit und die intensiven Diskussionen zur Entwicklung des LFRP-Frameworks danken. Unsere gemeinsame Konferenzteilnahme an der Stanford University mit anschließender Kalifornien-Tour und einem Besuch bei den Bären :) wird mir immer in Erinnerung bleiben.

Auch meinen anderen ehemaligen Kollegen des Lehrstuhls — insbesondere *Stefanie Sieber* — möchte ich meinen Dank für die gute Zusammenarbeit, die auflockernden Kickerspiele nach dem Mittagessen und die anregenden Gespräche zur Fertigstellung dieser Arbeit aussprechen.

Weiterhin danke ich Herrn *Professor Dr. Guido Wirtz* für die Übernahme des Vorsitzes meiner Promotionskommission sowie Herrn *Professor Dr. Otto K. Ferstl* für die Übernahme des Zweitgutachtens.

Ein herzlicher Dank gilt zudem meinen Korrekturlesern *Florian Müller*, *Karlheinz*

Morgenroth, Jasmin Weber und *Anja Sauer*. Ohne sie und ihre kritischen, aber wertvollen Fragen und Anmerkungen hätte ich wohl an dem einen oder anderen Punkt aufgegeben.

Schließlich möchte ich auf diesem Weg noch ganz besonders meinen Eltern, *Konrad und Bärbel Weber*, sowie meiner gesamten Familie dafür danken, dass sie immer für mich da sind und mich in einer persönlich sehr schwierigen Zeit ermutigt haben, diese Arbeit fertigzustellen!

Kurzfassung

Die Entwicklung technischer Systeme — im Weiteren auch als Produkte bezeichnet — steht durch den zunehmenden Einsatz von Informations- und Kommunikationstechnologien sowie den schnellen Technologiefortschritt vor neuen Herausforderungen. In immer kürzeren Zeitabständen müssen Unternehmen innovative Produkte auf den Markt bringen, deren Variantenvielfalt und Komplexität aufgrund der immer individuelleren Kundenanforderungen stetig zunimmt.

Dabei bestehen Produkte heutzutage nicht mehr nur aus rein mechanischen Komponenten; vielmehr handelt es sich um mechatronische Systeme, an deren Entwicklung neben der Mechanik diverse Disziplinen wie die Elektrotechnik, die Softwareentwicklung oder die Akustik beteiligt sind. Daraus resultiert ein enormer Administrations- und Kommunikationsaufwand, welcher durch die Forderung nach paralleler Bearbeitung von Unternehmensabläufen im Sinne eines Simultaneous Engineering sowie nach einer mit der Globalisierung einhergehenden verteilten Produktentwicklung zusätzlich erhöht wird. Gleichzeitig bleibt aber die Forderung nach hochwertigen und vor allem kostengünstigen Produkten bestehen, wodurch der Druck zur Kostensenkung insbesondere auch auf die Entwicklungsabteilungen verstärkt wird.

Die Aufgabe der Produktentwicklung selbst ist dabei als ein Problemlösungsprozess zu betrachten, in dem die Lösungsfindung nur mittels intensiver Informationsverarbeitung möglich ist. Täglich werden unterschiedlichste Arten von Informationen erstellt, benötigt und verarbeitet, die primär in digitaler Form vorliegen. Diese werden in heterogenen Systemen, wie z. B. in CAD-, PDM-Systemen oder Dateiverzeichnissen, erzeugt und verwaltet, was eine Wiederverwendung bereits existierender Informationen erschwert. Folglich nimmt die Tätigkeit der Informationssuche immer noch einen erheblichen Anteil der Entwicklungszeit in Anspruch. Dies ist unter anderem darauf zurückzuführen, dass bereits vorhandene Suchfunktionalitäten meist nur bestimmte Informationsbedürfnisse fokussieren und übergreifende Suchmechanismen im Sinne von Enterprise Search-Lösungen die Charakteristiken produktentwicklungsspezifischer Informationen nicht berücksichtigen. Da jedoch aufgrund verkürzter Produktlebenszyklen immer weniger Zeit zur Erwirtschaftung der Entwicklungskosten bleibt, birgt gerade das Auffinden und Wiederverwenden bereits existierender Informationen ein enormes Potential zur Reduzierung nicht nur der Entwicklungszeit, sondern vor allem auch der Entwicklungskosten.

Die vorliegende Arbeit beschäftigt sich aus diesem Grund mit der Verbesserung der Informationsversorgung im Bereich der technischen Produktentwicklung. Dazu wird

ein interaktives Information Retrieval-System — das LFRP-Framework — vorgestellt, welches die vier Basiskonzepte der multiplen Ebenen, der facettierten Suche, des Rankings und der parallelen Koordinaten kombiniert, um die aus der Analyse der Informationsbedürfnisse von Produktentwicklern resultierenden Suchsituationen zu befriedigen. Da letztere eine Vielzahl an Artefakten wie beispielsweise Produkte, Dokumente oder Werkstoffe adressieren, für die jeweils unterschiedliche charakteristische Suchkriterien benötigt werden, sind diese durch eine hohe Komplexität geprägt. Darüber hinaus konkretisiert sich der Informationsbedarf meist erst im Prozessverlauf, wodurch ein exploratives Suchvorgehen unter Verwendung von facettenbasierten Suchkriterien erforderlich ist. Dadurch können Produktentwickler einerseits durch Stöbern im Datenbestand auf hilfreiche Informationen stoßen und andererseits zielgerichtet nach einer benötigten Information suchen.

Neben einer geeigneten Suchoberfläche erfordert die Realisierung des LFRP-Frameworks vor allem die Entwicklung einer Indexierungskomponente, welche die im Produktentwicklungsbereich bereits existierenden Informationen in eine für das Information Retrieval-System rechnerverarbeitbare Form, eine sogenannte rechnerinterne Repräsentation, transformiert. Da dieser als Indexierung bezeichnete Prozess prinzipiell die Grundvoraussetzung für das Funktionieren eines Suchsystems darstellt, wird dieser Teilaspekt des LFRP-Frameworks in den Fokus der vorliegenden Arbeit gestellt. Dementsprechend wird ein Framework vorgestellt, das eine Indexierung bereits vorhandener Informationen in Form facettenbasierter Suchkriterien ermöglicht. Dabei werden jedoch nicht nur Artefaktinformationen aus den in der Produktentwicklung eingesetzten heterogenen Systemen, sondern insbesondere entwicklungsspezifische Dokumente, wie technische Zeichnungen oder Stücklisten, als Informationsquellen berücksichtigt. Neben erforderlichen Modulen zum Datenimport und zur Konsistenzprüfung enthält das Indexierungsframework somit ein Modul zur Repräsentationserstellung, das maßgeblich Methoden zur Extraktion von Informationen aus diesen entwicklungsspezifischen Dokumenten bereitstellt.

Insgesamt betrachtet wird dabei die zu beherrschende Gesamtkomplexität in drei Teilkomplexitäten zerlegt, welche sich erstens auf die Informationsbedürfnisse bzw. Suchobjekte, zweitens auf die Informationsquellen und drittens auf die Informationsextraktion und ihre Methoden beziehen und nur durch ein mindestens ebenso komplexes Indexierungsframework beherrscht werden können. Folglich wird seine Realisierung abschließend mittels einer kritischen, sowohl prozess- als auch ergebnisorientierten Reflexion bewertet.

Inhaltsverzeichnis

1. Motivation	1
1.1. Problemstellung	3
1.1.1. Herausforderungen in der technischen Produktentwicklung . .	4
1.1.2. Analyse des Informationsbedarfs von Produktentwicklern . . .	7
1.2. Einordnung der Arbeit	15
1.2.1. Informations- und Wissensmanagement	16
1.2.2. Information Retrieval	18
1.2.3. Rolle des Information Retrieval im Informations- und Wissens- management	20
1.3. Forschungsfragen und Zielsetzung der Arbeit	21
1.4. Aufbau der Arbeit	22
2. State of the Art	25
2.1. Bewährte Systematiken als Suchhilfen	25
2.1.1. Ordnungsschemata	26
2.1.2. Klassifikationssysteme	28
2.2. Industrielle Anwendungssysteme und ihre Suchfunktionalitäten	33
2.2.1. Datenbanksysteme	33
2.2.2. Dokumentenmanagementsysteme	37
2.2.3. Enterprise Resource Planning-Systeme	39
2.2.4. Produktdatenmanagementsysteme	41
2.3. Kommerzielle CAD-IR-Systeme	42
2.4. Enterprise Search-Systeme	44
2.5. Konzepte und Ansätze aus der Forschung	47
2.5.1. Spezialisierte Konzepte mit niedrigem Unterstützungsgrad . .	48
2.5.2. Generelle Konzepte mit hohem Unterstützungsgrad	52
2.6. Unterstützte Suchparadigmen	54
3. Interaktives Retrievalmodell für komplexe Suchsituationen	57
3.1. Anforderungen an ein Information Retrieval-System für die Produkt- entwicklung	57
3.2. Basiskonzepte zur Erfüllung der Anforderungen	60
3.2.1. Facettierte Suche	60
3.2.2. Parallele Koordinaten	64

3.2.3.	Konzept der multiplen Ebenen	65
3.2.4.	Konzept des Ranking	66
3.3.	Integration der Basiskonzepte in ein übergreifendes Gesamtsystem — das LFRP-Framework	68
3.3.1.	Architektur des Gesamtsystems	68
3.3.2.	Retrievalprozess des LFRP-Frameworks	70
3.3.3.	Indexierungsprozess des LFRP-Frameworks	75
4.	Datenimport — Entwicklungsspezifische Artefakte und ihre Besonderheiten	79
4.1.	Betrachtung ausgewählter entwicklungsspezifischer Artefakte	79
4.1.1.	Artefakttyp Produkt	81
4.1.2.	Artefakttyp Dokument	87
4.1.3.	Artefakttyp Werkstoff	93
4.1.4.	Artefakttyp Person	95
4.1.5.	Artefakttyp Projekt	97
4.2.	Beziehungen zwischen Artefakten	99
4.3.	Quellen für Artefaktinformationen	100
4.4.	Importieren von Artefaktinformationen	103
5.	Repräsentationserstellung — Informationsextraktion aus XML-Datei und Dokumenten	109
5.1.	Schema zur Definition der suchbaren Facetten und ihrer Abhängigkeiten	109
5.2.	Szenario 1: Import einer XML-Datei	118
5.3.	Szenario 2: Import von XML-Datei und Dokumentverzeichnis	123
5.4.	Szenario 3: Import eines Dokumentverzeichnisses	124
5.4.1.	Lasten- und Pflichtenheft	128
5.4.1.1.	Informationsgehalt eines Lastenheftes	129
5.4.1.2.	Informationsgehalt eines Pflichtenheftes	131
5.4.1.3.	Extraktorkomponenten für Lasten- bzw. Pflichtenhefte	134
5.4.2.	CAD-Modell	141
5.4.2.1.	3D-CAD-Systeme: Historie und Charakteristik	142
5.4.2.2.	Austauschformate für 3D-Modelle	147
5.4.2.3.	Informationsgehalt von CAD-Modellen	151
5.4.2.4.	Extraktion und Repräsentation von CAD-Daten	157
5.4.2.5.	Evaluation implementierter Ansätze für einen 3D / 3D-Vergleich	171
5.4.3.	Berechnungsergebnis und Testbericht	180
5.4.3.1.	Berechnungsergebnis: Informationsgehalt und Repräsentationsmöglichkeiten	181
5.4.3.2.	Testbericht: Informationsgehalt und Repräsentationsmöglichkeiten	185

5.4.4.	Technische Zeichnung	188
5.4.4.1.	Erstellung und Archivierung technischer Zeichnungen	189
5.4.4.2.	DXF als Austauschformat	191
5.4.4.3.	Informationsgehalt technischer Zeichnungen	192
5.4.4.4.	Extraktion und Repräsentation von Zeichnungsdaten	194
5.4.4.5.	Konzepte für einen 2D / 2D-Vergleich und Evaluations- ergebnisse implementierter Ansätze	200
5.4.5.	Stückliste	206
5.4.5.1.	Informationsgehalt von Stücklisten	207
5.4.5.2.	Extraktion und Repräsentation von Stücklisteninfor- mationen	209
5.4.6.	Zusammenfassung	210
6.	Konsistenzprüfung und Indexverwaltung	213
6.1.	Konsistenzprüfung zur Validierung extrahierter Artefaktbeschreibungen	214
6.2.	Indexverwaltung und Konsistenzprüfung zur Verifikation von Artefakt- beschreibungen	216
6.2.1.	Aktualisierungsszenario 1: Identische Informationsquellen . . .	221
6.2.2.	Aktualisierungsszenario 2: Unterschiedliche Informationsquellen	222
6.2.3.	Zusammenfassung	224
7.	Kritische Reflexion	225
7.1.	Prozessorientierte Reflexion: Leitfaden für eine unternehmensspezifi- sche Realisierung	225
7.2.	Ergebnisorientierte Reflexion: Nutzen und kritische Erfolgsfaktoren .	230
8.	Zusammenfassung und Ausblick	235
A.	Vergleich von Vorgehensmodellen	241
B.	Attribute der Artefakttypen	243
B.1.	Produkt	243
B.2.	Dokument	249
B.3.	Werkstoff	251
B.4.	Person	252
B.5.	Projekt	254
C.	Exemplarische XML-Datei für den Datenimport	255
D.	XML-Schema für das Facettenschema	259
E.	Ergebnisse der Dokumenttypanalyse	267
F.	IGES 5.3 – Spezifikationsauszug	279

G. Recall-Precision-Diagramme der FV-basierten Konzepte	281
Abbildungsverzeichnis	283
Tabellenverzeichnis	289
Verzeichnis der Code-Listings	291
Literaturverzeichnis	293
Abkürzungsverzeichnis	327
Stichwortverzeichnis	333

1. Motivation

In den vergangenen 20 Jahren bzw. seit der Entwicklung der Grundlagen für das World Wide Web (WWW) durch den britischen Informatiker Tim Berners-Lee im März 1989 am Schweizer Kernforschungsinstitut CERN¹ [Foc09] hat sich unser aller Leben bedeutend verändert. Durch den zunehmenden Einsatz von Informations- und Kommunikationstechnologien (IKT) in allen Lebensbereichen — insbesondere geprägt durch den großen Erfolg des Internets mit seinen verschiedenen Diensten (WWW, E-Mail usw.) — ist eine Informationsgesellschaft entstanden, für die der Umgang mit digitalen Informationen alltäglich geworden ist. Es gibt kaum Haushalte oder Unternehmen, die nicht mit dem Internet verbunden sind. *Surfen, googeln, bloggen, skypen* oder *twittern* sind nur einige Beispiele für Tätigkeiten, die heutzutage den Lebens- und Arbeitsalltag eines jeden Einzelnen prägen. Zusätzlich fördert die Entwicklung immer besserer und günstigerer Technologien einen Transformationsprozess hin zu einer Digitalisierung der unterschiedlichsten Anwendungen und Bereiche, wie beispielsweise in der Fotografie, der Telefonie oder beim Fernsehen. Des Weiteren findet in Unternehmen eine zunehmende Automatisierung der Geschäftsprozesse statt, wozu diverse Anwendungssysteme (AwS) entwickelt und eingesetzt werden. Unzählige neue Informationen in Form von Anwendungsdaten oder Dokumenten sowie Webseiten, Bildern, Videos und vielem mehr werden somit täglich erstellt, vervielfältigt und archiviert.

Die Gesamtheit dieser digitalen Informationen, die in IDC²-Studien bezeichnenderweise als *Digitales Universum* beschrieben wird [GRC⁺07, GCM⁺08, GR09, GR10], hat im Laufe der Zeit enorme Ausmaße angenommen. Trotz globaler Wirtschaftskrise wurden allein im Jahr 2009 nahezu 800 Exabytes³ an neuen Informationen erstellt [GR10]. Verglichen mit Studien aus den Jahren 2008 (487 Exabytes), 2007 (281 Exabytes) und 2006 (161 Exabytes) [GRC⁺07, GCM⁺08, GR09] kann weiterhin von einer Informationsexplosion ausgegangen werden, bei der sich etwa alle 18 Monate die Menge der erzeugten Daten verdoppelt. Wie Abbildung 1.1 zeigt, ist zwar der Großteil dieses Universums dem Bereich der Unterhaltungsindustrie zuzuschreiben, doch auch andere Branchen tragen einen wesentlichen Anteil dazu bei. So wird laut der IDC-Studie vom Jahr 2008 der Bereich Fertigungsindustrie und Versorgungswirtschaft als zweitgrößter Produzent digitaler Informationen gesehen [GCM⁺08]. Dies lässt sich

¹Conseil Européen pour la Recherche Nucléaire (<http://www.cern.de>).

²Ein international tätiges Marktforschungs- und Beratungsunternehmen (<http://www.idc.com>).

³Ein Exabyte entspricht etwas mehr als einer Million Terabytes.

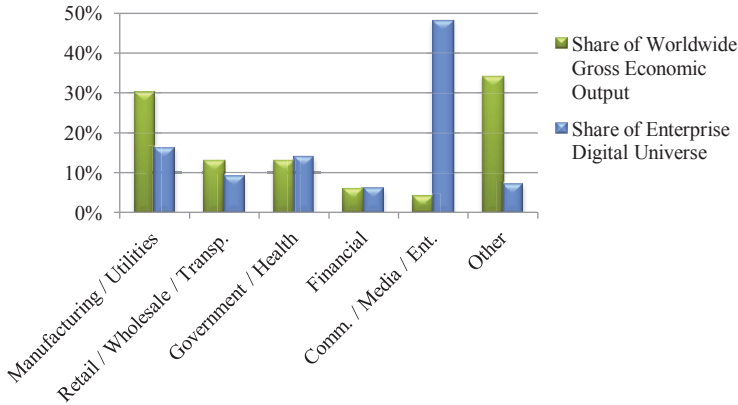


Abbildung 1.1. — Produzenten des *Digitalen Universums* gegliedert nach Industrien (nach [GCM⁺08, S. 7])

u. a. auch darauf zurückführen, dass gerade bei der Entwicklung technischer Systeme⁴ zunehmend rechnerunterstützte Methoden im Sinne des Computer Integrated Manufacturing (CIM)-Ansatzes sowie für Simulations- und Visualisierungskonzepte verwendet werden. Allein die Umstellung von zweidimensionalen (2D) auf dreidimensionale (3D) Computer Aided Design (CAD)-Systeme in den letzten 10–15 Jahren hat wesentlich zur Vergrößerung der digitalen Informationsmenge im Bereich der technischen Produktentwicklung (PE) beigetragen.

Gleichzeitig ist zu berücksichtigen, dass Wissen neben Arbeit, Kapital und Boden einen vierten Produktionsfaktor darstellt [Dru93, Prz08b], dessen Nutzung wesentlich zum Erfolg eines Unternehmens beiträgt. Dabei spielt insbesondere die Wiederverwendung bereits existierender Informationen eine bedeutende Rolle, da hierdurch Redundanzen vermieden, Mehrfacharbeiten und somit Kosten reduziert und Wettbewerbsvorteile realisiert werden können. Allerdings gestaltet sich gerade im Unternehmensumfeld die Suche nach nützlichen und hilfreichen Informationen schwierig. Trotz der Entwicklung von Websuchmaschinen und Enterprise Search-Systemen (ESS), welche die Möglichkeiten der Informationssuche erheblich revolutionierten, konnte sich insbesondere im Bereich der technischen PE bislang keiner der existierenden Ansätze wirklich durchsetzen. Deshalb wird noch immer ein zu großer Anteil der Produktent-

⁴Ein technisches System ist gemäß Pahl et al. ein Gebilde, das über Eingangs- und Ausgangsgrößen mit seiner Umwelt verbunden ist und je nach Komplexitätsgrad bspw. als Anlage, Apparat, Maschine, Gerät, Baugruppe, Maschinenelement oder Einzelteil bezeichnet wird [PBFG07]. Da es sich hierbei um einen Oberbegriff für komplexe technische Produkte handelt, wird in der vorliegenden Arbeit der Begriff *Produkt* als Synonym hierzu verwendet.

wicklungszeit mit der Suche nach Informationen verbracht [BEE⁺97, Con05]. Folglich werden innovative Konzepte benötigt, die ein einfaches und zielgerichtetes Auffinden der jeweils erforderlichen Information ermöglichen und somit zur Realisierung eines effektiven und effizienten Informations- und Wissensmanagements (IWM) im Unternehmen beitragen.

Die vorliegende Arbeit leistet hierzu ihren Beitrag, indem sie sich mit Aspekten der Entwicklung eines Information Retrieval-Systems (IRS), welches im Rahmen des Forschungsverbundes FORFLOW⁵ am Lehrstuhl für Medieninformatik der Otto-Friedrich-Universität Bamberg entwickelt wurde, beschäftigt. Dabei handelt es sich um ein interaktives Retrievalmodell für komplexe Suchsituationen, das speziell die Informationsbedürfnisse von Produktentwicklern⁶ adressiert, jedoch so generisch ist, dass es auch für den Einsatz in anderen Anwendungsbereichen geeignet ist. Während die Konzepte zur graphischen Benutzeroberfläche (GUI) sowie zur Anfrageverarbeitung nicht im Fokus dieser Arbeit liegen⁷, wird im Speziellen der Indexierungsprozess zur Bereitstellung der suchbaren Informationen betrachtet. Diesbezüglich wurde ein Indexierungsframework für multiple Artefakte auf Basis von Facetten entwickelt, welches den Schwerpunkt der vorliegenden Arbeit bildet.

Im Folgenden wird zunächst die zugrunde liegende Problemstellung erläutert, indem der in der Arbeit fokussierte Anwendungsbereich der technischen PE näher betrachtet wird. Dabei werden insbesondere die in diesem Bereich zu bewältigenden Herausforderungen beleuchtet und der Informationsbedarf von Produktentwicklern analysiert. Darauf folgen eine Einordnung der Arbeit in den Bereich des IWM, eine Detaillierung der Forschungsfragen und Ziele sowie des damit verbundenen Aufbaus der Arbeit.

1.1. Problemstellung

Die Entwicklung innovativer Produkte ist für viele Unternehmen ein wesentlicher Erfolgsfaktor, um nicht nur am Markt bestehen, sondern um sich v. a. gegenüber Konkurrenten am Markt behaupten zu können. Dabei ist die Aufgabe der technischen PE allerdings mit zahlreichen Herausforderungen verbunden, die im Folgenden dargestellt werden.

⁵Bayerischer Forschungsverbund für Prozess- und Workflowunterstützung zur Planung und Steuerung der Abläufe in der Produktentwicklung (<http://www.forflow.org>).

⁶Für einen vereinfachten Sprachgebrauch wird im Weiteren durchgängig die männliche Bezeichnung zur Berufsbeschreibung verwendet, wobei aber immer auch das weibliche Geschlecht mit eingeschlossen ist.

⁷Detaillierte Informationen hierzu enthält die Dissertation von Raiko Eckstein am gleichnamigen Lehrstuhl.

1.1.1. Herausforderungen in der technischen Produktentwicklung

Der erhöhte Wettbewerbsdruck zwingt Unternehmen qualitativ hochwertige Produkte oder Produktvarianten in möglichst kurzer Zeit auf den Markt zu bringen. Als wesentlicher Aspekt hierbei ist die stetig wachsende Produktkomplexität zu berücksichtigen, welche gemäß Abbildung 1.2 Folge diverser Ursachen ist, selbst aber auch herausfordernde produkt- und prozessseitige Auswirkungen nach sich zieht. So ruft der rasante Technologiefortschritt den Wunsch nach immer besserer und hochwertigerer Produktfunktionalität hervor, was dazu führt, dass moderne Produkte nicht mehr nur aus rein mechanischen Bauteilen bestehen [Les01]. Vielmehr sind sie dem Bereich der Mechatronik⁸ zuzuschreiben, da zunehmend elektronische und elektrotechnische Komponenten integriert werden. So wird beispielsweise laut einer Studie von McKinsey & Company⁹ der Wertanteil von elektronischen Systemen in Fahrzeugen bis zum Jahr 2015 auf einen Wert von mehr als 40 % steigen [MT03]. Aber nicht nur die Elektronik, auch andere Disziplinen wie die Softwareentwicklung, die Akustik oder die Optik sind immer häufiger an der Entwicklung eines Produktes beteiligt. Demzufolge ist die Entwicklung von Produkten zwischenzeitlich als eine interdisziplinäre Aufgabenstellung zu verstehen, die mit einem hohen Administrations- und Koordinationsaufwand verbunden ist [Moe09, Sch02a]. Dieser Effekt wird zusätzlich durch die Tatsache verstärkt, dass im Rahmen eines Simultaneous Engineering eine parallele Bearbeitung der mit der PE gekoppelten Abläufe (z. B. mit der Produktionsmittelpaltung) zur Verkürzung der Gesamtentwicklungszeit stattfinden muss. Hierbei sind aber nicht nur die eigenen, sondern auch die Prozesse zahlreicher Zulieferer einzubeziehen, was durch eine mit der Globalisierung einhergehende verteilte Produktentwicklung verkompliziert wird. [Con05, PBFG07, Sch05b]

Ferner entsteht die Komplexität der Produkte aber auch durch die vermehrte Individualisierung der Kundenanforderungen, was in vielen Unternehmen eine deutliche Verschiebung des Leistungsprozesses von der Massenproduktion hin zu einer kundenindividuellen Produktion bewirkt. Anstatt Produkte in großen Stückzahlen zu produzieren, müssen Unternehmen abhängig von den jeweiligen Wünschen und Forderungen der Kunden unterschiedliche Varianten eines Produktes anbieten können [SW08, Les01]. Gleichzeitig bleibt jedoch die Forderung nach kostengünstiger Entwicklung und Fertigung bestehen. Demzufolge sind Unternehmen nicht nur einem hohen Qualitäts- und Zeit-, sondern auch einem hohen Kostendruck ausgesetzt. Dieser schlägt sich insbesondere auf die PE nieder, da laut Gusig etwa 70 % der Herstellkosten eines Produktes allein im Bereich der Entwicklung und Konstruktion festgelegt werden [Gus08].

⁸Der Begriff Mechatronik wurde 1969 ursprünglich als Markenname vom Präsidenten des Elektronikonzerns YASKAWA, Ko Kikuchi, aus den beiden Begriffen „Mechanism“ und „Electronics“ gebildet [Moe09].

⁹<http://www.mckinsey.com>

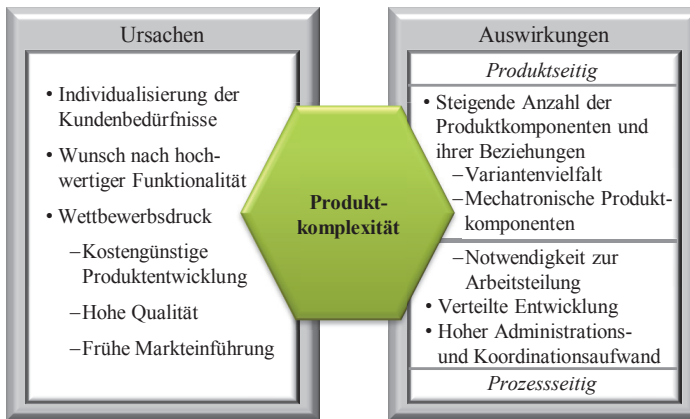


Abbildung 1.2. — Ursachen und Auswirkungen der Produktkomplexität (nach [Les01, S. 1])

Gleichzeitig bewirkt die zunehmende Produktkomplexität auch eine höhere Prozesskomplexität. Der Produktentwicklungsprozess (PEP) besteht im Allgemeinen aus mehreren, nicht-sequentiell ablaufenden und schwer strukturierbaren Prozessphasen, in denen die Beteiligten nach einer konstruktiven Lösung für ein Problem — die Entwicklung eines neuen oder verbesserten Produktes — suchen und diese bestenfalls realisieren [Ver79, PBFG07]. Dabei handelt es sich bei der Lösungsfindung selbst um einen Prozess intensiver Informationsverarbeitung, in dem Informationen von unterschiedlicher Art, unterschiedlichem Inhalt und unterschiedlichem Umfang benötigt, verarbeitet und erstellt werden [PBFG07, Dör87]. Da diese, bedingt durch eine zunehmende Rechnerunterstützung sowie vorgeschriebene Langzeitarchivierungspflichten (vgl. [Rat85]), zum größten Teil ebenfalls in digitaler Form archiviert werden, finden sich im Entwicklungsumfeld zahlreiche heterogene Systeme der Informationstechnik (IT), welche sich in erzeugende und prozessbegleitende Systeme differenzieren lassen. Während Erzeugersysteme zur Generierung konkreter Informationsarten in bestimmten Prozessphasen dienen und sich somit auf Bearbeitungsaufgaben fokussieren, unterstützen prozessbegleitende Systeme die gesamte Produktentstehung¹⁰. Sie sind nicht auf einzelne Phasen ausgerichtet und werden vorrangig zu Kommunikations- und Dokumentationszwecken verwendet [Les01, Ver02, Pul04]. Tabelle 1.1 gibt einen Überblick über die am häufigsten eingesetzten IT-Systeme mit ihrer jeweiligen Einordnung in die beiden Gruppen. Zur Informationsspeicherung nutzen diese Systeme

¹⁰Der Produktentstehungsprozess umfasst alle Vorgänge bis zur Auslieferung eines Produktes an den Kunden bzw. Markt und bezieht somit neben der Produktentwicklung auch die Produktherstellung mit ein [PBFG07, Reu08].

jeweils eigene, spezifische Datenformate, die üblicherweise nicht kompatibel zueinander sind und durchaus auch identische Informationen enthalten können. Demzufolge sind Produktentwickler heutzutage zwar mit einer Flut an Informationen konfrontiert, allerdings beseitigt dies den häufig als Schwachstelle identifizierten Informationsmangel nicht. Stattdessen wird dieser aufgrund von Redundanzen, Inkonsistenzen und Versionsvielfalt eher noch verstärkt [BEE⁺97, Vaj05].

Tabelle 1.1. — Funktionsorientierte Einordnung von IT-Systemen der PE als erzeugende oder prozessbegleitende Systeme: + Hauptfunktionalität, o Teilfunktionalität und - keine Funktionalität im Aufgabenbereich der jeweiligen Systemgruppe

IT-System	erzeugend	prozessbegleitend
CAD-System	+	-
Enterprise Resource Planning (ERP)-System	o	+
Dokumentenmanagementsystem (DMS)	-	+
Content Management System (CMS)	o	+
Produktdatenmanagementsystem (PDMS)	-	+
Datenbanksystem (DBS) / Data Warehouse	o	+
Workflowmanagementsystem (WfMS)	-	+
Projektmanagementsystem (PMS)	o	+
Supplier Relationship Management (SRM)-System	o	+
Customer Relationship Management (CRM)-System	o	+
Virtual Reality (VR)-System	+	o
Simulations- / Berechnungswerkzeug	+	-
Bürosoftware	+	-
Groupware-System	o	+
Dateiverwaltungssystem	-	+

Obwohl viele der existierenden Systeme über individuelle Suchfunktionalitäten verfügen, ist das Auffinden von Informationen noch immer mit einem relativ großen Zeitaufwand verbunden. Da dieser sowohl von der Branche als auch von der jeweiligen Aufgabenstellung abhängt, finden sich in der Literatur unterschiedliche Schätzungen hierzu. So veranschlagen beispielsweise sowohl Beitz et al. als auch Ehrlenspiel etwa 20–30 % der gesamten Entwicklungszeit [BEE⁺97, Ehr07], Wallace und Ahmed 26 % [WA03] und Przywara 5–15 % [Prz08a] der täglichen Arbeitszeit eines Produktentwicklers allein für die Tätigkeit der Informationsbeschaffung. Dies ist u. a. darauf zurückzuführen, dass sich die vorhandenen Suchmechanismen nur auf einzelne, spezielle Informationsbedürfnisse, wie beispielsweise die Suche nach Produkten anhand von Produktdaten, konzentrieren. Somit stehen Insellösungen bereit, die zwar für

bestimmte Suchsituationen durchaus hilfreich sind, welche aber keine übergreifende Suche ermöglichen. Ein weiterer Grund für den erheblichen Suchaufwand ist die Tatsache, dass es bislang eigentlich keine, von allen Beteiligten einheitlich angewandte Datenspeicherung gibt. Trotz oder gerade aufgrund der großen Systemvielfalt erfolgt die Datenablage häufig noch stark personenorientiert entweder im eigenen, meist lokalen Dateiverzeichnis oder zumindest in einem gemeinsamen Projektverzeichnis — allerdings unter Verwendung wenig strukturierter Hierarchiepfade und wenig aussagekräftiger Bezeichnungen. Folglich ist Kollegen die Existenz einer Information meist gar nicht bekannt [Prz08a]. Dies wiederum führt auch aufgrund mangelnder Kommunikation zu einer geringen Transparenz im Unternehmen, die in einer redundanten Ausführung von Tätigkeiten und damit einer redundanten Generierung und Speicherung von Informationen resultiert. Dieses »das Rad neu erfinden« verursacht somit höhere Kosten als die Nutzung bereits vorhandener Produkte, deren Wiederverwendung jedoch ein schnelles Auffinden mit allen zugehörigen und benötigten Informationen voraussetzt. Obwohl seit einigen Jahren Anstrengungen im Bereich Enterprise Search (ES) unternommen werden, um vorhandenes Wissen unternehmensweit verfügbar zu machen, scheitern die bisherigen Ansätze insbesondere an den Besonderheiten des Produktentwicklungsbereiches (vgl. hierzu Kapitel 2.4). Somit liegt Information bzw. Wissen brach und bleibt ungenutzt. Da durch verkürzte Produktlebenszyklen jedoch immer weniger Zeit bleibt die Entwicklungskosten zu erwirtschaften, birgt gerade das Auffinden und Wiederverwenden dieser bereits vorhandenen Informationen ein enormes Potenzial zur Reduzierung nicht nur der Entwicklungszeit, sondern v. a. auch der Entwicklungskosten.

Um allerdings eine verbesserte Informationsversorgung von Produktentwicklern gewährleisten zu können, ist es zunächst notwendig, deren Informationsbedarf genauer zu analysieren.

1.1.2. Analyse des Informationsbedarfs von Produktentwicklern

Das Auffinden einer Lösung für ein durch Kunden- bzw. Marktanforderungen und -wünsche gegebenes Problem und deren Umsetzung in ein verkaufsfähiges Produkt erfordert spezifische Informationen über den Problembereich, welche durch bewusstes Denken (d. h. durch die Anwendung geistiger Operationen, Regeln usw.) ausgewertet, verarbeitet und verändert werden [Dör87]. Da es allerdings in der Natur des Menschen liegt, nicht allwissend zu sein, erfordert jeder Problemlösungsprozess — und somit auch der PEP — mindestens eine Phase der Informationsbeschaffung, in der durch eine Suche die vorhandenen Informationslücken geschlossen werden. Demnach ist die Analyse des Informationsverhaltens von Problemlösern zu einem eigenständigen Forschungsfeld avanciert, in dem seit vielen Jahrzehnten zahlreiche Studien durchgeführt werden. Einen Überblick hierzu liefern King et al. in [KCJ94], wobei

insbesondere das Kommunikations- und Suchverhalten sowie der Informationsbedarf von Ingenieuren und Wissenschaftlern im Allgemeinen untersucht wurden. Obwohl die aus diesen Studien resultierenden primären Ergebnisse (z. B. der Bedarf an Fachinformationen aus Büchern und Zeitschriften oder die Wichtigkeit des persönlichen Gesprächs mit Experten) selbstverständlich auch für Produktentwickler zutreffen, soll ihr Informationsbedarf in der vorliegenden Arbeit genauer spezifiziert werden. Hierzu ist grundsätzlich zwischen den beiden Arten des objektiven und subjektiven Informationsbedarfs zu differenzieren, welche im Folgenden näher betrachtet werden.

Aufgabenspezifische Betrachtung des objektiven Informationsbedarfs

Gemäß Stock und Koreimann ist der objektive Informationsbedarf dadurch charakterisiert, dass dieser nur von der zugrunde liegenden Aufgabenstellung und nicht von der konkreten, diese Aufgabe ausführenden Person abhängt [Sto07, Kor76]. Der objektive Informationsbedarf ist daher allein aus der Tätigkeit, die ein Produktentwickler im PEP erfüllen muss, vorhersehbar. Dabei ist jedoch der Begriff der Tätigkeit bzw. der Aufgabenstellung sehr allgemein gehalten, so dass dieser für eine detaillierte Analyse näher spezifiziert werden muss. Dafür werden die folgenden drei Faktoren herangezogen:

- die *Art der Gesamtaufgabe* (Konstruktionsart), in deren Kontext die Aufgabenstellung zu erfüllen ist,
- der *Zeitpunkt*, zu dem die Tätigkeit im PEP durchzuführen ist, und
- die *Rolle*, die für die Erledigung der jeweiligen Tätigkeit verantwortlich ist.

Prinzipiell ist bei der Entwicklung von Produkten zwischen den drei Konstruktionsarten Neukonstruktion, Anpassungskonstruktion und Variantenkonstruktion zu unterscheiden. Zweck der Neukonstruktion ist es, ein noch nicht existierendes, innovatives Produkt durch die Erarbeitung eines neuartigen Lösungsprinzips zu entwickeln. Dabei sind meist neue und unbekannte Themenfelder, Problembereiche und technische Gesetzmäßigkeiten zu erforschen sowie konkrete Teilprobleme, z. B. bei der Ausgestaltung des Produktes, anzugehen. Um allerdings zu letzteren zu gelangen, sind zunächst neue Ideen zu generieren, weshalb Produktentwickler anhand unterschiedlichster Informationen nach inspirierendem Material suchen [MBA07]. Somit umfasst der objektive Informationsbedarf bei derartigen Aufgabenstellungen eine Vielfalt an sowohl internen als auch externen Informationen zu unterschiedlichen Bereichen und Domänen. Markt- und Konkurrenzanalysen, Abbildungen und Funktionsprinzipien existierender technischer sowie natürlicher Systeme, Berichte über neuartige Werkstoffe und Fertigungsverfahren, Richtlinien und Gesetze, Umweltvorschriften oder technische Spezifikationen sind nur einige Beispiele hierfür [Hab08]. Solche Neukonstruktionen kommen jedoch relativ selten vor, da laut Sandler und Waver 80 % der Kundenanforderungen durch Anpassungs- und Variantenkonstruktionen, welche auf bereits existieren-

den Entwicklungen aufbauen, realisiert werden können [SW08]. Während bei einer Anpassungskonstruktion ein bestehendes Produkt an neue Anforderungen anzupassen ist und dabei Neuentwicklungen einzelner Baugruppen oder -teile erforderlich sein können, beschränkt sich die Variantenkonstruktion auf die Modifikation einzelner Abmessungen oder Anordnungen bei sonst gleichbleibender Gestalt der Systemelemente. Somit setzen die beiden letzteren Konstruktionsarten erst später im PEP an und erfordern zu ihrer Erfüllung eher weniger, dafür aber umso spezifischere Informationen, wie beispielsweise formähnliche Bauteile oder Lieferantendatenblätter. [AB⁺07, Reu08, PBFG07, Ver79]

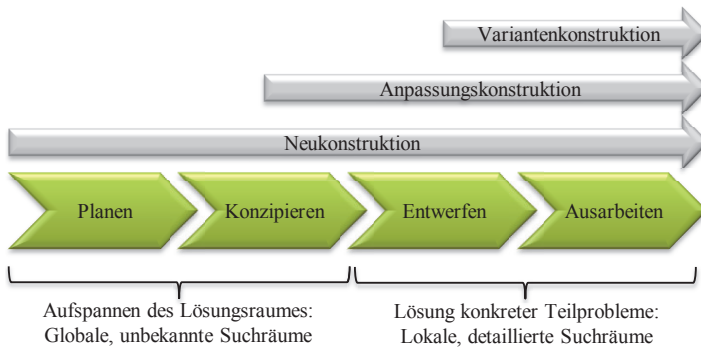


Abbildung 1.3. — Einordnung der Konstruktionsarten und Suchräume anhand der vier von Pahl et al. in [PBFG07] definierten Prozessphasen

Die Unterscheidung des Bedarfs in *viele, eher allgemeine* und *wenige, aber konkrete* Informationen bzw. die damit einhergehende von Sarkar und Chakrabarti in [SC07] identifizierte Differenzierung in zunächst *globale, unbekannte* und schließlich *lokale, detaillierte* Suchräume geht dabei mit der Abhängigkeit der Aufgabe von ihrem Durchführungszeitpunkt im PEP einher (vgl. Abbildung 1.3). Betrachtet man sich den PEP anhand verschiedener, die Makro-Logik des Prozesses beschreibender¹¹ Vorgehensmodelle (VM), wie beispielsweise den vom Verein Deutscher Ingenieure (VDI) festgelegten Richtlinien 2221 - 2223 [Ver86, Ver97, Ver77, Ver99] und 2206 [Ver04] oder den Konstruktionsmethodiken von Pahl/Beitz [PBFG07] und Koller [Kol98], so empfehlen alle im Prinzip das gleiche Vorgehen — eine Entwicklung vom Groben zum Detail¹². Dabei konzentrieren sich die frühen Phasen der Entwicklung zunächst auf

¹¹Im Sinne des Systems Engineering bestehen VM u. a. aus einem Phasenkonzept (Makro-Logik), das den Ablauf eines Prozesses in seiner Gänze beschreibt, und einem Problemlösungszyklus (Mikro-Logik), dessen elementare Handlungsabläufe in jeder Phase zur konkreten Problemlösung dienen. [BBH⁺02]

¹²siehe Anhang A zum Vergleich der Vorgehensmodelle

strukturorientierte Tätigkeiten, in denen die Aufgabenstellung zu präzisieren sowie mögliche Lösungsprinzipien zu konzipieren sind. Dies erfordert v. a. allgemeine Informationen, um Informationsdefizite beseitigen, das Wissen über den Problembereich erweitern und den möglichen Lösungsraum aufspannen zu können. In den späteren Phasen hingegen sind eher geometrieorientierte und detaillierende Aufgaben zur Ausgestaltung des ausgewählten Lösungsprinzips und zur Anfertigung der Fertigungsunterlagen durchzuführen. Dabei sind einzelne Teilprobleme bzw. -funktionalitäten (z. B. die Gestaltung oder Simulation der einzelnen Systemelemente) umzusetzen, für die wiederum speziellere, überwiegend interne Informationen (z. B. Werksnormen, Maßangaben zu Produkten oder Bauräumen) benötigt werden. [PBF07, Sch02b]

Des Weiteren ist zur Konkretisierung der Aufgaben eines Produktentwicklers zu berücksichtigen, dass mit dem Begriff Produktentwickler in der vorliegenden Arbeit alle an der Entwicklung eines Produktes mitwirkenden Personen gemeint sind. Allerdings ist jede dieser Personen abhängig von der Aufbauorganisation des Unternehmens nur für eine bestimmte Rolle im PEP — eventuell auch für mehrere — und den sich daraus ergebenden Aufgaben verantwortlich. Während z. B. ein Konstrukteur zur Erstellung von CAD-Modellen Informationen über bereits existierende Produkte und ihre Schnittstellen benötigt, erfordert die Durchführung von Simulationen von einem Tester Wissen über den verwendeten Werkstoff und seine Belastungsgrenzen sowie über physikalische Gesetzmäßigkeiten. Folglich ist der objektive Informationsbedarf abhängig von der jeweils eingenommenen Rolle unterschiedlich ausgeprägt.

Allgemeine Betrachtung des objektiven Informationsbedarfs

Neben dieser ersten, doch sehr aufgabenspezifischen Charakterisierung des objektiven Informationsbedarfs können noch andere Ansätze angewendet werden, die allerdings auf einer eher allgemeinen Betrachtung der Entwicklungstätigkeit basieren. So werden generell die bereits im Unternehmen vorhandenen Entwicklungsinformationen als wichtige Teilmenge des objektiven Informationsbedarfs gesehen [SD01, SC07, WA03]. Dabei werden unter dem Begriff Entwicklungsinformationen in Anlehnung an Baya et al. alle Daten verstanden, die während des gesamten PEP erstellt, benutzt, referenziert und konsultiert wurden [BGB⁺92]. Durch ihre Verwendung können nicht nur Entwicklungskosten und -zeit reduziert, sondern auch das Risiko fundamentaler Fehler aufgrund bereits durchgeführter Tests und vorhandener Erfahrung vermindert und somit die Produktqualität erhöht werden. Auch Ullman konstatiert in [Ull03], dass es sich bei den meisten Entwicklungsaufgaben um Modifikationen auf Basis bereits vorhandener Problemlösungen, Produkte, Testergebnisse usw. handelt, was ein Auffinden dieser vorhandenen Informationen mittels eines IRS erforderlich macht. Ein wesentlicher Aspekt hierbei ist die Suche nach ähnlichen Objekten (Problemen, Ideen, Produkten, ...), welche von Carstensen als einer von zwei wesentlichen Suchbelangen in der technischen PE identifiziert wurde [Car97]. Allerdings interessieren

hier häufig nicht nur die gefundene Lösung, sondern auch die Gründe, die zu dieser geführt haben [WA03, Prz08a]. Diese als Design Rationales (DR) oder auch Design History bezeichnete Entscheidungsdokumentation ist somit ebenfalls ein wichtiger Bestandteil des Informationsbedarfs von Produktentwicklern. Daneben stellt, wie bereits zu Beginn genannt, die Suche nach Personen, die Erfahrungen bzw. Wissen in bestimmten Bereichen haben und die man bei Problemen kontaktieren kann, das zweite wesentliche Suchvorgehen dar. Deshalb wird viel häufiger die Frage *Wer könnte etwas darüber wissen?* als die Frage *Wo finde ich Informationen darüber?* gestellt [Car97, HP00, WA03]. Weiterhin sind bei jeder Entwicklungsaufgabe gewisse Gestaltungsrichtlinien — sogenannte Design for X (DfX)-Kriterien — hinsichtlich verschiedener Aspekte wie Kosten, Sicherheit, Ergonomie, Umwelt usw. einzuhalten [Con08a]. Da hierfür viel Spezialwissen erforderlich ist, benötigen insbesondere unerfahrene Produktentwickler Informationen darüber, wie sie diese Richtlinien erfüllen können. Die Bereitstellung von entsprechendem Methodenwissen, Lessons Learned oder Best Practices kann hier zu einer besseren Informationsversorgung beitragen.

Um die in den bisherigen Abschnitten aufgezeigte Informationsvielfalt einigermaßen handhaben zu können, wurden in der Literatur Klassifizierungsversuche vorgenommen, welche im Folgenden kurz aufgezeigt werden. So differenziert beispielsweise Mewes in [Mew73] zwischen funktions-, fertigungs-, gebrauchts-, kosten- und organisationsorientierten Informationen. Da erstere zur Ermittlung von Alternativlösungen und somit zum direkten Fortgang der Entwicklung dienen (z. B. Problemlösungen, Fachliteratur oder Berechnungsinformationen), sind sie in allen Phasen der Entwicklung notwendig. Fertigungs-, gebrauchts-, kosten- und organisationsorientierte Informationen dagegen weisen einen einschränkenden Charakter auf und dienen vorrangig zur Bewertung der erstellten Alternativlösungen anhand des jeweiligen Aspektes. Sie sind folglich nur in bestimmten Phasen des PEP erforderlich. Carstensen hingegen gruppiert in [Car97] die bei der Entwicklung benötigten Informationen nach Fachbereichen in technische Informationen (Spezifikationen, Standards, Normen, technische Zeichnungen, Testergebnisse, ...), Kundeninformationen (Anforderungen, Wünsche, Erwartungen, ...), Marketinginformationen (Markt-, Konkurrenzanalysen, Preise, ...), Produktionsinformationen (Informationen zu Herstellungsverfahren, Werkstoffen, Kosten, Sicherheitsvorschriften, ...), Wartungs- bzw. Instandhaltungsinformationen (Betriebs-, Montage-, Wartungsanleitungen, ...) und Entsorgungsinformationen (Umweltschutzauflagen, -richtlinien, -gesetze, ...). Des Weiteren besitzen aber auch an Problemlösungsprozessen allgemein orientierte Klassifikationen für den Bereich der technischen PE Gültigkeit, wie beispielsweise die von Müller in [Mül90] vorgestellte Klassifikation zeigt. Hier unterscheidet Müller anhand der Funktion, die eine Information im Prozess erfüllt, zwischen Ziel-, Sach-, Prozedur- und Steuerinformationen. Während Zielinformationen die Aufgabenstellung charakterisieren und als Anforderungen zum Klären des zu erreichenden Ziels beitragen, umfassen Sachinformationen alle Daten, die Fakten hinsichtlich der Aufgabe beschreiben und festlegen

(z. B. technische Sachverhalte in Form von Kennwerten oder Beschaffenheitsmerkmalen). Prozedurinformationen hingegen geben Hinweise darüber, wie das Ziel erreicht werden kann (Methoden, Regeln, Vorschriften, . . .), und Steuerinformationen, unter welchen organisatorischen Rahmenbedingungen (z. B. Projekt-, Terminpläne, . . .) dies erfolgt.

Zusammenfassend haben die vorherigen Abschnitte gezeigt, dass der objektive Informationsbedarf und damit das Spektrum der von Produktentwicklern nachgefragten Informationen sehr weit gefasst ist und eine Vielzahl an sehr unterschiedlichen Informationen enthält. Welche davon jedoch konkret benötigt werden, ergibt sich erst im Prozessverlauf — also abhängig von der konkreten Entwicklungssituation — und kann nicht eindeutig vorhergesagt werden. Jedoch gilt auch hier, wie bereits allgemein von Byström und Järvelin festgestellt [BJ95], dass der objektive Informationsbedarf umso komplexer ist, je höher die Komplexität der Aufgabenstellung selbst ist. Außerdem lässt sich zeigen, dass dieser immer durch eine Menge von Informationskomplexen charakterisiert ist, welche zur Durchführung der jeweiligen Aufgabe(n) erforderlich sind. Ein Informationskomplex — im Weiteren als *Artefakt* bezeichnet — setzt sich dabei gemäß Mewes aus mehreren Informationseinheiten zusammen, die sich aus der Kombination von Informationselementen, bestehend aus Informationssymbolen (Buchstaben, Ziffern, Symbolen), ergeben [Mew73]. So stellen beispielsweise die Beschreibung eines technischen Systems oder auch die Informationen einer Zeichnung einen Informationskomplex dar. Folglich umfasst der objektive Informationsbedarf von Produktentwicklern verschiedene, sowohl extern als auch intern verfügbare (und damit auch wiederverwendbare) Artefakte wie beispielsweise Produkte, Dokumente, Werkstoffe, Projekte, Personen oder Methoden. Diese domänenspezifischen Artefakte lassen sich in sogenannte *Artefakttypen* gruppieren, indem gleichartig strukturierte Artefakte, welche sich nur durch die Ausprägungen ihrer Eigenschaften voneinander unterscheiden, zu einem Typ zusammengefasst werden.

Betrachtung des subjektiven Informationsbedarfs

Wie bereits im vorherigen Abschnitt erwähnt, resultiert der objektive Informationsbedarf allein aus der Aufgabensicht und ist somit unabhängig von der ausführenden Person. In einer konkreten Suchsituation entscheidet der Produktentwickler allerdings aus seiner Perspektive heraus, d. h. basierend auf seinem Hintergrundwissen, seinen Erfahrungen und Kenntnissen, seinen vergangenen Aufgaben, seinen psychologischen und kognitiven Eigenschaften sowie seinem individuellen Arbeitsstil, welche Artefakte er zur Durchführung der jeweiligen Aufgabe für notwendig erachtet [Ste01]. Dieser subjektive Informationsbedarf leitet sich somit sowohl von der Aufgabe als auch von der konkreten, anfragenden Person bzw. der sie auszeichnenden Kontextfaktoren ab und wird daher häufig auch als Informationsbedürfnis bezeichnet [Sto07, Kor76]. Dabei kann das Informationsbedürfnis durchaus eine Teilmenge des objektiven Infor-

mationsbedarfs abdecken; es kann aber auch Artefakte umfassen, die zwar objektiv betrachtet nicht zur Aufgabenerfüllung erforderlich sind, die aber von der jeweiligen Person als notwendig dafür angesehen werden. Ein typischer Anwendungsfall hierfür ist die Nutzung von Assoziationen [Sto07] zwischen Artefakten, um ausgehend vom subjektiven bzw. geäußerten Informationsbedarf sozusagen über Umwege schließlich zur Befriedigung des objektiven Informationsbedarfs zu gelangen. Dieses Verhalten zeigt sich auch bei Produktentwicklern, welche hauptsächlich produktbezogen denken. Sie beginnen in den meisten Fällen mit einer Suche nach hilfreichen Produkten, um schließlich in einem zweiten Schritt die tatsächlich benötigten Artefakte, z. B. zugehörige Dokumente oder Ansprechpartner, zu finden. Auch die bereits erwähnte Suche nach Personen ist eigentlich nur ein Zwischenschritt, um über den persönlichen Kontakt die tatsächlich benötigten Dokumente oder Informationen zu erhalten [HP00].

Grundsätzlich sollten IRS in der Lage sein, sowohl den objektiven als auch den subjektiven Informationsbedarf eines Nutzers befriedigen zu können. Allerdings hängt die Akzeptanz derartiger Systeme v. a. von ihrer Fähigkeit, die vom Nutzer als notwendig erachteten Informationen zu liefern, ab. Dabei ist zu beachten, dass der subjektive Informationsbedarf und die tatsächliche Anfrage einer Person — also der von ihr geäußerte Informationsbedarf — nicht übereinstimmen müssen. Gerade in Situationen, in denen konkrete Informationen, d. h. Fakten, benötigt werden und somit eine zielgerichtete Vorstellung über das Suchobjekt vorliegt, können auf relativ einfache Weise exakte, dem subjektiven Informationsbedarf entsprechende Anfragen formuliert werden. Die technische PE ist jedoch eher durch problemorientierte Suchsituationen geprägt, in denen Produktentwickler nicht eindeutig wissen, wonach sie suchen. Dies liegt daran, dass das Gesuchte eben meist nicht oder zumindest noch nicht genau bekannt ist oder dass es erst bewusst wird, wenn man es zufällig findet bzw. darauf stößt [WA03, Ahm06a]. Demzufolge können Informationsanfragen in derartigen Situationen nur vage formuliert werden und repräsentieren damit oft nur einen Bruchteil des eigentlichen subjektiven Informationsbedarfs.

Beispiele für komplexe Suchsituationen

Der Informationsbedarf von Produktentwicklern lässt sich, wie bereits erläutert, nicht eindeutig definieren. Vielmehr wird er durch verschiedene Randbedingungen beeinflusst, was in unterschiedlichen Ausprägungen bzw. Zuständen des konkreten Informationsbedarfs resultiert. Da die Eigenschaft, eine große Anzahl an unterschiedlichen Zuständen annehmen zu können, als Komplexität bezeichnet wird [Sch05b], sind Suchsituationen in der technischen PE demzufolge mit einer hohen Komplexität zu charakterisieren. Dies ist insbesondere auf die Vielzahl der benötigten Artefakte, ihre unterschiedlichen Erscheinungsformen und damit verbundenen Informationsstrukturen, ihre Archivierung in unterschiedlichsten Softwaresystemen sowie ihre Beziehun-

gen zueinander zurückzuführen. Zur Verdeutlichung dieser Komplexität werden im Folgenden beispielhafte Suchszenarios erläutert.

Suchszenario A: Herr A hat die Aufgabe, zwei Systemelemente miteinander zu verbinden. Grundsätzlich gibt es zur Lösung dieser Aufgabe verschiedene Möglichkeiten, so dass der objektive Informationsbedarf sämtliche Informationen zu allen Lösungsmöglichkeiten umfasst. Aufgrund seiner Erfahrungen weiß Herr A allerdings, dass nur *Schraubverbindungen* mit einer *Schraubenlänge* von *mindestens 5,6 cm* und einer bestimmten *Gewindeausführung* (z. B. einem metrischen Normgewinde) hierfür geeignet sind. Demzufolge besteht sein subjektiver Informationsbedarf in der Menge aller Schraubverbindungen, welche die entsprechenden Eigenschaften aufweisen.

Suchszenario B: Bei der Entwicklung eines neuartigen, innovativen Kotflügels handelt es sich um eine komplexe Aufgabe, bei der Produktentwickler mit unterschiedlichen Fragestellungen konfrontiert sind (vgl. hierzu die Beispiele in Abbildung 1.4). Im vorliegenden Szenario ist Herr B mit der Erstellung des Pflichtenheftes beauftragt. Da es sich bei Herrn B um einen jungen und unerfahrenen Produktentwickler handelt, der erst wenige Monate im Unternehmen arbeitet, benötigt er Hilfestellungen, die ihm bei der Bewältigung dieser Aufgabe nützlich sein könnten. Folglich ist der subjektive Informationsbedarf von Herrn B dadurch charakterisiert, dass Herr B bereits im Unternehmen vorhandene und freigegebene *Pflichtenhefte* sucht, an deren Struktur und Layout er sich orientieren kann und die eventuell bereits ähnliche Anforderungen enthalten.



Abbildung 1.4. — Beispiele für mögliche Fragestellungen bei der Entwicklung

Suchszenario C: Herr C hat die Aufgabe ein Produkt zu entwickeln, das eine *Hitzebeständigkeit bis zu 75°C* aufweist. Folglich besteht der objektive Informationsbedarf darin, nach *Werkstoffen* zu suchen, die diese Hitzebeständigkeit gewährleisten. Da Herr C keine Erfahrungen mit der in Frage kommenden Werkstoffgruppe besitzt, fokussiert sich sein subjektiver Informationsbedarf auf die Suche nach bereits existierenden *Testberichten* zu den in Frage kommenden Werkstoffen, um deren Ergebnisse schließlich für den Werkstoffauswahlprozess heranzuziehen.

Suchszenario D: Aufgabe von Herrn D ist die Konstruktion eines Lüfters im CAD-System. Die Zusammenstellung des Lüfters aus seinen Einzelteilen erfordert dabei die Verwendung einer *ringförmigen Dichtung*, die über einen *Außenradius von exakt 2,25 cm* und einen *Innenradius von mindestens 1,5 cm* verfügt. Allerdings benötigt Herr D nicht das reale, physische Produkt zur Erfüllung seiner Aufgabe, sondern das zugehörige *CAD-Modell*, welches er im CAD-System in das Gesamtprodukt an entsprechender Stelle einfügen kann. Zusätzlich hat Herr D bei der Verwendung von Standardprodukten allgemein darauf zu achten, möglichst Produkte von Lieferanten zu verwenden, bei denen bereits Bestellungen geplant sind, um durch Nutzung von Bestellmengenrabatten die Einkaufskosten senken zu können.

Suchszenario E: Aufgrund einer schlechten konjunkturellen Lage ist Herr E gezwungen, sich eine neue Arbeitsstelle zu suchen. Als erfahrener Entwickler muss er zwar nicht lange suchen, kann allerdings nicht mehr in seinem Spezialgebiet arbeiten. Nach jahrelanger Entwicklung von Bremssystemen muss sich Herr E einem völlig neuen Aufgabengebiet, der Entwicklung von Tüllen für die Verlegung von Kabelleitungen in Bordnetzsystemen, widmen. Da Herr E sehr wenige Kenntnisse auf diesem Gebiet hat, besteht sein subjektiver Informationsbedarf darin, möglichst viele Informationen bezüglich der *Tüllenentwicklung* zu erlangen. Folglich sucht er zum einen nach *Ansprechpartnern*, die er kontaktieren kann, und zum anderen nach *Dokumenten*, die Informationen zu Tüllen enthalten.

1.2. Einordnung der Arbeit

„Wenn Siemens wüsste, was Siemens weiß.“ [Pie00, 2. Abschnitt] Dieses aus dem Jahr 2000 stammende Zitat von Heinrich von Pierer, dem ehemaligen Vorstandsvorsitzenden der Siemens AG¹³, verdeutlicht das enorme Potential, welches in der Nutzung des in Unternehmen vorhandenen Wissens steckt; welches jedoch in vielen, insbesondere kleinen und mittelständischen Unternehmen selbst heute

¹³<http://w1.siemens.com/entry/de/de/>

noch immer nicht oder zumindest nicht in vollem Maße ausgeschöpft wird¹⁴. Dabei hatte es Siemens damals schon geschafft: Durch die Entwicklung eines Wissensmanagementsystems im unternehmenseigenen Intranet, welches alle Projekte des weltweiten Konzerns, zusammen mit ihren Projektbearbeitern und zugehörigen Lösungen, suchbar machte, war der erste Schritt zur Erreichung des Ziels „*Siemens weiß, was Siemens weiß*.“ [Pie00, 2. Abschnitt] gemacht. Obwohl insbesondere das Management von Wissen bereits seit Mitte der 1990er Jahre als Bestandteil der allgemeinen Managementtheorien gesehen wird [Pon07], gehört es erst seit wenigen Jahren zu den wesentlichen Kernaufgaben von Unternehmen aller Branchen. Dabei wird die Versorgung der Mitarbeiter mit den richtigen Informationen zur richtigen Zeit immer wichtiger. Da diese Problematik, angewendet auf den Bereich der technischen PE, im Fokus der vorliegenden Arbeit steht, werden im Folgenden die beiden Fachgebiete IWM und Information Retrieval (IR), in die diese Arbeit einzuordnen ist, sowie deren Zusammenhang ausführlicher betrachtet.

1.2.1. Informations- und Wissensmanagement

Das IWM umfasst grundsätzlich zwei eigenständige Konzepte, die allerdings nicht losgelöst voneinander betrachtet werden können. Dies zeigt sich bereits in der Wortbildung der beiden Komposita. Zum einen enthalten beide das Substantiv *Management* als Grundwort und beschäftigen sich demnach mit der Gestaltung, Lenkung und Kontrolle einer Sache. Zum anderen sind die Konzepte eben anhand dieser Sache, die gemanagt wird, voneinander abzugrenzen. Dabei handelt es sich im einen Fall um *Information*, im anderen Fall um *Wissen* — zwei Begrifflichkeiten, die häufig synonym verwendet werden. Folglich ist zunächst die Frage nach dem Unterschied zwischen Information und Wissen zu klären. Obwohl sich bereits zahlreiche Forscher aus unterschiedlichen Bereichen wie der Philosophie, der Informationstheorie, der Informationswissenschaft oder auch der Nachrichtentechnik mit dieser Frage beschäftigten, gibt es bis heute keine eindeutigen, allgemein anerkannten Definitionen. Aus diesem Grund wird zur besseren Einordnung der Arbeit das ihr zugrunde liegende Begriffsverständnis erläutert. Für eine ausführliche Diskussion wird neben den zahlreichen Literaturquellen zum Thema IWM u. a. auf Court verwiesen, der in [Cou97] den Informationsbegriff in Relation zu den vier Konzepten Daten, Wissen, Gedächtnis und Kommunikation betrachtet.

Information wird grundsätzlich mit Hilfe von Daten repräsentiert, welche selbst als Zeichenfolgen aus der Kombination von Zeichen eines Zeichenvorrats unter Anwendung bestimmter, festgelegter Syntaxregeln entstehen [FSV01]. Derartige Zeichenfolgen stellen für sich alleine betrachtet allerdings noch keine Information dar; erst durch

¹⁴Laut Lehner werden nur etwa 30 % des real vorhandenen Wissens einer Organisation auch tatsächlich genutzt [Leh09].

ihre Einbettung in einen Kontext können diese von einem Individuum als Information aufgefasst und verarbeitet werden [FS08, Krc05, Kuh04]. So stellt die Zeichenfolge *M10x1.5x30* zunächst eine beliebige Aneinanderreihung von alphanumerischen Zeichen ohne jegliche Bedeutung dar. Vor dem Hintergrund der technischen PE wird mit dieser Zeichenfolge allerdings eine Information, nämlich die Beschreibung einer Normschraube anhand bestimmter Parameter (metrisches Gewinde, 10 mm Nenndurchmesser, 1,5 mm Gewindesteigung und 30 mm Schraubenlänge), assoziiert [Cou97]. Demzufolge werden unter Informationen im Weiteren Daten verstanden, die für einen bestimmten Zweck übertragen, gespeichert, interpretiert und verarbeitet werden. Letzteres wirkt sich dabei auf das Wissen eines Individuums oder einer Organisation aus, indem dieses durch Aufnahme und Vernetzung der neuen Information mit dem bereits vorhandenen Wissen erweitert wird. Besonders deutlich wird diese Beziehung in Brookes' elementarer Gleichung $K(S) + \Delta I = K(S + \Delta S)$ ausgedrückt [Bro80]. Das Wissen K wird hierbei als eine Struktur S von Begriffen, Aussagen und ihren Beziehungen verstanden, welche zur Lösung von Problemen eingesetzt wird. Sobald eine neue Information ΔI — welche einen kleinen Ausschnitt einer derartigen Struktur repräsentiert — hinzu kommt, wird diese Struktur verändert (ausgedrückt durch ΔS). Dabei ist allerdings zu berücksichtigen, dass dieselbe Information bei unterschiedlichen Empfängern durchaus unterschiedliche Strukturänderungen oder eventuell auch keine Änderung (in dem Fall würde $\Delta S = 0$ gelten) bewirken kann. Wissen ist damit als Eigenschaft eines Individuums zu sehen, die ganzheitlich ausgerichtet und somit schwer greifbar ist [Mac80]. Um diese jedoch näher beschreiben zu können, findet man in der Literatur eine Reihe von Klassifikationen. Eine sehr weit verbreitete und auf Polanyis Veröffentlichung zum impliziten Wissensbegriff aus dem Jahr 1966 basierende Klassifikation ist die von Nonaka und Takeuchi vorgenommene Differenzierung in explizites und implizites Wissen [Pol66, NT97]. Während man ersteres eindeutig beschreiben und/oder formalisieren und damit dokumentieren kann (z.B. mathematische Sachverhalte oder wissenschaftliche Erkenntnisse), ist implizites Wissen — häufig auch als *tazites* Wissen bezeichnet — verborgen. Es handelt sich vielmehr um das Wissen, das man weiß ohne es in Worte fassen zu können, das „(...) tief verankert [ist] in der Tätigkeit und der Erfahrung des Einzelnen sowie in seinen Idealen, Werten und Gefühlen.“ [NT97, S. 18f] Folglich ist implizites Wissen schwer zugänglich, weshalb nur explizites Wissen — das jedoch nicht direkt übertragbar ist — weitergegeben werden kann [Sto07]. Vielmehr wird dieses nur über Informationen ausgetauscht bzw. nachgefragt, so dass Information gemäß Kühlen die Menge an Wissen ist, die in einer aktuellen Handlungssituation benötigt wird, über die der Handelnde jedoch nicht verfügt [Kuh04]. Diese werden letztendlich wiederum durch Daten (z.B. die Schmelzpunkte bestimmter Werkstoffe) repräsentiert, welche somit eigentlich die benötigte Information und damit das benötigte Wissen darstellen [Wal86].

In Anlehnung an diese Abgrenzung sind auch das Informations- (IM) und Wissensmanagement (WM) differenziert zu betrachten. So besteht die Aufgabe des IM darin,

die Informationsversorgung und damit das informationswirtschaftliche Gleichgewicht im Unternehmen sicherzustellen. Hierfür ist das betriebliche Informationssystem (IS) derart zu planen und zu gestalten, dass Informationen optimal zur Erreichung der Unternehmensziele eingesetzt werden [Krc05]. IM umfasst somit alle Aufgaben zur Beschaffung, Speicherung, Übertragung, Bereitstellung, Verarbeitung und Verteilung von Informationen [HS09], damit Mitarbeiter mit der richtigen Information zur richtigen Zeit am richtigen Ort in der richtigen Menge und Qualität (Sachziel der Informationslogistik) versorgt sind. Allerdings beschränkt sich das IM dabei v. a. auf diejenigen Aufgaben, die mit Hilfe maschineller Aufgabenträger (AT) (d. h. mittels AwS) realisiert werden können [FS08]. Demzufolge stehen hier vorrangig auf explizitem Wissen basierende Tätigkeiten im Fokus. Aufgaben, deren Ausführung personelle AT und damit implizites Wissen erfordern, werden hingegen zum Aufgabengebiet des WM gezählt. Sein Ziel liegt in der Unterstützung der personellen AT mit Informationen, wozu es sich sowohl mit der geeigneten Gestaltung und Bereitstellung der Wissensbestände im Unternehmen als auch mit individuellen und organisationalen Lernprozessen befassen muss [PRR06]. Damit ist WM gemäß Herbst „(...) ein Führungskonzept, mit dem ein Unternehmen sein Wissen bewusst aktiv und systematisch gestaltet.“ [Her00, S. 9] Wie dieses realisiert werden kann, erläutern verschiedene WM-Modelle¹⁵, von denen sich der sogenannte Wissenskreislauf von Probst et al. als Referenzmodell im deutschsprachigen Raum etabliert hat. Dieser setzt sich aus den sechs Kernprozessen *Wissensidentifikation*, *Wissenserwerb*, *Wissensentwicklung*, *Wissens(ver)teilung*, *Wissensnutzung* und *Wissensbewahrung* zusammen, die mehr oder weniger eng miteinander verbunden sind und über eine Verknüpfung mit den strategischen Bausteinen *Wissensziele* und *Wissensbewertung* dem Regelkreisprinzip des Managements folgen [PRR06].

Obwohl in den obigen Absätzen aufgezeigt wurde, dass sich die beiden Fachgebiete Informations- und Wissensmanagement in bestimmten Aspekten unterscheiden, sollte trotzdem keine strikte Trennung zwischen ihnen vorgenommen werden. Schließlich können letztlich auch WM-Systeme für Aufgaben des IM eingesetzt werden und umgekehrt, so dass WM vielmehr als eine Erweiterung des IM zu sehen ist.

1.2.2. Information Retrieval

Erstmals 1950 von Calvin N. Mooers in seiner Abschlussarbeit über Zatoncoding¹⁶ am MIT¹⁷ in Cambridge erwähnt, wurde der Begriff IR für das Finden von Informatio-

¹⁵Weitere Modelle sind u. a. das SECI (Socialization, Externalization, Combination, Internalization)-Modell von Nonaka und Takeuchi [NT97] oder der geschäftsprozessorientierte Ansatz GPO-WM von Heisig [Hei05].

¹⁶Eine dem Hashcode-Verfahren ähnliche Methode, mit der erstmals eine große Anzahl von Dokumentbeschreibungen auf einer einzigen Kerblockkarte gespeichert werden konnte [Gar97].

¹⁷Massachusetts Institute of Technology (<http://mit.edu>)

nen, deren Existenz oder Fundort a priori unbekannt ist, verwendet [Gar97]. Dabei beschäftigte sich die IR-Forschung in ihren Anfängen hauptsächlich mit der Aufgabe, aus einer bestehenden Kollektion von Textdokumenten diejenigen (wieder) auffindbar zu machen, die für einen Nutzer zur Befriedigung seines Informationsbedürfnisses hilfreich und relevant sind [BR⁺99]. Durch die große Verbreitung des Internets sowie neuer Technologien ist Wissen jedoch nicht mehr nur in Textdokumenten enthalten. Folglich haben sich im Laufe der Zeit aus dem reinen *Text Retrieval* zahlreiche spezielle Forschungsbereiche wie beispielsweise das *Multimedia Retrieval* oder das *Web IR* gebildet [IJ05]. Demzufolge steht der Begriff IR heutzutage eher allgemein für die Ermittlung von benötigten Informationen aus einem Wissensbestand beliebiger Art und Komplexität und umfasst dabei alle Aspekte der dazu notwendigen IS. Letztere dienen also dazu, den „(...) Prozeß des Wissenstransfers vom menschlichen Wissensproduzenten zum Informations-Nachfragenden (...)“ [Fuh96, 3. Absatz] zu unterstützen. Sie werden als IRS bezeichnet und haben, wie im vorherigen Zitat zu erkennen ist, zwei aufeinander abgestimmte Aufgaben zu erfüllen [Sto07].

Zunächst ist im Rahmen der *Indexierung* das vom Wissensproduzenten erzeugte Wissen, welches üblicherweise in Form von Artefakten beschrieben ist, in eine vom IRS verarbeitbare Form — eine Repräsentation — zu bringen. Als Ergebnis entsteht eine Menge von Artefaktrepräsentationen, die das Wissen im System darstellen. Anschließend muss im eigentlichen *Retrieval* die vom Nutzer an das IRS gestellte Anfrage ebenfalls in eine geeignete Anfragerepräsentation überführt und mit den vorhandenen Artefaktrepräsentationen verglichen werden [IJ05, Rij79]. Dies erfolgt mit Hilfe spezieller IR-Modelle (z. B. dem Vektorraummodell [SWY75]), welche die Relevanz der Artefakte in Bezug auf die gestellte Anfrage ermitteln. Obwohl man mit Relevanz diverse Erscheinungsformen wie beispielsweise Pertinenz, Nützlichkeit, System- oder Nutzerrelevanz verbindet [Sar06, Soe94], wird in der vorliegenden Arbeit keine explizite Begriffsdifferenzierung vorgenommen. Vielmehr wird Relevanz gemäß Saracovic als Maß einer auf eine Eigenschaft bezogene Relation zwischen zwei Objekten betrachtet, die von jemandem oder etwas bestimmt wird [Sar06, S. 11]: „*Relevance is the A of a B existing between a C and a D as determined by an E.*“ Allerdings kann dieses Maß entweder genau zwei oder mehrere Ausprägungen besitzen [Tun09]. Im ersten Fall werden nur die beiden Mengen der relevanten und der nicht-relevanten Artefakte differenziert, was bei der Anwendung entsprechender IR-Modelle (z. B. dem Booleschen Retrievalmodell) in einer unsortierten Menge relevanter Ergebnisse resultiert. Aus diesem Grund wird hier auch vom sogenannten *Set Retrieval* gesprochen. Demgegenüber wird der Relevanzbegriff beim *Ranked Retrieval* in das Intervall von 0 bis 1 (bzw. von 0% bis 100%) unterteilt, was im Ergebnis eine Sortierung — ein Ranking — der relevanten Ergebnisse nach ihrem Relevanzwert ermöglicht.

Grundsätzlich gilt dabei aber immer, dass ein IRS in der Lage sein sollte, so viele relevante Informationen wie möglich zu liefern. Dies erhöht nicht nur seine Recall- und Precisionwerte, welche die Fähigkeit des Systems vollständig und genau zu arbei-

ten beschreiben [Sto07], sondern v. a. seine Akzeptanz beim Nutzer, wie es bereits seit 1959 implizit in Mooers' Gesetz¹⁸ postuliert wird. Einen problematischen Aspekt hierbei stellen allerdings die vom Nutzer an das IRS gestellten Anfragen dar. Da diese nur ein Mittel sind, das tatsächliche Informationsbedürfnis zu beschreiben, sind sie immer mit einem gewissen Maß an Vagheit und Unvollständigkeit verbunden [IJ05]. Folglich werden Suchparadigmen benötigt, die dem Nutzer speziell bei der Formulierung seiner Anfrage behilflich sind. Beispiele sind die Ähnlichkeitssuche durch Schlagworte oder die explorative Suche mittels Browsing, auf die im Weiteren noch ausführlicher eingegangen wird.

1.2.3. Rolle des Information Retrieval im Informations- und Wissensmanagement

Wie im vorherigen Abschnitt gesehen, handelt es sich bei einem IRS um ein IS, das zur Suche und zum Auffinden von nützlichen Wissensbestandteilen (Informationen) aus einer existierenden Wissensbasis eingesetzt wird. Da Wissen nur dann zum Unternehmenserfolg beiträgt, wenn es verwendet und zwischen den Mitarbeitern ausgetauscht wird [PRR06], leisten IRS hierzu einen wesentlichen Beitrag. Zum einen bieten sie die Möglichkeit, das Wissen eines Unternehmens zu bündeln, langfristig zu speichern und v. a. zugreifbar zu machen. Demnach werden sowohl die *Wissensbewahrung* als auch die *Wissens(ver)teilung* im Rahmen des zuvor beschriebenen Wissenskreislaufs von Probst et al. unterstützt. Bieten die hierzu entwickelten und eingesetzten Systeme eine einfach zu bedienende GUI, schnelle Antwortzeiten sowie qualitativ hochwertige (d. h. für den Nutzer relevante) Ergebnisse wird ihre Akzeptanz gesteigert, wodurch zusätzlich eine Verbesserung der *Wissensnutzung* erzielt wird. Zum anderen trägt IR auch zur *Wissensidentifikation* bei. So kann ein Mitarbeiter beispielsweise nach Experten oder Mitarbeitern suchen, die bereits Erfahrung in dem von ihm zu bearbeitenden Problembereich haben. Auf diese Weise wird eine höhere Transparenz im Unternehmen erzeugt, so dass etwaige Synergien effektiv zur Problemlösung genutzt werden können. Des Weiteren führt das Auffinden und Verarbeiten neuer Informationen beim Nutzer zur Erweiterung des individuellen Wissens, was in manchen Fällen — z. B. bei der Entstehung neuer Lösungsprinzipien oder neuer Produkte — eine Vermehrung des organisationalen Wissens bewirkt. Somit dienen IRS in gewisser Weise auch der *Wissensentwicklung*. Insgesamt betrachtet sind IRS auf die Unterstützung personeller AT fokussiert und stellen hierbei ein wichtiges Werkzeug zur Realisierung des gesamten WM-Prozesses in einem Unternehmen dar.

¹⁸„An information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not have it.“ (Calvin N. Mooers, 1959); aus [Gar97]

1.3. Forschungsfragen und Zielsetzung der Arbeit

Um eine verbesserte Informationsversorgung für Produktentwickler zu gewährleisten, muss ein IRS möglichst viele der in Kapitel 1.1.2 aufgezeigten Aspekte ihres Informationsbedarfs berücksichtigen. Daraus ergeben sich diverse Fragestellungen, von denen in der vorliegenden Arbeit aufgrund ihrer Fokussierung auf die Indexierungskomponente eines derartigen Systems einige detaillierter und andere weniger ausführlich geklärt werden. Grundsätzlich jedoch geht es um drei Fragestellungen, die zu beleuchten sind.

Als erstes ist zu klären, wonach gesucht werden soll. Folglich steht die Frage nach dem *Was?* im Vordergrund, d. h. :

- Was wird in der technischen PE als Information gesehen bzw. welche Informationen werden von Produktentwicklern benötigt?
- Durch welche Aspekte bzw. Eigenschaften lassen sich diese Informationen charakterisieren?
- Welche dieser Aspekte bzw. Eigenschaften sollen suchbar sein?

Anschließend muss analysiert werden, wo sich die Information im Unternehmen befindet, womit die Frage nach dem *Woher?* einhergeht. Hierfür sind folgende Fragen zu erörtern:

- In welchen Quellen befindet sich welche Information?
- Welche der Quellen müssen bzw. sollen bei der Suche einbezogen werden?
- Kann auf die Informationen aus den Quellen zugegriffen werden bzw. welche Rechte sind dabei zu berücksichtigen?

Letztlich stellt sich noch die Frage nach den Methoden und Konzepten, die für die beiden notwendigen Teilprozesse Indexierung und Retrieval verwendet werden, also *Wie?* das IRS bzw. seine Komponenten zu realisieren sind. Zu klärende Teilaspekte hierbei sind:

- Wie suchen Produktentwickler bisher nach Informationen?
- Welche Suchmöglichkeiten werden tatsächlich benötigt und durch welche Suchparadigmen können diese bereitgestellt werden?
- Wie müssen Informationen im Suchergebnis präsentiert werden, damit sie genutzt bzw. wieder verwendet werden?
- Wie kann die Information aus den identifizierten Quellen gewonnen werden?

- Wie muss die gewonnene Information im System aufbereitet (indexiert) werden, damit sie mit den, im Rahmen des Retrievalprozesses bereitgestellten Möglichkeiten effektiv gesucht werden kann?

Somit ist das primäre Ziel der vorliegenden Arbeit ein Konzept für ein Indexierungsframework zu entwickeln, welches zur Beantwortung der obigen Fragen beiträgt und somit eine effektive Befriedigung von vagen Anfragen im Bereich der technischen PE ermöglicht.

1.4. Aufbau der Arbeit

Zur Erreichung des in Kapitel 1.3 erläuterten Ziels wird mehrstufig vorgegangen, was sich in dem in Abbildung 1.5 gezeigten Aufbau der Arbeit aus insgesamt acht Kapiteln widerspiegelt.

Nachdem in *Kapitel 1* die der Arbeit zugrunde liegende Problemstellung erörtert, eine Einordnung in die zugehörigen Fachgebiete IWM und IR vorgenommen sowie die zu behandelnden Forschungsfragen detailliert wurden, gibt *Kapitel 2* einen Überblick über bereits existierende Lösungen zur Verbesserung der Informationsversorgung von Produktentwicklern. Dazu werden zum einen vorhandene Ordnungsschemata und Klassifikationssysteme näher betrachtet. Zum anderen werden gängige industrielle AWS sowie kommerziell verfügbare CAD-IR-Systeme und existierende ES-Lösungen auf ihre bereitgestellten Suchfunktionalitäten hin untersucht. Nach dieser eher praxisorientierten Sichtweise wird noch ein Überblick über die in der Forschung hierzu entwickelten Konzepte gegeben, so dass letztlich eine Gegenüberstellung aller betrachteten Ansätze hinsichtlich der unterstützten Suchparadigmen vorgenommen werden kann.

Im Anschluss stellt *Kapitel 3* das entwickelte IRS zur Informationssuche in der Produktentwicklung vor. Hierbei handelt es sich um ein interaktives Retrievalmodell, welches komplexe Suchsituationen beherrschbar macht. Dazu werden die Anforderungen an das Retrievalmodell im Allgemeinen und an den Indexierungsprozess im Speziellen definiert. Da ihre Realisierung auf den vier Basiskonzepten *Facettierte Suche*, *Parallele Koordinaten*, *Ranking* und *Ebenenkonzept* basiert, wird auf diese zunächst einzeln eingegangen. Die Gesamtarchitektur des IRS erläutert schließlich die Kombination dieser Basiskonzepte zu einem Gesamtkonzept, wobei die beiden Teilprozesse des Retrievals und der Indexierung mittels der jeweils zugehörigen Komponenten GUI und Indexierungsframework im Detail betrachtet werden. Letzteres bildet den Schwerpunkt der vorliegenden Arbeit und wird aus diesem Grund in den folgenden drei Kapiteln ausführlich behandelt.

Kapitel 4 beschäftigt sich mit den Fragen, welche Informationen für Produktentwickler relevant sind, wie diese charakterisiert und schließlich in ein IRS importiert werden können. Dazu werden zum einen ausgewählte nachgefragte und entwicklungsspezifische Artefakttypen im Hinblick auf ihre jeweiligen Eigenschaften sowie den zwischen ihnen vorherrschenden Beziehungen untersucht. Zum anderen erfolgt für die daraus resultierenden Informationsarten eine Analyse der zu berücksichtigenden Informationsquellen, so dass schließlich ein Lösungsansatz zum Import der Informationen in das IRS vorgestellt werden kann.

Die auf diesem Datenimport basierende Weiterverarbeitung der Informationen ist Gegenstand von *Kapitel 5*, in dem drei Importszenarios differenziert und erläutert werden. Dabei wird der Schwerpunkt des Kapitels auf den Import von Dokumenten und damit auf die Betrachtung ausgewählter Dokumenttypen gelegt. Der Grund hierfür ist in dem besonderen Stellenwert zu sehen, den Dokumente als Informationsquellen in der PE einnehmen. Insbesondere spezielle Dokumenttypen wie beispielsweise CAD-Modelle oder technische Zeichnungen dokumentieren wichtige Informationen im Verlauf des PEP und sind nur in dieser Anwendungsdomäne zu finden. Folglich stehen in diesem Kapitel die Möglichkeiten zur Extraktion von Informationen aus diesen Quellen im Vordergrund. Dazu werden ausgewählte Dokumenttypen näher untersucht, indem zunächst eine Analyse ihres Informationsgehalts erfolgt und anschließend zugehörige Extraktions- und Repräsentationsmöglichkeiten — gegebenenfalls mit zugehörigen Problemen sowie Evaluationsergebnissen — aufgezeigt werden.

Nach der Erfassung der Informationen aus den verschiedenen Quellen sind diese schließlich einem Index hinzuzufügen. Dabei ist zu berücksichtigen, dass es unter Umständen multiple Quellen für dieselbe Information geben kann oder dass Daten unvollständig sind und somit als Nullwerte vorliegen. Beide Situationen können Konsistenzprobleme verursachen, die mit Hilfe von Mechanismen zur Validierung und Verifikation der Informationen zu beheben sind, bevor das tatsächliche Hinzufügen zum Index erfolgt. Demnach beschreibt *Kapitel 6* zum einen wie sich diese Konsistenzprüfung im vorgestellten Indexierungsframework realisieren lässt. Zum anderen wird die Struktur des Index betrachtet, welcher sich aus mehreren Subindizes zusammen setzt. Die Notwendigkeit für die Bildung eines Index aus mehreren Subindizes ergibt sich aus der Tatsache, dass die nachgefragten Informationen von sehr unterschiedlicher Art sind und somit differierende Anforderungen an einen Index stellen.

Kapitel 7 gibt schließlich ein Resümee der vorangegangenen Kapitel, indem die Realisierung des vorgestellten Indexierungsframeworks einer kritischen Reflexion unterzogen wird. Dazu wird zunächst der Prozess zur Umsetzung des Frameworks selbst betrachtet und ein Leitfaden als Orientierungshilfe für ein konkretes Realisierungsvorhaben gegeben. Im Anschluss wird diese prozessorientierte Sicht um eine ergebnisorientierte Sicht ergänzt, in der zusätzlich die Auswirkungen eines realisierten Indexie-

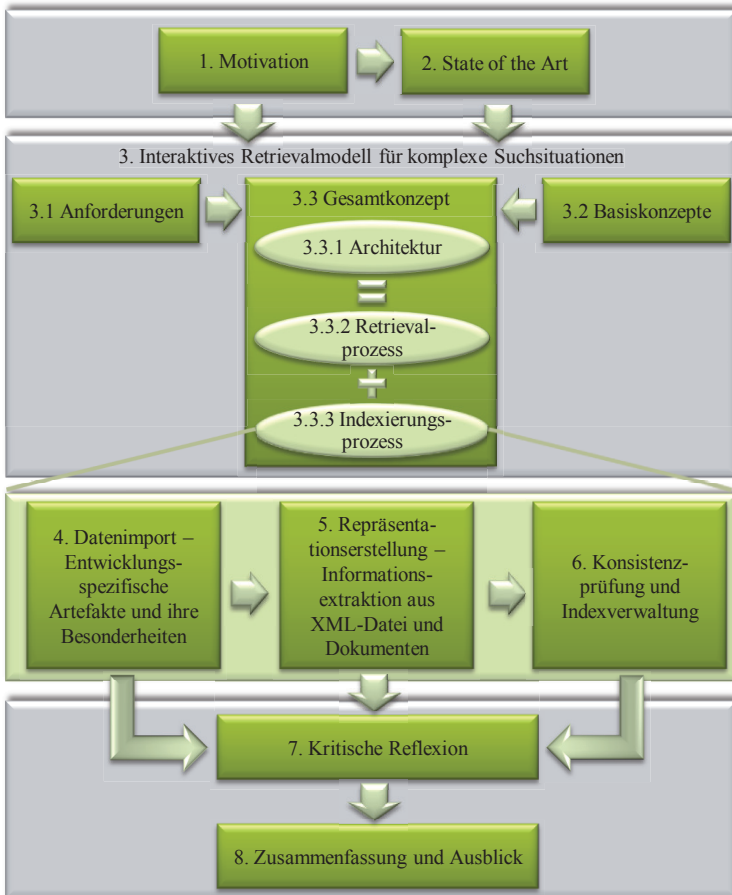


Abbildung 1.5. — Aufbau der Arbeit

rungsframeworks bezüglich seines Nutzens sowie kritischer Erfolgsfaktoren betrachtet werden.

Mit einer Zusammenfassung sowie einem Ausblick auf weiteren Forschungsbedarf wird die Arbeit in *Kapitel 8* abgeschlossen.

2. State of the Art

Die Verwaltung von Daten und Dokumenten und damit verbunden die Gewährleistung der Informationsversorgung von Mitarbeitern im Allgemeinen, aber auch von Produktentwicklern im Speziellen, gehören bereits seit vielen Jahren zu den wesentlichen Herausforderungen von Industrie und Forschung. So entstanden im Laufe der Zeit zahlreiche Ansätze, Methoden und Konzepte, welche mittels unterschiedlicher Suchparadigmen den Zugriff auf bereits existierende Daten bzw. Informationen ermöglichen. Aus diesem Grund beschäftigen sich die folgenden Kapitel mit der Frage, über welche Suchmöglichkeiten Produktentwickler bisher verfügen. Dabei werden in den Kapiteln 2.1 und 2.2 zunächst Ansätze betrachtet, die sich in der industriellen Praxis etabliert haben und sozusagen als Quasi-Standard zu betrachten sind. Hierzu gehören die bereits vor dem Computerzeitalter entwickelten Ordnungsschemata und Klassifikationssysteme sowie die Suchfunktionalitäten gängiger industrieller AwS. Daneben findet man in jüngerer Zeit verstärkt IT-Systeme, die sich entweder auf einen ganz speziellen Aspekt der Suche, nämlich den Vergleich von Produkten, konzentrieren oder aber ein umfassendes und damit unternehmensweites Konzept in den Fokus stellen. Einen Überblick über derartige Systeme in Zusammenhang mit den von ihnen angebotenen Funktionalitäten geben die Kapitel 2.3 und 2.4. Allerdings hat sich bisher keines der entwickelten Konzepte als ganzheitliche und allumfassende Lösung offenbart, weshalb sich selbstverständlich auch die Forschung weiterhin intensiv mit dieser Thematik beschäftigt. In Kapitel 2.5 werden daher verschiedene Ansätze aus der Literatur vorgestellt. Abschließend fasst Kapitel 2.6 diese Betrachtung mit einem Überblick der jeweils unterstützten Suchparadigmen zusammen.

2.1. Bewährte Systematiken als Suchhilfen

Bereits vor der Einführung computerunterstützter Methoden wurden Ordnungsschemata und Klassifikationssysteme entwickelt, die durch eine Systematisierung und geordnete Darstellung von Informationen ein schnelleres und v. a. einfacheres (Wieder-) Auffinden ermöglichen und bis heute in der industriellen Praxis Verwendung finden [PBFG07, Con05]. Prinzipiell stellt dabei jedes Ordnungsschemata in gewisser Weise auch ein Klassifikationssystem dar, da bei diesem eine Klassifikation der jeweils betrachteten Artefakte auf der obersten Gliederungsebene erfolgt. Demzufolge werden

in der vorliegenden Arbeit Ordnungsschemata als Klassifikationssysteme im weiteren Sinne (vgl. Kapitel 2.1.1) und Klassifikationssysteme im engeren Sinne differenziert. Letztere nehmen eine detaillierte Klassifikation der Artefakte auch auf unteren Ebenen vor und werden in Kapitel 2.1.2 erläutert.

2.1.1. Ordnungsschemata

Unter Ordnungsschemata versteht man im Allgemeinen ein- oder mehrdimensionale Matrizen, deren Spalten und Zeilen mit Hilfe von Parametern spezifiziert werden [PBF07]. Ein Beispiel hierfür sind die häufig in der Konzeptionsphase verwendeten Konstruktionskataloge, welche bekannte und bewährte Sachverhalte anhand diverser Gesichtspunkte und Merkmale (z. B. anhand von Energiearten, physikalischen Effekten oder Funktionen) beschreiben. Dabei wird abhängig davon, ob sich der Kataloginhalt auf technische Objekte, Vorgehensweisen, prinzipielle Lösungen zur Funktionserfüllung oder Beziehungen zwischen technischen Objekten bezieht, zwischen Objekt-, Operations-, Lösungs- und Beziehungskatalogen unterschieden [Rot01, Ver77]. Von verschiedenen Experten erstellt, sind sie hauptsächlich in Fachbüchern und Richtlinien und damit in Papierform zu finden¹⁹. Allerdings weisen alle den gleichen, von Roth vorgeschlagenen und in der VDI-Richtlinie 2222 Blatt 2 festgelegten Aufbau aus vier Bestandteilen auf, um einen möglichst schnellen Zugriff auf ein nahezu vollständiges und breit einsetzbares Spektrum an Lösungsalternativen zu gewährleisten [Rot01, Ver77]. Demnach wird in einem sogenannten Gliederungsteil zunächst der systematische Aufbau des Katalogs anhand ordnender Gesichtspunkte festgelegt. Im Anschluss daran folgt der Hauptteil, der die eigentlichen Sachverhalte mit Hilfe von Abbildungen und / oder physikalischen Gleichungen beschreibt. Diese werden im sogenannten Zugriffsteil durch charakteristische, eine Auswahl ermöglichende Eigenschaften, wie beispielsweise Abmessungen, Elementanzahlen oder Kräfteangaben, sowie einen optionalen Anhang ergänzt. Abbildung 2.1 zeigt diesen Aufbau am Beispiel eines Beziehungskataloges, welcher einen Überblick über die fünf Grundbauformen (siehe Parameter *Benennung* im Hauptteil) von Welle-Nabe-Verbindungen gibt²⁰.

Ein weiteres Beispiel für Ordnungsschemata sind die bereits seit mehreren Jahrzehnten vom Deutschen Institut für Normung e. V. (DIN) im Rahmen der Normenreihe DIN-4000 standardisierten Sachmerkmalelisten. Ähnlich zu den Konstruktionskatalogen werden hier charakteristische Merkmale einer Produktgruppe zu einer Sachmerkmalenliste (SML) gruppiert, um ein Auffinden gleicher oder ähnlicher Produkte zu vereinfachen [PBF07, Ver97]. Bei den Merkmalen einer SML handelt es sich einerseits um sogenannte Sachmerkmale, d. h. um spezifische Eigenschaften, welche eine

¹⁹Eine ausführliche Auflistung vorhandener Kataloge findet sich in [Rot01] ab S. 447.

²⁰Hierzu existieren zusätzliche Detailkataloge, die in größerem Umfang die wichtigsten Varianten für diese Grundbauformen aufzeigen (vgl. [Rot01, S. 166ff])

2.1. Bewährte Systematiken als Suchhilfen

Gliederungsteil		Hauptteil			Zugriffsteil							Anhang	
Art des Flächen-schlusses	Art der Kraft-über-tragung	Gleichung	Benennung	Anordnungsbeispiel	Über-trag-bares Moment	Moment-über-tragung abhängig von	Auf-nahme von Axial-kräften	Wirkung bei Über-lastung	Verbin-dung zen-trier-bar	Nabe axial-ver-schieb-bar	Nabe-ver-setz-bar	Anmerkungen	
1	2	1	2	3	Nr	1	2	3	4	5	6	7	8
Normal (Form-schluß)	Un-mittel-bar	$M_t = \frac{d_m}{2} A_{\tau_{\text{ges}}} \tau_{\text{zul}}$ $M_t = \frac{d_m}{2} A_p \tau_{\text{zul}}$	Profil-welle		1	groß	Form-faktor	nein	Bruch	ja	möglich	in Stufen möglich	—
	Mittel-bar	$M_t = \frac{d_m}{2} A_p \tau_{\text{zul}}$ M_t übertragbares Moment mittlerer wirk-samer Durchmesser	Form-element-ver-bindung		2	klein		möglich				möglich	einfache Montage
Tangen-tial (Reib-schluß)	Un-mittel-bar	$M_t = M_r = F_r \cdot d_m / 2$ $= F_N \mu \cdot d_m / 2$ α Keilwinkel	Klemm-sitz		3	klein bis groß	Tem-pe-ratur, Rota-tions-kräften, Axial-kräften	ja	Rut-schen	möglich	ja	—	—
	Mittel-bar	M_t Reibmoment F_r Reibkraft d_{st} Stiftdurch-messer	Spann-element		4	mittel					nur bei $F_A \geq F_r$	stufen-los	Herstell- und Montageauf-wand klein
Tangen-tial und normal	Mittel-bar	—	Vorge-spannte Verbin-dung		5	klein	möglich	Bruch	nein	nein	möglich	—	

Abbildung 2.1. — Übersichtskatalog für Welle-Nabe-Verbindungen (entnommen aus [Rot01, S. 165])

neutrale Beschreibung eines Produktes unabhängig von seiner Herkunft oder Verwendung ermöglichen. Sie werden grundsätzlich in Beschaffenheits- und Verwendbarkeitsmerkmale differenziert [Con05]. Während erstere beschreiben, wie ein Produkt charakterisiert ist, also welche Abmessungen, Form und Farbe es besitzt, geben Verwendbarkeitsmerkmale an, was ein Produkt kann und was es benötigt. Damit werden Eigenschaften wie beispielsweise seine Leistung, sein Platz- oder Energiebedarf spezifiziert. Zum anderen kann eine SML auch sogenannte Relationsmerkmale enthalten, welche eine Beziehung des Produktes zu seinem Umfeld kennzeichnen. Beispiele hierfür sind Merkmale wie Herstellkosten, Bestellmenge oder auch die Einbauhöhe eines Produktes [Con05]. Insgesamt umfasst die Normenreihe laut des Deutsch-Chinesischen Normeninformationsportals²¹ derzeit 120 Produkte, wobei größtenteils Normteile charakterisiert werden [Sta09]. Die Beschreibung erfolgt dabei in tabellarischer Form mit üblicherweise neun Merkmalen²², die durch einen Kennbuchstaben (A bis H und J), eine Merkmalsbenennung sowie eine Einheitenangabe detailliert sind [Con05, ADEK05].

²¹<http://www.standards-portal.de>

²²Die festgesetzte Anzahl von neun Merkmalen basiert auf der Analyse, durch wie viele Merkmale einfache Produkte beschreibbar sind und wie viele davon zur gegenseitigen Abgrenzung auf einen Blick erfassbar sind [Con05].

Durch die klare Gliederung und Darstellung der in Ordnungsschemata enthaltenen Informationen kann ein Produktentwickler einerseits eine bestimmte Lösung zielgerichtet auffinden. Dazu definiert er Werte für einen oder mehrere Parameter, welche in der vorliegenden Arbeit aufgrund ihrer Funktion — der Darstellung von Daten über den gesuchten Sachverhalt [ABB⁺04] — auch als Metadaten verstanden werden, und vergleicht diese mit den entsprechenden Parameterwerten jedes einzelnen im Katalog bzw. in der SML enthaltenen Objektes. Als Ergebnis der Suche erhält er schließlich das Objekt (eventuell auch mehrere), bei dem der Wertevergleich in einer Übereinstimmung resultierte. Demzufolge spricht man bei diesem zielgerichteten Suchvorgehen auch von einer Metadaten- bzw. Parametersuche (MPS) [Tun09]. Andererseits hat der Nutzer v. a. bei Katalogen auch die Möglichkeit, sich zunächst einen Überblick zu verschaffen, was gerade in Situationen, in denen der Nutzer nicht genau weiß wonach er sucht, sehr hilfreich sein kann. Durch Browsing (B), d. h. durch Stöbern in den Informationen entlang eines bestimmten Aspektes, kann der Nutzer Alternativen entdecken, an die er ursprünglich nicht gedacht hat [SGR97]. Er kann aber auch auf Informationen stoßen, die sein Wissen erweitern und somit zu einer Anpassung seines Informationsbedürfnisses und damit einer Präzisierung seiner Suchanfrage beitragen. Da hierbei die tatsächlich benötigte Information meist erst nach mehreren Iterationsschritten — d. h. durch Erforschen des Suchraumes — gefunden wird, spricht man bei dieser Art des Suchens von einem explorativen Suchvorgehen [WR09]. Im Rahmen der Digitalisierung von Katalogen (insbesondere im Bereich elektronischer Zulieferer- bzw. Normteilkataloge) und Sachmerkmalleisten werden hierzu zunehmend Hypermedia-Systeme und DBS eingesetzt [PBFG07, Con05], welche eine Suche in derartigen Ordnungsschemata beispielsweise durch Verknüpfungen (Hyperlinks) zwischen den Inhalten erleichtern.

2.1.2. Klassifikationssysteme

Als Hilfsmittel zur Organisation und Strukturierung von Objekten oder von Wissen über jene werden Klassifikationssysteme bereits seit mehreren tausend Jahren verwendet. Einer der Ersten war Aristoteles, welcher die ihm bekannten Lebewesen überwiegend nach der Methode der Zweiteilung in immer speziellere Gruppen einordnete [Tun09]. Seitdem werden Klassifikationen v. a. aufgrund ihrer Erweiterbarkeit und ihres natürlichen Verständnisses nicht nur im Bibliothekswesen, sondern für diverse Anwendungsbereiche eingesetzt. So auch in der technischen PE, wo insbesondere der Aspekt der Wiederverwendung eine wichtige Rolle hinsichtlich Kosten und Entwicklungszeit spielt. Demnach finden sich hier diverse Klassifikationssysteme, deren Hauptaufgabe darin besteht, durch das Ordnen und Gruppieren von ähnlichen Elementen einer Menge einen schnellen Überblick zu verschaffen und somit das Suchen nach einem bestimmten Element deutlich zu vereinfachen [Man04, PBFG07]. Dies wird grundsätzlich durch die Bildung von Klassen und die Zuordnung der einzelnen

Mengenelemente zu diesen Klassen erreicht. Die Zuordnung erfolgt dabei mit Hilfe eines oder mehrerer klassifikatorischer Merkmale in der Art, dass in Bezug auf dieses Merkmal oder diese Merkmale zwar alle Elemente einer bestimmten Klasse gleich sind, sie sich jedoch von den Elementen anderer Klassen darin unterscheiden. Die meisten Klassifikationen stellen dabei eine hierarchische Untergliederung dar, die auf dem Prinzip der Spezialisierung (d. h. auf 'ist ein'-Beziehungen) basieren. Dabei werden abhängig vom Ordnungsgrad der jeweiligen Hierarchie zwei Arten von Klassifikationen differenziert [Man04, Gau05, Fer03]. Weist die Hierarchie eine starke Ordnung auf, d. h. gehört jedes Element einer Unterklasse zu genau einer Oberklasse (= Monohierarchie), so liegt eine analytische Klassifikation vor, die keine Überlagerungen besitzt. Wie das einfache Beispiel in Abbildung 2.2 unter a) zeigt, findet hier folglich eine schrittweise *top-down* Strukturierung statt, bei der vom Allgemeinen zum Speziellen verfeinert wird [ABB⁺04]. Obwohl diese Strukturierung eine Suche wesentlich vereinfacht, muss in Kauf genommen werden, dass Monohierarchien üblicherweise sehr umfangreich, unflexibel und damit schwierig zu handhaben sind.

Weiterhin ist eine 1:1-Zuordnung von Klassen für viele Anwendungsbereiche nicht immer möglich oder erwünscht. Vielmehr ist es notwendig, ein Element einer Unterklasse mehreren Oberklassen zuordnen zu können und somit Überlagerungen zuzulassen, wodurch eine Polyhierarchie — eine Hierarchie schwacher Ordnung — entsteht. Bei Vorliegen einer derartigen Unterteilung spricht man von einer analytisch-synthetischen Klassifikation, welche auch als Facettenklassifikation bezeichnet wird. Sie stellt, wie am Beispiel in Abbildung 2.2 unter b) gezeigt, für jeden zu betrachtenden Aspekt (= Facette) jeweils eine eigene Klassifikation bereit, so dass ein Element der zu klassifizierenden Menge zunächst in all diese Teilklassifikationen einzuordnen ist (Analyse). Anschließend ergibt sich die Gesamteinordnung des Elementes durch eine definierte Verknüpfung der dabei entstehenden Teilklassifikationszuweisungen (Synthese). Die bekannteste Klassifikation dieser Art ist die von S. R. Ranganathan im Jahr 1933 entwickelte Colon-Klassifikation, bei der Interpunktionszeichen (v. a. der Doppelpunkt) zur Verknüpfung der einzelnen Teilklassifikationen in der Facettenreihenfolge P-M-E-S-T (Personality, Matter, Energy, Space, Time) genutzt werden [ABB⁺04, Tun09]. Durch die Verwendung einer derartigen Polyhierarchie wird, wie Abbildung 2.2 verdeutlicht, nicht nur die Klassifikation an sich übersichtlicher und flexibler. Auf diese Weise können Elemente auch über unterschiedliche Wege gefunden werden.

Insgesamt betrachtet ermöglichen beide Klassifikationsstrukturen ein exploratives Suchen, wobei zwischen einem einfachen Navigieren in einer Monohierarchie ($\hat{=}$ Browsing) und einer Navigation in einer Polyhierarchie differenziert werden kann. Letztere wird auch als Facettierte Navigation (FN) bezeichnet [Tun09]. Zusätzlich wird die aus der Klassifizierung resultierende Einordnung eines Elementes häufig mit Hilfe einer Notation repräsentiert, welche in Form eines numerischen oder alphanumerischen Schlüssels angegeben wird. Dieser kann als Parameter sowohl manuell als auch digital einfach verarbeitet werden, was dazu beiträgt, dass auch ein zielgerichtetes Auffin-

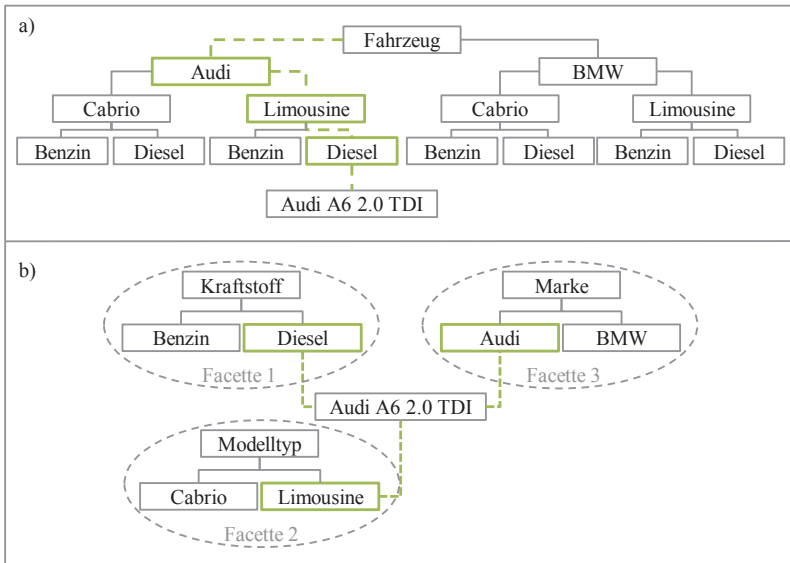


Abbildung 2.2. — Einordnung eines Fahrzeugs in eine a) Monohierarchie und eine b) Polyhierarchie

den von Informationen unterstützt wird [Man04, Wie08]. Im Folgenden werden drei der bekanntesten Klassifikationssysteme aus der technischen PE vorgestellt, welche allerdings maßgeblich der Klasse der Monohierarchien zuzuordnen sind.

Die internationale Patentklassifikation (IPC)

Bei der Entwicklung neuer, innovativer Produkte oder Verfahren ist es unumgänglich, bereits zu Beginn der Planungsphase eine Patentrecherche durchzuführen. Dies dient einerseits dazu, sich über den aktuellen Stand der Technik zu informieren, und andererseits dazu, das Risiko von Doppelentwicklungen und damit verbundenen Rechtsstreitigkeiten zu vermeiden. Da hierbei weltweit existierende Patente zu berücksichtigen sind, deren Anzahl stetig zunimmt, wurde im Jahr 1971 im Straßburger Abkommen die IPC als international einheitliches System zur Strukturierung von Patenten festgelegt. Diese wird von der Weltorganisation für geistiges Eigentum WIPO²³ betreut und ständig aktualisiert. Dabei werden Patente in eine Monohierar-

²³World Intellectual Property Organization (<http://www.wipo.int>)

chie aus acht Sektionen eingeordnet, wobei jede Sektion ein bestimmtes Sachgebiet der Technik repräsentiert. Während beispielsweise Sektion A alles Schutzwürdige aus dem täglichen Lebensbedarf umfasst, enthält Sektion F Patente aus dem Gebiet *Maschinenbau, Beleuchtung, Heizung, Waffen, Sprengen*. Weiterhin sind die Sektionen ihrerseits in Klassen, Unterklassen und Gruppen, welche Haupt- oder Untergruppen sein können, gegliedert. Die Einordnung eines Patentes in die so definierte Klassifikation resultiert schließlich in einer IPC-Notation, bestehend aus Buchstaben und Ziffern, anhand welcher geprüft werden kann, ob bereits Patente für ein bestimmtes technisches System existieren, von wem sie eingereicht wurden und welchen Inhalt diese aufweisen. So werden beispielsweise spezielle Mikroskoptypen mit der Notation G02B21/00 versehen, welche die Hauptgruppe *Mikroskope* (21/00) der Unterklasse *Optische Elemente, Systeme oder Geräte* (B) in der Klasse *Optik* (02) der Sektion *Physik* (G) repräsentiert. [Str08, Man04, Wor10, Deu10]

Das Klassifikationssystem von Opitz

Eine weitere häufig auftretende Situation ist die Suche nach sogenannten Wiederholteilen. Dabei handelt es sich um Produkte, die in verschiedenen Erzeugnissen in der immer gleichen Form benötigt werden und somit einen hohen Wiederverwendungsgrad besitzen [Sch87]. Um das Auffinden derartiger Produkte zu vereinfachen, wurde am Laboratorium für Werkzeugmaschinen und Betriebslehre der RWTH Aachen²⁴ ein Klassifikationssystem in Form sogenannter Formenschlüssel entwickelt [Opi66]. Wie der Name bereits andeutet, werden hier maßgeblich Formmerkmale verwendet, um fünfstellige numerische Schlüssel zu bilden. Die erste Stelle des Schlüssels identifiziert dabei die Klasse, in die ein Produkt aufgrund seiner Grundform einzuordnen ist. Diesbezüglich werden genau zwei Klassen, nämlich die der Rotationsteile (Ziffern 0–5) und die der Nichtrotationsteile (Ziffern 6–9), differenziert. Ihre detailliertere Unterteilung erfolgt anhand verschiedener Abmessungsverhältnisse wie beispielsweise dem Verhältnis von Länge (L) und Durchmesser (D) bei Rotationsteilen. Die übrigen Schlüsselstellen werden anschließend mit einer Zahl zwischen 0 und 9 zur Beschreibung der Außenform, der Innenform und weiterer Flächenmerkmale besetzt, wobei diese abhängig von der ersten Stelle im Formenschlüssel jeweils unterschiedliche Ausprägungen beschreiben. Beispielsweise charakterisiert der Schlüssel 12132 ein Rotationsteil, dessen Verhältnis L/D zwischen 0,5 und 3 liegt (1). Während seine Außenform einseitig steigend oder glatt mit Gewinde (2) ist, wird seine Innenform glatt oder einseitig steigend ohne Formelemente (1) charakterisiert. Außerdem verfügt das Teil über eine Nut und/oder einen Schlitz (3), hat keine Verzahnung und enthält axial angeordnete Hilfsbohrungen mit Teilung (2). Weiterhin besteht die Möglichkeit, den auf diese Weise erzeugten Formenschlüssel um einen sogenannten vierstelligen Ergänzungsschlüssel zu erweitern, der zusätzliche Informationen zu Abmessungen, Werkstoff, Ausgangs-

²⁴Rheinisch-Westfälische Technische Hochschule Aachen (<http://www.rwth-aachen.de>)

form und Genauigkeit enthält. Obwohl Formenschlüssel aufgrund ihrer begrenzten Stellenanzahl nur wenige Merkmale zur Klassifikation verwenden, sind sie aufgrund ihres einfachen Aufbaus und ihrer leichten Handhabung zumindest für bestimmte Anwendungsbereiche wie die Wiederholteilsuche weit verbreitet [Wie08].

Der eCl@ss-Standard

Mit dem Erfolg und der zunehmenden Verbreitung des elektronischen Handels nahm auch die Notwendigkeit von global einheitlichen Beschreibungen der angebotenen Waren zur Verbesserung des Datenaustausches und der Vereinfachung der Geschäftsabwicklung zu. Folglich wurde im Rahmen des vom Bundesministerium für Wirtschaft und Technologie²⁵ geförderten Projektes 'eCl@ss für den Mittelstand' ein internationaler Standard entwickelt, nach dem Produkte, Dienstleistungen und Werkstoffe entlang der gesamten Wertschöpfungskette klassifiziert werden können. Innerhalb von etwa 10 Jahren ist dabei — v. a. begründet durch die Mitarbeit vieler verschiedener internationaler Unternehmen, Verbände und Institutionen im Rahmen der Non-Profit-Organisation eCl@ss e. V.²⁶ — eine branchenübergreifende Klassifikation entstanden, die ständig weiterentwickelt wird, in verschiedenen Sprachen erhältlich ist und somit in der Industrie zunehmende Verbreitung findet. Die Klassifikation selbst erfolgt anhand eines hierarchischen Nummernsystems, das sich aus den vier Ebenen Sachgebiete, Hauptgruppen, Gruppen und Untergruppen zusammensetzt und durch einen achstelligen Warengruppenschlüssel mit jeweils zwei Stellen pro Ebene repräsentiert wird. Außerdem verfügt die Ebene der Untergruppen zum einen über Merkmalleisten, die eine detailliertere Beschreibung und damit Einordnung der zu klassifizierenden Waren auf Basis genormter Merkmale unterstützen. Hierzu greift das eCl@ss-Datenmodell u. a. auf den von der International Organization for Standardization (ISO) spezifizierten Parts Library (PLIB)-Standard (ISO 13584) zurück, welcher die Grundlagen für den digitalen Datenaustausch von Teilekatalogen zwischen Lieferanten und Kunden festlegt. Zum anderen sind die Untergruppen mit Schlagworten versehen, welche ein Auffinden benötigter Waren vereinfachen. Sucht man beispielsweise in dem vom eCl@ss e. V. bereitgestellten Suchsystem²⁷ nach dem Begriff Kreuzschlitzschraube, so gelangt man u. a. zur Warengruppe mit dem Schlüssel 23-11-01-02, welche die Untergruppe der flach aufliegenden Kopfschrauben mit Innenantrieb beschreibt. Besitzen alle auf dieser Untergruppenebene definierten Merkmale eindeutige Ausprägungen für eine bestimmte Ware, so wird diese Ware als eCl@ss-konform bezeichnet.

Mit derzeit 32832 Klassen (davon etwa 27952 Untergruppen) in 26 Sachgebieten im aktuellen²⁸ Release 6.2 deckt der eCl@ss-Standard bereits ein breites Spektrum zur

²⁵<http://www.bmwi.de>

²⁶<http://www.eclass.de>

²⁷<http://www.eclass.de/Home/Suche-in-eCl@ss/3463,de.html?no=intro>

²⁸Stand 17.10.2010

Beschreibung international handelbarer Waren ab. Allerdings ist er ständig bedarfsorientiert weiter auszubauen, um nicht nur den Bedürfnissen der verschiedensten Branchen, sondern v. a. auch denen der verschiedenen unternehmensinternen Bereiche (Entwicklung, Beschaffung, Vertrieb, ...) und damit der gesamten Wertschöpfungskette gerecht zu werden. Ziel ist daher mit dem für Ende 2010 geplanten Release 7.0 ein einheitliches Datenaustauschformat für alle produktdatenfokussierten AwS auf Basis der Extensible Markup Language (XML) zu schaffen, welches zusätzlich sowohl graphische Daten der Mechanik als auch der Elektronik berücksichtigt. Die Orientierung des grundsätzlichen Aufbaus des Standards an acht festgeschriebenen Grundsätzen, wie der Identitätsorientierung, der hersteller- und lieferantenneutralen Strukturierung oder der neutralen Berücksichtigung von Sprachen und landesspezifischen Besonderheiten, ist hierbei als wesentliche Stärke zu sehen, wodurch eine für die Zukunft weltweite Anerkennung und Anwendung dieses Standards begünstigt wird. [eCl07, eCl09, eCl10]

2.2. Industrielle Anwendungssysteme und ihre Suchfunktionalitäten

Die Vielfalt der in der Industrie eingesetzten IT-Systeme zur digitalen Informationsspeicherung und -bereitstellung ist heutzutage kaum vollständig zu erfassen. Die Bandbreite reicht von historisch bedingten Eigenentwicklungen bis hin zu Standardsoftwaresystemen, welche in den meisten Fällen an die jeweiligen Bedürfnisse eines Unternehmens angepasst sind. Aus diesem Grund betrachten die folgenden Kapitel DBS, DMS, ERP-Systeme und PDMS als vier wichtige, nahezu überall vorzufindende Vertreter industrieller AwS und erläutern die von ihnen zur Verfügung gestellten Suchfunktionalitäten.

2.2.1. Datenbanksysteme

Die IT-Landschaft nahezu aller Unternehmen — unabhängig ob es sich um kleine und mittlere Unternehmen (KMU) oder um Großunternehmen handelt — ist heutzutage durch den Einsatz von DBS, welche aus einer Datenbank (DB) und einem zugehörigen Datenbankmanagementsystem (DBMS) bestehen [FS08], geprägt. Dabei versteht man unter einer DB nach Elmasri und Navathe „(...) eine logisch zusammenhängende Sammlung von Daten (...)“, die „(...) für einen bestimmten Zweck entworfen (...)“ sowie „(...) von einer bestimmten Benutzergruppe in zweckbezogenen Anwendungen verwendet (...)“ wird [EN09, S. 18]. Ihre heutzutage fast ausschließlich rechnerbasierte Erzeugung, Pflege und Anwendung erfolgt mit Hilfe eines DBMS, welches die „(...) *Definition, Konstruktion und Manipulation* von Datenbanken (...)“ [EN09, S. 19]

erleichtert und darüber hinaus durch eine Bereitstellung verschiedener Sichten auf die Datensammlung deren gemeinsame Nutzung für unterschiedliche Anwendungen ermöglicht. Die Festlegung, welche Daten, also welche „(...) Fakten aus der »realen« Welt (...)“ [EN09, S. 35], in der DB zu verwalten sind, erfolgt üblicherweise in Form eines DB-Schemas, das mit Hilfe einer Datendefinitionssprache spezifiziert wird²⁹. Dieses Schema beschreibt die Struktur der DB inklusive aller Datentypen, Einschränkungen und Beziehungen für die zu verwaltenden Fakten auf Basis eines speziellen Datenmodells. Abhängig von letzterem werden unterschiedliche Arten von DBS differenziert [EN09, Ern08].

Das in der industriellen Praxis in vielen Bereichen am häufigsten verwendete Datenmodell basiert auf dem Konzept der mathematischen Relation. Hier werden die Daten in Form von Relationen (Tabellen) verwaltet, welche durch einen Relationsnamen sowie eine Liste zugehöriger Attribute definiert sind. Folglich bündelt eine Relation Daten gleicher Struktur, wobei die einzelnen Einträge der Tabelle die einzelnen, durch einen Primärschlüssel eindeutig identifizierbaren Datensätze repräsentieren. Diese Form einer relationalen DB findet man in der technischen PE beispielsweise bei einer Werkstoff-DB, welche u. a. über eine Relation *LEICHTMETALL* mit den Attributen Werkstoffnummer, Zugfestigkeit, Härte und weiteren verfügt. Ihre Datensätze lassen sich anhand der Werkstoffnummer (= Primärschlüssel) unterscheiden und stellen somit die individuellen zu verwaltenden Leichtmetall-Werkstoffe wie Aluminium, Titan oder Magnesium mit ihren jeweils charakteristischen Attributwerten dar. Obwohl mittels sogenannter Fremdschlüssel durchaus komplexere Sachverhalte (insbesondere Beziehungen zwischen Daten) abgebildet werden können, reicht das relationale Datenmodell zur Handhabung komplexer Anwendungen, wie beispielsweise der Speicherung graphischer Informationen in Form technischer Zeichnungen, nicht aus. Speziell für diesen Fall komplexer Objektstrukturen wurde das aus der Softwareentwicklung stammende Konzept der Objektorientierung auf den DB-Bereich übertragen. Die Modellierung von Objekten, welche über einen Zustand in Form von Attributen und ein Verhalten in Form von Methoden verfügen [FS08, FSV01], ermöglicht nicht nur die vereinfachte Integration einer DB in heute überwiegend objektorientierte Softwareprogramme, sondern darüber hinaus eine persistente Speicherung dieser Objekte, so dass sie trotz Programmbeendigung verfügbar sind. Obwohl dadurch eine erhöhte Flexibilität erreicht wird, konnten sich objektorientierte DB v. a. in klassischen, weniger komplexen Anwendungsbereichen bis jetzt nicht durchsetzen. Vielmehr versucht man durch eine Erweiterung der relationalen Systeme um objektorientierte Konzepte in Form objektrelationaler DBS oder durch Verwendung sogenannter

²⁹Eine Ausnahme diesbezüglich sind die in jüngster Zeit verfügbaren NoSQL-Datenbanken, welche explizit ohne eine Schemadefinition Dokumente in Form von Schlüssel-Wert-Paaren speichern. Sie wurden speziell für Web 2.0-Anwendungen entwickelt, weshalb sie gegenwärtig zwar noch keine weite Verbreitung im industriellen Umfeld erlangt haben, zukünftig aber sicherlich — gerade im Zuge der Enterprise 2.0-Einführung — eine Rolle spielen werden. Für weitere Informationen wird daher auf [Jan10] verwiesen.

O/R-Mapper, wie beispielsweise Hibernate³⁰, die sozusagen zwischen dem objektorientierten AwS und der relationalen DB vermitteln, die Vorteile beider Sichtweisen zu vereinen [EN09].

The screenshot shows the 'werkstoffe.de' website, which is the 'Bayerisches Online-Materialinformationssystem'. The main navigation bar includes 'Werkstoffsuche', 'Kompetenzsuche', 'werkstoffe.de', 'Nutzeranleitung', 'Forum', 'Werkstoff aktuell', 'Kontakt', and 'Sitemap'. The 'Werkstoffsuche' section is active, displaying 'Kennwertsuche nach Werkstoffen'. A sidebar on the left lists various search categories: 'Werkstoffdatenbank', 'Werkstoffbaum', 'Anwendungsgrenzen', 'Profilsuche', 'Kennwertsuche' (selected), 'Bauteilsuche', 'Normen, Bezeichnungen', 'Vollstetsuche', 'Thermisches Fügen', 'Werkstoffbeschreibung', 'Ihre Werkstoffdaten', 'Schadenanalyse', and 'Registrieren'. The main search area is titled 'Dieses Menü bietet die Möglichkeit der werkstoffübergreifenden Materialsuche.' and contains four search criteria, each with 'von' and 'bis' input fields and a unit: 'E-Modul [MPa]', 'Zugfestigkeit (Rm) [MPa]', 'Streckgrenze (Re) [0,2% Dehngrenze (Rp0,2)] [MPa]', and 'Dichte [kg/dm^3]'. A 'Suchen' button is located at the bottom of the search area. On the right, there are links for 'Anmelden' and 'Registrieren' (highlighted with a red box), and a link for 'Auskünfte über Werkstoffe und Verfahren'.

Abbildung 2.3. — Screenshot der Kennwertsuche in der Werkstoffdatenbank M-Line 2 [SP07]

Obwohl gerade im Bereich der objektrelationalen und -orientierten DBS die Realisierung diverser Suchansätze möglich ist, wird die folgende Betrachtung aufgrund der hohen Bedeutung relationaler DBS im Umfeld der technischen PE auf diese beschränkt. Hier stellen DBMS unterschiedliche Alternativen für eine Suche in einer DB bereit. Zum einen existieren Oberflächen, die den Nutzern Menüs oder Formulare zur Formulierung einer Anfrage anbieten [EN09]. Während im ersten Fall aus Listen von Optionen schrittweise einzelne Optionen ausgewählt werden können, füllt der Nutzer im zweiten Fall Felder eines vordefinierten Formulars aus. Letztere können unterschiedlich gestaltet werden, was wiederum verschiedene Suchmöglichkeiten unterstützt. So findet man häufig Eingabemasken, die Eingabefelder für die in der DB enthaltenen Parameter zur Verfügung stellen und deren Eingabewerte schließlich als Parameter bei der Suche übernommen werden (= MPS). Abbildung 2.3 zeigt beispielhaft eine Suchmaske der im Internet verfügbaren Werkstoffdatenbank des Bayerischen Online-Materialinformationssystems M-Line 2 der Technischen Universität München [SP07], welche Werkstoffe über Intervallangaben für die vier Parameter E-Modul, Zugfestigkeit, Streckgrenze und Dichte suchbar macht. Neben derartigen Eingabemasken können Formulare aber beispielsweise auch in Form von Tabellen gestaltet werden,

³⁰<http://www.hibernate.org>

wie es bei der von IBM³¹ entwickelten graphischen Anfragesprache Query-by-example erfolgt [Zlo76, RG02]. Hier kann der Nutzer in eine vordefinierte Tabelle Werte entweder als Konstanten oder als Beispielwerte (markiert durch einen vorangestellten Unterstrich) eintragen. Durch Vergleichen dieser Tabelle mit den DB-Tabellen erhält man schließlich im Suchergebnis alle Datensätze, welche die definierten Werte bzw. zu den Beispielwerten ähnliche Werte besitzen. Somit werden hier Objekte auf ihre Ähnlichkeit zueinander untersucht, was allgemein durch den Begriff der Ähnlichkeitssuche verdeutlicht wird. Hierfür werden Vergleichsobjekte benötigt, welche je nach Anwendungsfall unterschiedlich ausgeprägt sein können. Während im vorliegenden Fall eine Tabelle verwendet wird, können in anderen Fällen auch andere Formen (z. B. Skizzen, Dokumente) als Beispielobjekte dienen. Aus diesem Grund wird eine Ähnlichkeitssuche, deren Anfrage aus einem Beispielobjekt besteht, allgemein mit dem Begriff Query-by-Example (QbE) bezeichnet [BR⁺99].

Weiterhin gibt es Oberflächen, bei denen Anfragen entweder in einer speziellen Syntax oder auch in natürlicher Sprache gestellt werden können [EN09]. Allen gemeinsam ist, dass die Anfrage – entweder direkt vom Nutzer oder intern durch das System – in eine Datenmanipulationssprache transformiert werden muss. Dies gilt für eine objektorientierte ebenso wie für eine relationale DB, für welche die standardisierte Sprache SQL (Structured Query Language), die ebenfalls die zur Definition und Konstruktion einer DB benötigte Datendefinitionssprache in sich vereint [FS08], ein bekanntes Beispiel ist. Diese stellt zum Suchen die Anweisung

SELECT <Attributliste> **FROM** <Tabellenliste> **WHERE** <Bedingung>;

zur Verfügung, welche einen Zugriff auf die in einer DB stark strukturierten Attribute bzw. deren Werte aus den definierten Relationen unter Berücksichtigung der spezifizierten Bedingungen ermöglicht [Fer03]. Somit kann ein Nutzer in Situationen, in denen er weiß wonach er sucht, diese genau identifizierbaren Daten bzw. Fakten abfragen [SP05]. Dabei werden im Hinblick auf die Anfrage exakte Ergebnisse geliefert, weshalb dieses Suchparadigma auch als Faktensuche (FS) bezeichnet wird [Rij79, Fer03]. So kann z. B. in der oben genannten Werkstoff-DB nach dem Zugfestigkeitswert von purem Aluminium gesucht werden, was als Ergebnis den Wert 49 N/mm^2 [Wik10b] liefern würde. Natürlich lassen sich aber auch Bereichsanfragen bilden oder durch Verschachtelung mehrerer SELECT-FROM-WHERE-Anweisungen komplexere Anfragen formulieren, welche entweder in keinem oder genau einem Faktum oder einer Menge von Fakten resultieren [EN09, FSV01].

³¹<http://www.ibm.com>

2.2.2. Dokumentenmanagementsysteme

Eines der wichtigsten Elemente zur Abwicklung von Geschäftsprozessen war und ist das Dokument, welches in seinen diversen Ausprägungen nicht nur zur Regelung und Steuerung der internen Abläufe dient, sondern v. a. als Basis und Nachweis für die Vertragsvereinbarungen mit Lieferanten und Kunden fungiert. Der einzige Unterschied zu früher liegt allein darin, dass Dokumente heutzutage in weitaus geringerem Umfang in Papierform vorliegen. Doch auch in der Vergangenheit stand man wie heute v. a. aufgrund gesetzlicher Aufbewahrungspflichten vor der Frage, wie man immer größer werdende Dokumentarchive nicht nur im Hinblick auf den von ihnen benötigten Platzbedarf sondern auch in Bezug auf die Auffindbarkeit der in den Dokumenten enthaltenen Informationen zukünftig handhaben sollte. Insbesondere die fortschreitenden Erkenntnisse im Bereich der Speichertechnologien ermöglichten bereits Anfang der 1960er Jahre die Entwicklung von Systemen, die sich auf die Verwaltung der in Papierform vorliegenden Dokumente fokussierten [Sch04]. Dabei wurden die Papierdokumente mit Hilfe von Scannern digitalisiert, wodurch sie gespeichert und über den Bildschirm oder Drucker auch wieder ausgegeben werden konnten. Dies trug nicht nur zu einem geringeren Raumbedarf und damit verbunden einer Reduzierung der Kosten bei, sondern ermöglichte v. a. eine permanente und schnellere Verfügbarkeit der Dokumente für mehrere Nutzer unabhängig vom geographischen Ort ihrer Archivierung [Ber94, Sch04]. Allerdings wurden die dabei entstandenen Systeme im Laufe der Zeit immer weiter entwickelt, so dass man zwischenzeitlich von reinen Archivsystemen über Recherchesysteme zur Informationsrückgewinnung bis hin zu Vorgangsbearbeitungssystemen unterschiedliche Varianten am Markt findet [Ber94, Dan99]. Letztere enthalten neben Archivierung und Recherche zusätzliche Funktionalitäten wie beispielsweise die elektronische Verteilung bzw. Weiterleitung von Dokumenten zur Unterstützung von Freigabeprozessen, wodurch sie eine schnellere und v. a. fehlerfreie Abarbeitung betrieblicher Abläufe ermöglichen. Sie sind heutzutage unter dem Schlagwort der WfMS in diversen Anwendungsbereichen zu finden.

Eine Einordnung der am Markt verfügbaren DMS in diese drei Gruppen ist abhängig davon, welche der genannten Funktionalitäten in welchem Umfang unterstützt werden, nicht immer eindeutig möglich. Aus diesem Grund wird in der vorliegenden Arbeit ein DMS gemäß Dandl als „(...) ein hard- und softwaretechnisches Werkzeug des Informationsmanagements (...)“ [Dan99, S. 6] verstanden, welches den gesamten Lebenszyklus eines Dokumentes, d. h. „(...) das Erstellen, Digitalisieren, Indizieren, Archivieren, Wiederfinden, Weiterleiten und Vernichten (...)“ [Sch04, S. 339] unterstützt und dabei insbesondere die Verwaltung verschiedener Dokumentversionen sowie den kontrollierten Zugriff auf die Dokumente mit Hilfe von Berechtigungen ermöglicht [Fin05]. Hierzu sind die Dokumente unabhängig von ihrem Inhalt (d. h. Textdokumente, ebenso wie Bilder, Zeichnungen usw.) zunächst entweder digital aus anderen AwS zu importieren oder mit Hilfe von Scannern zu digitalisieren und in

Recherche formulieren

Veröffentlichungsnummer:

Titel:

Anmelder:

Erfinder:

Veröffentlichungsdatum:

Bibliographische IPC:

Reklassifizierte IPC:

Anmeldedatum:

Prüfstoff-IPC:

Suche im Volltext:

Trefferliste konfigurieren

☒ Veröffentlichungsnummer ☒ Titel ☐ Erfinder
☐ Anmelder ☐ Veröffentlichungsdatum ☐ Anmeldedatum
☐ Prüfstoff-IPC ☒ IPC-Hauptklasse

Trefferlisten sortierung nach:

Treffer/Seite

Abbildung 2.4. — Screenshot des in DEPATISnet verfügbaren Suchmodus für Einsteiger

das DMS einzupflegen. Dabei werden sie mit diversen Metadaten wie beispielsweise dem Dokumenttyp, dem Autor oder ihrem Erstellungsdatum näher beschrieben [Ber94, Dan99] und schließlich in einem üblicherweise aus einem optischen Speichermedium bestehenden Archivsystem abgelegt [Sch04]. Die Metadaten werden dabei zusammen mit einem Verweis auf diesen Speicherort in einer relationalen DB verwaltet, was eine MPS mittels Suchmasken im DMS erlaubt. Ferner werden neben Metadaten auch die Inhalte von Textdokumenten — gegebenenfalls über Optical Character Recognition (OCR)-Verfahren — erschlossen und indiziert [Sch04]. Somit hat der Nutzer zusätzlich die Möglichkeit, eine Volltextsuche in diesen Inhalten durchzuführen. Grundsätzlich versteht man unter einer Volltextsuche eine Variante der Ähnlichkeitssuche, bei der mittels einer Anfrage in Form eines Beispieltextes die Ähnlichkeit von Texten bzw. Textfragmenten ermittelt wird. Solche Query-by-Text (QbT)-Anfragen können zwar aus einem Langtext bestehen, werden aber in den meisten Fällen nur aus einzelnen Schlagwörtern gebildet. Aus diesem Grund wird hierfür häufig auch die Bezeichnung einer Schlagwortsuche verwendet. So findet sich beispielsweise in dem im WWW verfügbaren elektronischen Archiv- und Recherchesystem DEPATISnet³² des Deutschen Patent- und Markenamtes ein Suchmodus für Einsteiger (vgl. Abbildung 2.4), welcher neben der Definition der bei der Suche zu verwendenden Metadaten (z. B. Veröffentlichungsnummer, Titel, Erfinder oder IPC-Notation) im letzten Eingabefeld auch eine Suche in den Inhalten der Patentdokumente ermöglicht. Eine dritte Variante des Suchens in DMS ergibt sich schließlich durch

³²<http://depatisnet.dpma.de>

die Ablage der Dokumente im Archivsystem. Sie erfolgt unter Nutzung hierarchisch verknüpfter Verzeichnisse, welche in Anlehnung an die manuelle Papierablage typischerweise als Schränke, Schubladen, Ordner und Mappen bezeichnet werden [Sch04] und somit ein Browsing durch diese Struktur ermöglicht.

2.2.3. Enterprise Resource Planning-Systeme

Die Planung von Ressourcen ist seit jeher eine wichtige Aufgabe zur Herstellung von Produkten, welche durch die Einführung von IKT wesentlich beeinflusst wurde. Angefangen mit der computerunterstützten Planung des Materialbedarfs, also der für ein Produkt benötigten Einzelkomponenten und Rohstoffe durch Auflösung von Stücklisten (Material Requirements Planning, MRP I), wurden bereits wenige Jahre später alle für die Fertigung benötigten Hilfsmittel, d. h. auch personelle und finanzielle Ressourcen, in die Produktionsplanung und -steuerung einbezogen (Manufacturing Resource Planning, MRP II). Allerdings wurde schnell deutlich, dass die Planung von Ressourcen allein für den Fertigungsbereich zur Sicherung eines langfristigen Unternehmenserfolges nicht ausreicht. Vielmehr ist es notwendig alle Ressourcen eines Unternehmens für dessen gesamten betrieblichen Ablauf möglichst effizient einzusetzen. Demzufolge fand eine Erweiterung der existierenden IS statt, die eine übergreifende Verwendung der Unternehmensressourcen zur Abwicklung der Geschäftsprozesse in diversen Unternehmensbereichen (Beschaffung, Vertrieb, Personal usw.) unterstützt. Wesentliche Merkmale der dabei entstandenen ERP-Systeme sind die Zusammenfassung diverser Insellösungen in einem AwS unter Verwendung einer gemeinsamen, konsistenten Datenbasis sowie eine transaktionsorientierte Unterstützung der betrieblichen Abläufe [FGS08].

Die Integration der einzelnen Systeme kann dabei sowohl horizontal, d. h. über mehrere Unternehmensbereiche hinweg, als auch vertikal erfolgen, so dass neben Planungs- und Kontrollaufgaben auch operative Tätigkeiten (Administration und Disposition) berücksichtigt werden. Je nach Art und Umfang der Integration weisen ERP-Systeme folglich unterschiedliche Funktionalitäten auf, wobei im Wesentlichen drei primäre Funktionsbereiche für ihren Einsatz identifiziert werden können [Gro04, Con08d]. Neben dem Finanz- und Personalwesen existieren dabei speziell auf Produktion und Logistik ausgerichtete Systeme, welche im Rahmen der vorliegenden Arbeit von besonderem Interesse sind. Sie beschäftigen sich überwiegend mit der Beschaffung und Auftragsabwicklung, mit Mengen-, Bestands- und Kapazitätsplanungen im Rahmen der Fertigungsvorbereitung sowie der Stücklisten- und Stammdatenverwaltung [Wie08, Ver02] und stellen gerade deshalb eine wichtige Informationsquelle für Produktentwickler dar. Da die Anzahl der am Markt existierenden Anbieter für ERP-Systeme

jedoch sehr groß ist³³, wird für die weitere Betrachtung stellvertretend eines der führenden Systeme, das von der SAP AG³⁴ entwickelte System SAP ERP (vorher SAP R/3) mit seinem Modul SAP ERP Operations als Beispiel herangezogen. Die Verwaltung von Ressourcen erfolgt hier mit Hilfe von Stamm- und Bewegungsdaten. Dabei werden unter Stammdaten alle Daten zusammengefasst, die langfristig definiert sind und von den diversen Unternehmensbereichen benötigt werden [SAP09]. Hierzu gehört beispielsweise der sogenannte Materialstamm, welcher manuell definierbare Informationen zu beschreibenden Eigenschaften (z. B. Größe, Abmessungen, Gewicht oder Materialart) sowie automatisch vom System erzeugte Daten (z. B. Bestände) zu einem bestimmten Material³⁵ enthält [SAP09]. Aber auch grundlegende Lieferanten- und Kundeninformationen wie Anschrift, Bankverbindung, Zahlungskonditionen oder Lieferzeiten werden in zugehörigen Kreditoren- und Debitorenstammsätzen gespeichert. Bewegungsdaten hingegen sind Daten, die sich auf konkrete betriebliche Abläufe beziehen und damit nur eine kurzfristige Gültigkeit besitzen [SAP09]. Hierbei handelt es sich beispielsweise um Bestellungen, Buchungsbelege oder auch Fertigungsaufträge, die schließlich den entsprechenden Stammdaten zugeordnet werden. Beide Datengruppen enthalten folglich wichtige Informationen, wie beispielsweise Materialnummern, Einkaufspreise, Lieferzeiten oder Bestellmengen, die Produktentwickler zur Erfüllung ihrer Aufgaben benötigen (vgl. [Sch08] für einen ausführlichen Überblick). Die Suche nach diesen Daten im ERP-System erfolgt anhand des zugrunde liegenden Datenbestandes, welcher üblicherweise durch eine relationale DB realisiert ist. Demzufolge können ERP-Systeme prinzipiell alle vom jeweils zugehörigen DBS unterstützten Suchfunktionalitäten anbieten, wobei der Fokus jedoch auf die Metadaten- bzw. Parameter- und auf die Faktensuche gelegt wird. So bietet beispielsweise SAP ERP eine sogenannte Matchcodesuche an, bei der mit Hilfe von Eingabemasken der Datenbestand nach bestimmten Eingabewerten (Matchcodes) für bestimmte Felder durchsucht werden kann [FGS08]. Darüber hinaus wird häufig auch die Verwendung eines DMS ermöglicht [Fin05, SAP10], um für Geschäftsprozesse notwendige Zusatzinformationen in Form von Dokumenten bei den Stammdaten hinterlegen zu können. Das wiederum ermöglicht eine Einbindung der von diesen zur Verfügung gestellten Suchfunktionalitäten, wie sie in Kapitel 2.2.2 erläutert wurden.

³³Allein im deutschsprachigen Raum existierten laut Gronau im Jahr 2008 über 600 Anbieter von ERP-Systemen speziell für KMU verschiedener Branchen [Gro08].

³⁴<http://www.sap.com>

³⁵SAP definiert ein Material allgemein als ein „Gut, das Gegenstand der Geschäftstätigkeit ist.“ [SAP09, Eintrag *Material (LO)*] und somit sowohl ein Produkt als auch einen Werkstoff beschreiben kann.

2.2.4. Produktdatenmanagementsysteme

Obwohl die bisher betrachteten Systeme bereits viele Aktivitäten der Wertkette eines Unternehmens unterstützen, sind ihre Funktionalitäten jedoch für den Bereich der technischen PE nicht ausreichend. Hier steht das zu entwickelnde Produkt mit all seinen zugehörigen Informationen im Mittelpunkt, welche aufgrund des zunehmenden Einsatzes von CAX- und VR-Technologien einen immer höheren Grad an Digitalisierung aufweisen. Aus diesem Grund hat man bereits Ende der 1980er Jahre unter der Bezeichnung EDM / PDM ein neues Konzept entwickelt, welches im Laufe der Jahre als Produktdatenmanagement (PDM) in immer mehr Unternehmen Einzug fand [Ver02]. Dabei liegt das Ziel dieses Konzeptes im Einsatz eines DB- und Kommunikationssystems, mit dem sämtliche Produktdaten über den gesamten PEP hinweg effizient verwaltet und bereitgestellt werden [Les01]. Hierzu sind einerseits „(...) die das Produkt beschreibenden Strukturen, Dokumente und Daten (...)“ [Ver02, S. 4] selbst (→ PDM), aber auch der „(...) Arbeitsprozess in der Entwicklung und die ihn beschreibenden Daten (...)“ [Ver02, S. 7] (→ Engineering Data Management (EDM)) einzubeziehen. Folglich integrieren PDMS alle im Laufe der PE in den verschiedenen AwS erzeugten Informationen und stellen dieses Wissen durch definierte Workflowprozesse nicht nur Mitarbeitern innerhalb der Entwicklungsabteilung sondern auch anderen Unternehmensbereichen zur Verfügung [Sch02a, SW08]. Um dies zu gewährleisten, hat ein PDMS zahlreiche sowohl anwendungsbezogene als auch übergreifende Aufgaben zu erfüllen. Beispiele sind das Produktdaten- und Dokumentenmanagement, das Konfigurationsmanagement, die Versionierung, die Teileklassifizierung, das Workflowmanagement, die Kommunikation oder das Benutzermanagement — um nur einige zu nennen. Ein ausführlicher Überblick der unterstützten Funktionalitäten wird beispielsweise von Sandler und Waver in [SW08] gegeben, die diese detailliert am Beispiel des PDMS PRO.FILE der PROCAD GmbH & Co. KG³⁶ erläutern.

Grundlage all dieser Funktionalitäten ist ein virtuelles Produktdatenmodell, mit dessen Hilfe sämtliche Informationen objektorientiert abgebildet werden. Dazu wird grundsätzlich zwischen Geschäfts-, Daten- und Relationsobjekten unterschieden [Sch99]. Erstere repräsentieren die im PDMS zu verwaltenden Daten wie Produkte (Eigen- und Fremdprodukte), Dokumente, Kunden oder Projekte. Diese werden mit Hilfe von Metadaten, wie beispielsweise der Produktnummer, einer Produkttypbeschreibung und spezifischen Eigenschaften wie Höhe oder Durchmesser, näher beschrieben und in ein Klassifikationssystem eingeordnet. Einem solchen Geschäftsobjekt können zusätzlich andere Geschäftsobjekte (z. B. Produkte als Unterkomponenten oder beschreibende Dokumente) zugeordnet werden, was mit Hilfe von Relationsobjekten realisiert wird. Handelt es sich bei den zu verknüpfenden Informationen um Dokumente, so werden diese — sofern es sich um digitale oder digitalisierte (gescannte)

³⁶<http://www.procad.de>

Dokumente handelt — durch Datenobjekte realisiert³⁷. Diese enthalten neben der Datei des Dokumentes ebenfalls beschreibende Metadaten. Während letztere gemeinsam mit den Geschäftsobjekten bzw. deren Metadaten in einer DB verwaltet werden [Sch99, SW08], findet die Archivierung der Dokumentdateien ähnlich zu DMS in separaten Dateiverzeichnissen, den sogenannten Vaults, statt [Ver02, ES01]. Damit können PDMS drei grundlegende Recherchemöglichkeiten zur Verfügung stellen [Sch99]. Neben einem Browsing in der Klassifikationsstruktur, steht hierbei die MPS auf Basis der in der DB enthaltenen Informationen im Fokus. Allerdings wird durch das zusätzliche Dokumentenmanagement auch eine Suche in den Dokumentinhalten möglich, wobei derzeit allerdings nur die Textinhalte der Dokumente im Rahmen einer Volltextsuche berücksichtigt werden.

Insgesamt betrachtet sind PDMS somit — ebenso wie die in Kapitel 2.2.3 beschriebenen ERP-Systeme — IT-Plattformen für die Zusammenfassung einzelner Insellösungen in einem speziellen Anwendungsbereich, um die in diesem Bereich stattfindende Datenverarbeitung und -verwaltung effizienter zu gestalten. Damit dabei jedoch nicht nur einzelne Bereiche des Produktlebenszyklus abgedeckt werden, sondern im Rahmen eines effektiven Produktlebenszyklusmanagements Informationen aus allen Lebensphasen eines Produktes, d. h. von der Produktidee bis zum Recycling, mit einbezogen werden, stellen sie zusammen mit ERP- und anderen Systemen wesentliche Bestandteile zur Umsetzung eines solchen Konzeptes dar [SW08].

2.3. Kommerzielle CAD-IR-Systeme

Wie in Kapitel 1.1.2 erläutert, sind Produktentwickler überwiegend mit Anpassungs- oder Variantenkonstruktionen konfrontiert, was ein schnelles Auffinden von verwendbaren Produkten mit zugehörigen CAD-Dokumenten erfordert. Aus diesem Grund wurden speziell auf diesen Informationsbedarf fokussierte CAD-IR-Systeme entwickelt, die als kommerzielle Lösungen am Markt angeboten werden. Diese unterstützen nahezu durchgängig eine Suche anhand produktbeschreibender Daten, indem sie u. a. die Metadaten industrieller AwS über Schnittstellen integrieren. Allerdings ist diese rein textuelle Beschreibung nicht immer ausreichend, um ein passendes Produkt zu finden. Vielmehr spielt die geometrische Beschreibung eines Produktes in Form seiner Gestalt eine entscheidende Rolle, weshalb zusätzlich geometriebasierte Ansätze zur Suche nach ähnlichen Produkten benötigt werden. Verstärkt wird diese Forderung auch durch den bereits in Kapitel 1 erwähnten Übergang von der 2D- auf die 3D-Technologie beim Konstruieren, wodurch immer weiter wachsende Dokumentbestände an 3D-CAD-Modellen verfügbar sind. Letztere beschreiben die Geometrie

³⁷Eine Zuordnung von nur in Papierform vorliegenden Dokumenten erfolgt lediglich durch ein Geschäftsobjekt mit zugehörigem Stammdatensatz, welcher den manuellen Ablageort des Dokumentes enthält [ES01].

verdeutlichen, gibt Tabelle 2.1 einen beispielhaften Überblick der derzeit bekanntesten Systeme. Die von ihnen unterstützten Funktionalitäten wurden dabei hinsichtlich folgender Fragestellungen untersucht:

- Welche Artefakttypen (Produkte, Dokumente) können gefunden werden?
- Welche Art der Anfrageformulierung (Metadaten, 3D, 2D) wird unterstützt?
- Welche Ähnlichkeitskriterien (Gestalt, Text, weitere) werden verwendet?
- Erfolgt eine Ergebnisvorschau in Form von Thumbnails?
- Wird ein Ranking der Ergebnisse ermittelt?

Wie bei diesem Vergleich zu erkennen ist, bietet das System GEOMETRICAL SEARCH Desktop von CADENAS als einzige Lösung die meisten Funktionalitäten, bleibt aber rein auf Produkte bzw. CAD-Dokumente beschränkt. Nur das Unternehmen CAD-Partner GmbH ermöglicht durch eine integrierte Lösung der beiden Anwendungen Smap3D PartFinder und Smap3D SmartSearch neben der Produktsuche auch eine Suche nach anderen Dokumenttypen wie beispielsweise Auftragsbestätigungen oder Rechnungen. Allerdings muss im Gegenzug z. B. auf die Möglichkeit einer 2D-Anfrageformulierung verzichtet werden. Insgesamt betrachtet stellen sämtliche am Markt verfügbaren CAD-IR-Systeme Insellösungen dar, die nur bestimmte Aspekte des Informationsbedarfs eines Produktentwicklers befriedigen.

2.4. Enterprise Search-Systeme

Obwohl Produktentwickler, wie die vorangegangenen Kapitel gezeigt haben, über diverse Suchmöglichkeiten zum Auffinden einer benötigten Information verfügen, sind diese allerdings als reine Insellösungen zu betrachten. Der Grund hierfür liegt in ihrer jeweiligen Fokussierung auf bestimmte Informationsbedürfnisse (z. B. Suche nach Produktinformationen), wodurch lediglich Suchsituationen unterstützt werden können, in denen ein Nutzer bereits über eine gewisse — wenn auch nicht vollständig konkrete — Vorstellung seines Suchergebnisses verfügt. Fehlt eine derartige Zielvorstellung oder ist nicht bekannt wie bzw. wo gesucht werden kann, wird eine umfassende Suche sowohl in allen unternehmensintern verfügbaren als auch in extern zugreifbaren Informationsbeständen (z. B. WWW, Lieferantendatenbanken, ...) notwendig. Um hierfür jedoch nicht jede einzelne Insellösung separat anwenden zu müssen, wurden spezielle IRS — sogenannte ESS — entwickelt, die einen gebündelten Zugang für eine übergreifende Suche darstellen (*single point of access*) und sich dabei speziell den im unternehmerischen Umfeld bestehenden Herausforderungen widmen [Haw04, MM04]. Hierzu gehört primär die Einbeziehung der bereits in Kapitel 1.1.1 angesprochenen zahlreichen, heterogenen IT-Systeme, in denen drei unterschiedliche Arten von Informationen er-

Tabelle 2.1. — Übersicht und funktionelle Charakterisierung der kommerziell verfügbaren CAD-IR-Systeme

CAD-IR-System (Anbieter-Internetadresse)	Suchbare Artefakte		Anfrage		Ähnlichkeitskriterien			Vor- schau	Ran- king
	Produkte	Dokumente	Metadaten	3D	2D	Gestalt	Text	sonstige	
simusclassmate FINDER (www.simus-systems.com)	✓	✓ ^a	✓				✓		✓
SMAP3D PartFinder (www.snap3d.com/CAD-partfinder.html)	✓ ^b	✓	✓				✓		✓
SMAP3D SmartSearch (www.snap3d.com/CAD-smartsearch.html)	✓	✓ ^a		✓		✓			✓
Geolus Search (www.plm.automation.siemens.com/en_us/products/open/geolus)	✓	✓ ^a	✓	✓		✓	✓	✓ ^c	✓
GEOMETRICAL Search Desktop (geosearch.cadenas.de)	✓	✓ ^a	✓	✓	✓	✓	✓	✓ ^d	✓
CADSeek (www.iseekcorp.com)	✓	✓ ^a	✓	✓		✓	✓		✓
3DSearchIT (3dsearchit.geometriglobal.com)	✓	✓ ^a	✓	✓		✓	✓		?
SQ Similarity Query (www.reknow.de)	✓	✓ ^a	✓		✓	✓	✓		✓
CADFind sketch & search (www.sketchandsearch.com)	✓	✓ ^a		✓	✓		✓ ^e		✓

^anur CAD-Dokumente

^bnur Normteile

^cTeilegröße

^dLochbild, Profilquerschnitt, Größe

^enur in Form eines Group Technology (GT) Codes

stellt und verwaltet werden. Dabei handelt es sich einerseits um strukturierte Daten, wie sie beispielsweise in DB-, PDM-, CRM- oder ERP-Systemen enthalten sind. Der weitaus größere Anteil der Informationen in einem Unternehmen besteht jedoch aus semi-strukturierten und unstrukturierten Daten [GR09], die sich durch einen geringen (Beispiele: Internet-/Intranetseiten, Word-Dokumente, Tabellen, ...) bzw. keinen Strukturierungsgrad (Beispiele: Präsentationen, E-Mails, Blogeinträge, ...) auszeichnen. Demzufolge muss ein ESS eine kombinierte Suche nach diesen drei Informationsarten unabhängig von ihrem jeweiligen Quellsystem ermöglichen, wobei zusätzliche Aspekte, wie die Unterstützung zahlreicher Dateiformate und Sprachversionen sowie die im Unternehmen vorhandenen Zugriffsrechte der Nutzer, zu berücksichtigen sind.

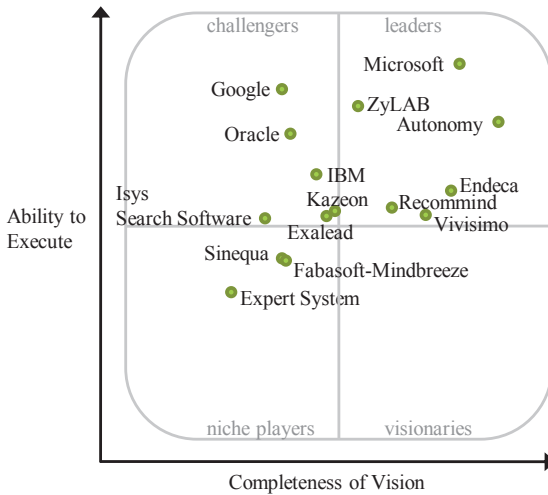


Abbildung 2.6. — Kommerzielle Anbieter von ES-Lösungen gemäß dem *Magic Quadrant for Information Access Technology* von Gartner, Inc. (nach [Whi09])

Wie in Abbildung 2.6 zu sehen ist, gibt es zahlreiche Unternehmen, die derartige ES-Lösungen kommerziell am Markt anbieten. Allerdings zeigt eine genauere Analyse dieser Softwareprodukte, dass sich die AwS nicht nur im Hinblick auf die Anzahl und Art der unterstützten Quellsysteme, Dateiformate³⁸ oder Sprachversionen unterscheiden, sondern vielmehr in den Möglichkeiten, die dem Nutzer mittels der GUI zum Suchen angeboten werden. Die Bandbreite reicht diesbezüglich von sehr einfachen Oberflächen, bestehend aus nur einem Textfeld zur Schlagworteingabe (vgl. Sucho-

³⁸Je nach Systemanbieter werden zwischen 220 (Google Search Appliance [Goo09]) und 1000 (Autonomy IDOL Retina [Aut10c]) verschiedene Dateiformate unterstützt.

berfläche der Google Search Appliance in Abbildung 2.7 links), bis zu aus mehreren Bereichen zusammengesetzten Oberflächen (vgl. Suchoberfläche der Endeca Information Access Platform in Abbildung 2.7 rechts), welche eine detailliertere Analyse der zurückgelieferten Suchergebnisse und damit eine Reformulierung der Suchanfrage anhand zusätzlicher Filterkriterien, Tag Clouds oder Karten- und Diagrammdarstellungen ermöglichen. Unabhängig davon ist aber allen ESS gemeinsam, dass sie sich rein auf textuelle Informationen konzentrieren, für deren Auffinden verschiedene Suchparadigmen (Volltext-, Metadatensuche, Browsing, Facettierte Navigation) angewendet werden.

Obwohl ESS somit einen Zugriff auf eine Vielzahl unterschiedlicher Informationsarten ermöglichen, sind sie für den Bereich der technischen PE nur bedingt geeignet. Dies liegt v. a. darin begründet, dass einerseits spezielle Dokumenttypen wie beispielsweise CAD-Modelle oder technische Zeichnungen mit ihren Dateiformaten bisher kaum und andererseits eine Suche nach ähnlichen Objekten – insbesondere nach Produkten – auf Basis graphischer Beschreibungen nicht von den bestehenden Systemen unterstützt werden. [Aut10d, End10, Exa08, Goo10b, Int10a, Viv10]

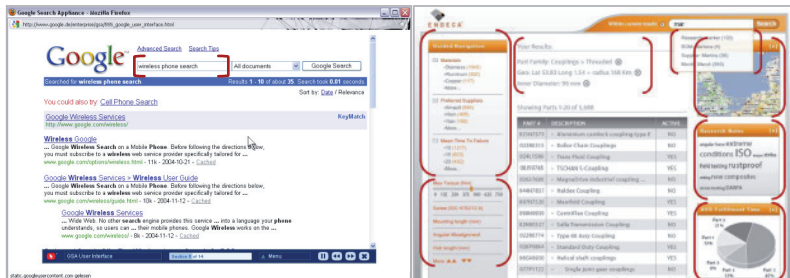


Abbildung 2.7. — Bandbreite der von ES-Lösungen bereitgestellten Suchmöglichkeiten. Links: GUI der Google Search Appliance mit einer einfachen Schlagwortsuche (entnommen aus [Goo10a]); Rechts: GUI der Endeca Information Access Platform mit zahlreichen Möglichkeiten zur Anfrageverfeinerung (entnommen aus [End09])

2.5. Konzepte und Ansätze aus der Forschung

Da bis heute keine allumfassende und ganzheitliche Lösung zur vollständigen Informationsversorgung von Produktentwicklern existiert, beschäftigt sich auch die Forschung weiterhin intensiv mit dieser Thematik. Aufgrund der bereits in Kapitel 1.1 erläuterten hohen Komplexität ist eine hohe Heterogenität an Ansätzen und Konzepten, die unterschiedliche Aspekte fokussieren, entstanden. Abhängig vom Grad der jeweils

geleisteten Unterstützung lassen sich diese in spezialisierte und generelle Konzepte differenzieren.

2.5.1. Spezialisierte Konzepte mit niedrigem Unterstützungsgrad

Konzepte dieser Gruppe zeichnen sich durch eine Fokussierung auf einen einzelnen Aspekt des Gesamtproblems, nämlich die Befriedigung bestimmter Informationsbedarfe, aus. Dadurch ist ihr Unterstützungsgrad zwar als eher niedrig einzustufen, jedoch liefern sie optimierte Lösungen für diese Teilprobleme, welche wiederum für eine Gesamtlösung von essentieller Bedeutung sind.

Suche nach ähnlichen Produkten

Betrachtet man sich die Forschung in diesem Bereich, so werden derzeit zwei Ähnlichkeitsaspekte intensiv erforscht. Dabei handelt es sich zum einen um die geometrische Ähnlichkeitssuche, die trotz erster kommerzieller Lösungen (vgl. Kapitel 2.3) noch immer große Optimierungspotentiale bietet. Zwar existieren bereits diverse Ansätze zum Erstellen der für einen rechnerinternen Vergleich benötigten Repräsentationen (siehe hierzu Kapitel 5.4.2.4); allerdings sind diese auf eine einheitliche Beschreibung von Anfrage und Datenbasis und damit eine homogene Ergebnismenge beschränkt. Nur wenige Konzepte sind hinsichtlich einer Heterogenität beider Aspekte bereits weiter fortgeschritten. Dazu gehören die an der Purdue bzw. der Princeton University entwickelten Systeme ShapeLab und 3D Model Search Engine, wobei letztere als Vorreiter der vorgestellten kommerziellen Systeme betrachtet werden kann. Als Websuchmaschine für 3D-Modelle von der Princeton Shape Retrieval and Analysis Group³⁹ entwickelt [FMK⁺03, MHKF03], beschränkt sie sich zwar rein auf 3D-Modelle, bietet dem Nutzer aber die Möglichkeit, Beispielobjekte für die Anfrage nicht nur in 3D, sondern auch in 2D anzugeben. Dazu können bis zu drei 2D-Ansichten oder eine 3D-Ansicht des benötigten Produktes mit Hilfe des Skizzierungstools Teddy [MCF02, IMT99] erstellt werden. Abhängig von der Form der Anfrage wird anschließend der Ähnlichkeitsvergleich entweder im 2D- oder im 3D-Raum anhand jeweils geeigneter Repräsentationen durchgeführt. Das daraus resultierende Ergebnis kann anschließend durch Veränderung der Ursprungsskizze oder durch Verwendung eines Elementes der Ergebnismenge verfeinert werden. Diese Funktionalität der 2D- / 3D-Anfrageformulierung wird auch von dem, speziell auf die Bereiche der Produktentwicklung und der Biologie zugeschnittenen System ShapeLab unterstützt. Im Vergleich erweitert allerdings die PRECISE⁴⁰-Forschungsgruppe um K. Ramani ihr Konzept um die Berücksichtigung verschiedener Geometriebeschreibungen in der Ergebnismenge [PR05, PR06]. Folglich

³⁹<http://www.cs.princeton.edu/gfx/proj/shape/>

⁴⁰Purdue Research and Education Center for Information Systems in Engineering (<https://engineering.purdue.edu/PRECISE/>)

wird nicht mehr nur auf einer homogenen Datenbasis operiert, so dass unabhängig von der Form der Anfrage sowohl 2D- als auch 3D-Modelle gefunden werden. Dies wird durch eine generelle Verlagerung des Ähnlichkeitsvergleiches in den 2D-Raum erreicht, wofür sämtliche 3D-Modelle durch eine Menge von auf ihren Hauptachsen basierenden 2D-Ansichten beschrieben werden. Allen bisher betrachteten Systemen ist außerdem gemeinsam, dass sie die Existenz eines Beispielobjektes zur Anfrage voraussetzen. Dies ist jedoch nicht in jeder Situation gegeben, da entweder keine konkrete Vorstellung über das gesuchte Produkt besteht, kein passendes CAD-Modell existiert oder eine Skizze aufgrund der Produktkomplexität zu aufwändig ist. Aus diesem Grund wurden Ansätze entwickelt, die anstelle einer QbE-Suche eine navigierende Ähnlichkeitssuche bieten. Mittels geeigneter Clusteralgorithmen werden dazu die Produkte anhand ihrer Ähnlichkeit in Kategorien (Cluster) eingruppiert, welche in Form von Baum- oder Verzeichnisstrukturen visualisiert werden. Folglich kann der Nutzer nach einem, seinen Erwartungen entsprechendem Produkt stöbern und bekommt bei Erfolg gleichzeitig auch ähnliche Produkte mitgeliefert. Die dabei erzeugten Cluster können zum einen hierarchisch strukturiert werden, wie es bei dem von Kriegel et al. entwickelten Prototyp BOSS⁴¹ anhand des dichte-basierten Algorithmus CLUSS⁴² [BKK⁺04, KBK⁺06] oder auch bei dem von der PRECISE-Forschungsgruppe erweiterten ShapeLab-System 3DESS⁴³ mittels einer Self-Organizing Map [LPR04, LJI⁺03] erfolgt. Da jedoch v. a. bei komplexen Produkten normalerweise keine hierarchische Struktur existiert, präsentiert Chakraborty in [Cha05] einen einfachen Algorithmus, der in nur einem Durchgang die Cluster anhand lokaler Ähnlichkeitsbedingungen unter Nutzung der diskreten Fouriertransformation bestimmt.

Der zweite, in der Forschung intensiv betrachtete Ähnlichkeitsaspekt bezieht sich auf die Fertigungskosten eines Produktes. Da diese maßgeblich durch die im Herstellungsprozess durchzuführenden Operationen zur Realisierung notwendiger Form-Features⁴⁴ (= Formelemente, wie Bohrungen, Fasen, Rundungen, Nuten, . . .) beeinflusst werden, sind sie für den weiteren Entwicklungsverlauf von entscheidender Bedeutung. Für eine schnelle Abschätzung dieser Kosten wird daher eine Suche nach fertigungsähnlichen Produkten benötigt. Dabei kann der Begriff der Fertigungsähnlichkeit allerdings unterschiedlich beurteilt werden. So ist gemäß Karnik et al. ein Produkt P_2 fertigungsähnlich zu einem Anfrageprodukt P_1 , wenn P_1 unter Verwendung von ausschließlich materialentfernenden Operationen aus P_2 hergestellt werden kann. Zur Prüfung dieser Bedingung schlagen Karnik et al. in [KGM05] einen dreistufigen Algorithmus vor, der prüft, ob die Geometrie von P_2 unter Berücksichtigung bestimmter Bedingungen (z. B.

⁴¹Browsing OPTICS-Plots for Similarity Search

⁴²Cluster Hierarchies for Similarity Search (eine Optimierung des ursprünglichen Clusteralgorithmus OPTICS = Ordering Points to Identify the Clustering Structure)

⁴³3D Engineering Shape Search

⁴⁴Im Maschinenbau beschreibt ein Feature gemäß Pahl et al. den Zusammenhang zwischen geometrischen und nicht-geometrischen Informationen, wobei abhängig vom Konkretisierungsgrad zwischen Funktions-, Prinzip-, Bauteil- und Form-Features differenziert wird [PBFG07].

umgebender Zylinder, parallel zur Rotationsachse gefertigte Features) die Geometrie von P_1 vollständig beinhaltet (*geometric containment analysis*). Ist dies der Fall, so wird der Ähnlichkeitswert von P_2 mit Hilfe der Volumendifferenz von P_1 und P_2 ermittelt. Im Gegensatz dazu kann eine Fertigungsähnlichkeit zweier Produkte auch durch ähnliche Form-Features gegeben sein. Diese Sichtweise wird u. a. von Cardone et al. vertreten, die jedes Produkt durch die Menge ihrer Formelemente und jedes Formelement wiederum durch einen Vektor im fünfdimensionalen Raum, charakterisiert durch die Parameter Zugriffsrichtung, Typ, Volumen, Toleranz und Kardinalität, beschreiben [CGDK06]. Auch Ramesh et al. folgen diesem Ansatz und bestimmen für die Form-Features eines Produktes sieben charakteristische Kennzahlen (u. a. die Häufigkeitsverteilung der enthaltenen Featuretypen) [RYD01]. Diese werden zunächst separat im n-dimensionalen Raum verglichen und schließlich zur Gesamtähnlichkeit addiert. Daneben gibt es Konzepte, die nicht die Ähnlichkeit der Form-Features an sich, sondern vielmehr deren Abhängigkeiten voneinander zur Beurteilung der Fertigungsähnlichkeit betrachten. Um die Abfolge der Features im Fertigungsprozess zu vergleichen, werden hierzu Graphen, wie der von Cicirello und Regli in [CR01] vorgestellte Model Dependency Graph — ein gerichteter azyklischer Graph mit Features als Knoten und gerichteten Abhängigkeiten als Kanten — eingesetzt. Insgesamt ist jedoch allen Konzepten gemeinsam, dass die Features bekannt sein müssen. Sie können zwar manuell definiert werden, sollten aber möglichst automatisch aus den Geometriedaten erkannt und extrahiert werden. Han et al. geben diesbezüglich in [HPR00] einen ausführlichen Überblick entwickelter Ansätze, die auf drei grundlegenden Vorgehensweisen, nämlich dem Einsatz graphbasierter Algorithmen, der Nutzung von Techniken der Volumenzerlegung oder dem geometrischen Schlussfolgern anhand von Hinweisen, basieren.

Suche nach Design Rationales

Wie in Kapitel 1.1.2 erwähnt, handelt es sich bei Design Rationales um Informationen darüber, welche Entscheidungen im Entwicklungsprozess aus welchen Gründen wie getroffen bzw. warum bestimmte Entscheidungen nicht gefällt wurden [BB98]. Auf Basis dieser Information kann einerseits eine zurückliegende Entwicklung einfacher nachvollzogen und andererseits eine zukünftige Entwicklung aufgrund von Lessons Learned- und Best Practice-Erfahrungen effektiver und effizienter durchgeführt werden. Da dies aber nicht nur speziell für den Bereich der technischen PE, sondern auch für andere Domänen (z. B. die Softwareentwicklung) gilt, wurden seit den 1980er Jahren diverse DR-Systeme in der Forschung entwickelt. Sowohl Regli et al. [RHAS00] als auch Burge und Brown [BB98] geben hierzu eine ausführliche Übersicht, wobei u. a. die Art der Informationserfassung und die unterstützten Retrievalmechanismen analysiert wurden. Davon ausgehend sind bezogen auf die Aufgabe der Informationserfassung prinzipiell zwei Vorgehensweisen denkbar. Zum einen kann unter Nutzung von IT-Werkzeugen wie Telefon-, Videokonferenz-, Mailsystemen u. a. die Kommunikation und Koope-

ration während der Entwicklung überwacht und dabei die Information automatisch dokumentiert werden. Allerdings besitzen die daraus resultierenden Aufzeichnungen aufgrund der Tatsache, dass eine Kommunikation zwischen Personen nicht nach festen Regeln abläuft, üblicherweise keine feste Struktur, weshalb die damit verbundene Semantik verloren geht. Aus diesem Grund erfordern die meisten DR-Systeme einen manuellen Eingriff, bei dem die Entwicklungshistorie bestehend aus Tätigkeiten, Arbeitsablauf, Kommunikationen usw. rein textuell in Form von Dokumenten festgehalten wird. Da dadurch allerdings nur eine Volltextsuche in den Dokumentinhalten realisierbar ist, wurden Benutzeroberflächen entwickelt, welche eine graphische Beschreibung in Form eines Netzes aus Knoten (= Problem/Entscheidung, Option, Argument, ...) und Kanten (= Beziehungen zwischen den Knoten) unter Verwendung spezieller argumentationsbasierter Modelle wie IBIS oder QOC⁴⁵ ermöglichen. Damit können generell zwei Mechanismen zum Auffinden von DR unterstützt werden: Ein unspezifisches Browsing in der Netzstruktur sowie die gezielte Beantwortung von Fragen nach Gründen als auch Auswirkungen von Entscheidungen durch Verfolgen von Beziehungskanten. Findet zusätzlich zur automatischen Informationserfassung eine Prozessüberwachung statt, so können Entscheidungen mit vordefinierten Regeln oder Kriterien verglichen und bei Abweichungen die entsprechenden DR dem Nutzer proaktiv angezeigt werden. [RHAS00, BB98, WJB09]

Suche nach Dokumentinhalten

Da insbesondere Entwicklungsdokumente wie Lasten-, Pflichtenhefte oder Testberichte zahlreiche Seiten umfassen, entwickelten Liu et al. das Engineering Document Content Management System (EDCMS) zum effizienten Zugriff auf deren Inhalte [LMD⁺06, LMD⁺07]. Dazu werden für jedes Dokument zunächst multiple Dokumentstrukturen auf Basis vordefinierter Zerlegungsschemata erstellt, welche unterschiedliche Sichten (z. B. physikalisch $\hat{=}$ Kapitel, Unterkapitel, ...; logisch $\hat{=}$ Ziel, Einleitung, Fazit, ...) auf die Struktur repräsentieren. Diese Dokumentstrukturen werden anschließend mit Hilfe von Auszeichnungssprachen (XML für Textdokumente, Hypertext Markup Language (HTML) für Pixelbilder, Scalable Vector Graphics (SVG) für Vektorgraphiken) formalisiert und gemeinsam mit den aus den Strukturfragmenten extrahierten Inhalten in ein facettierte Hierarchieschema eingeordnet. Das hierzu eingesetzte System Waypoint [MLC04] nimmt dabei u. a. unter Prüfung bestimmter Bedingungen eine Zuordnung zu einem Themen- und einem Strukturindex vor, was neben einer Volltextsuche ein Navigieren im Datenbestand ermöglicht. Obwohl dieser Ansatz vielversprechende Ergebnisse liefert, ist zu berücksichtigen, dass eine manuelle Transformation der Dokumente in die Auszeichnungssprachen vorausgesetzt wird. Zwar schlagen die Autoren hierfür die Verwendung spezieller Werkzeuge vor; für deren Anwendung muss allerdings zusätzlich sichergestellt sein, dass alle Dokumente eines

⁴⁵Für eine detaillierte Erläuterung dieser Notationen wird auf [RHAS00] verwiesen.

bestimmten Typs dieselbe Struktur aufweisen. Dies könnte prinzipiell über Dokumentvorlagen erreicht werden, jedoch werden diese — obwohl sie meist in den Unternehmen verfügbar sind — in der alltäglichen Praxis äußerst selten verwendet.

Suche nach Problemlösungen

Die Entwicklung eines neuen Produktes ist in den meisten Fällen mit einem neuen Problem verbunden, welches gelöst werden muss. Hierzu kann es hilfreich sein, bereits früher gelöste Probleme, die eine gewisse Ähnlichkeit zur neuen Problemstellung aufweisen, näher zu betrachten. Demzufolge wird eine Suche nach ähnlichen Problemen benötigt, wofür auf den aus dem Bereich der Künstlichen Intelligenz stammenden Ansatz des Fallbasierten Schließens — Case-Based Reasoning (CBR) — zurückgegriffen werden kann. Die Idee, Erfahrungswissen in Form von Fällen, bestehend aus einer Problembeschreibung und der zugehörigen Lösung, in einer Falldatenbank zu speichern und über einen Ähnlichkeitsvergleich zugänglich zu machen, ist nicht nur Basis zahlreicher Anwendungen in den Bereichen der medizinischen Therapie oder des Kundenservice [BK09]. Auch Systeme für den Bereich der technischen PE greifen diesen Ansatz auf. Neben dem auf die konzeptionelle Entwicklung ausgelegten System CADET, welches ein Graph-Matching zum Ähnlichkeitsvergleich durchführt, sind hier noch zahlreiche andere Systeme entstanden, weshalb für eine ausführliche Betrachtung auf [MP97, SSR01, QR03] verwiesen wird.

2.5.2. Generelle Konzepte mit hohem Unterstützungsgrad

Im Vergleich zu den spezialisierten Konzepten sehen generelle Forschungsarbeiten ihren Schwerpunkt in einer umfassenden Unterstützung von Produktentwicklern und versuchen dazu möglichst viele Informationsbedarfe während des PEP zu befriedigen. Folglich besteht deren Ziel darin, Wissen, welches sich aus diversen Informationsarten zusammensetzt, auffindbar zu machen. Dazu kann die Suche in einer heterogenen Wissenssammlung auf zwei Arten unterstützt werden. Zum einen existieren Konzepte, die sich an den aus dem WM bekannten Knowledge Organization Systems (KOS) orientieren und das gesamte Wissen in einer einheitlichen Form strukturieren. Gemäß Hodge lassen sich hierzu generell Termlisten, Klassifikationen / Kategorisierungen oder Beziehungsstrukturen nutzen [Hod00], wobei Termlisten⁴⁶ für die betrachtete Zielsetzung allein nicht ausreichend sind. Folglich finden sich insbesondere klassifikations- und beziehungsstrukturbasierte Ansätze. Unter ersteren versteht man dabei Konzepte, die das Wissen in Objektmengen organisieren. Wie bei Pavković et al. [PMHR09] werden dazu vorrangig Taxonomien eingesetzt, welche, wie im vorliegenden Beispiel, jedoch nicht notwendigerweise vordefiniert sind. Vielmehr bieten Pavković et al. dem

⁴⁶Dazu zählen Glossare, Wörterbücher, einfache Synonymlisten oder Gazetteers (= geographische Ortsverzeichnisse) [Hod00].

Nutzer die Möglichkeit, diese graphisch in Form einer Verzeichnisstruktur in der GUI zu erstellen und zu modifizieren, wodurch das manuelle Einpflegen von Wissen vereinfacht, die Verständlichkeit verbessert und damit die Motivation zur Bewahrung von Wissen erhöht werden sollen. Das zur Taxonomie zugeordnete Wissen wird hier schließlich mit Hilfe von Tags, welche die einzelnen Taxonomieelemente repräsentieren, in einer relationalen DB gespeichert. Somit kann bei einer Suchanfrage, deren Formulierung durch Markieren von Taxonomieelementen erfolgt, eine SQL-Abfrage nach den zugehörigen Tags durchgeführt werden. Allerdings zeigen Studien von Ahmed und Wallace, dass die Verwendung einer einzigen Taxonomie zur Beschreibung des PEP und seiner Informationen nicht ausreicht [AW03]. Sie empfehlen die Verwendung von insgesamt vier Taxonomien, die aufgrund von Interviews mit Entwicklern definiert und für eine allgemeine Anwendung optimiert wurden. Es handelt sich hierbei um eine Prozess-, eine Produkt-, eine Funktions- und eine Thementaxonomie, wobei letztere sämtliche, bei der Entwicklung zu beachtenden Aspekte wie Stückkosten, Gewicht, Recycling usw. abdeckt. Bei beziehungsstrukturbasierten Konzepten werden zusätzlich zu einer Strukturierung in Objektmengen explizit auch die Beziehungen zwischen den Objektmengen berücksichtigt. Dies kann, wie im Rahmen des Design Repository-Projektes des National Institute of Standards and Technology, beispielsweise in Form einer objektorientierten Repräsentation des Wissens erfolgen [SSBR98]. Da ein Design Repository ein „(...) intelligent knowledge-based design artifact modeling system used to facilitate the representation, capture, sharing, and reuse of corporate design knowledge.“ [SSB⁺00, S. 2] darstellt, muss das Wissen hier erst über einen entsprechenden Editor manuell eingepflegt werden. Als Wissen werden dabei die im PEP entwickelten Artefakte (= Produkte) mit all ihren zugehörigen Informationen aufgefasst, welche hinsichtlich ihrer Form, ihrer Funktion, ihres Verhaltens sowie deren Beziehungen zueinander zu definieren sind [SRSB00]. Die einzelnen Objektstrukturen können schließlich über Hyperlinks in einer webbasierten GUI erforscht werden. Anstelle derartiger Objektstrukturen werden zur Repräsentation allerdings häufiger semantische Netze [CDK⁺07, WPS04, Ahm06b] oder Ontologien eingesetzt. Hierbei werden Begriffe / Konzepte als Knoten durch Beziehungen, die durch Kanten abgebildet werden, zu einem Netzwerk verbunden. Da abhängig von den bei der Erstellung jeweils betrachteten Informationen (z. B. Produkt-Prozess-Material-Funktion oder Funktion-Verhalten-Struktur) zahlreiche unterschiedliche Ontologien entstanden sind, wird für weitere Informationen auf die umfassende Literatur zu diesem Bereich verwiesen [LFB96, BMM⁺06, LAR05, LSL06, LRR07]. Insgesamt betrachtet sind alle diese Ansätze schwerpunktmäßig auf ein Browsing im Wissensbestand ausgerichtet; erweitern dieses jedoch teilweise mit Funktionalitäten einer parameterbasierten Suche.

Zusätzlich zu einer derartigen Wissensorganisation sollten für eine Suche in heterogenen Wissensbeständen jedoch auch die Ideen oder Vorgehensweisen der spezialisierten Konzepte aufgegriffen und für eine Gesamtlösung mit integriert werden. Ein erster

Versuch hierzu wurde von Karnik et al. mit dem für den Militärbereich entwickelten System mit dem Namen Design Navigator [KGA⁺05] unternommen. Hierbei handelt es um ein System, in dem die gesamte Entwicklung eines Produktes durchzuführen ist. Dazu setzt es einerseits auf dem verwendeten CAD-System auf und verfügt andererseits über zusätzliche Modellierungswerkzeuge, welche die Eingabe von speziellem Wissen, wie der Entwicklungshistorie, der Funktionsstruktur eines Produktes oder Design Rationales, unterstützen. Für die Suche nach Informationen stellt das System vier Suchwerkzeuge bereit, die einerseits anhand von Taxonomien eine Suche nach Funktionen und Design Rationales sowie eine Parametersuche nach administrativen Daten ermöglichen. Darüber hinaus kann neben einer geometriebasierten Suche auf Basis eines Ähnlichkeitsvergleiches, für den jedes Produkt durch vier Signaturen beschrieben wird, auch visuell nach Produkten gesucht werden [KAE⁺05]. Für Letzteres können die Produkte entweder nach ihrer Größe (= Radius der umgebenden Kugel) oder nach ihrer Komplexität (= Anzahl der Oberflächen) sortiert angezeigt und gefiltert werden. Obwohl dieser Ansatz interessante Möglichkeiten aufzeigt, findet hier — mit Ausnahme von CAD-Modellen — keine Berücksichtigung von in Unternehmen existierenden Dokumenttypen und deren Inhalten statt.

2.6. Unterstützte Suchparadigmen

Zusammenfassend zeigt Tabelle 2.2 die in den Kapiteln 2.1 bis 2.5 erläuterten Ansätze zusammen mit den von ihnen jeweils unterstützten Suchparadigmen. Da in der Literatur keine allgemein gültige Klassifizierung der im IR üblichen Suchparadigmen existiert, werden im Weiteren die für die vorliegende Arbeit relevanten Suchparadigmen in die drei nicht-disjunkten Paradigmen der zielgerichteten Suche, der explorativen Suche und der Ähnlichkeitssuche differenziert. Diese Unterscheidung wurde auf Basis der in der Literatur vorzufindenden zahlreichen Definitionen und Begrifflichkeiten, welche insbesondere in Forschungsarbeiten bezüglich des Such- bzw. Informationsbeschaffungsprozesses (*information seeking process*) entstanden sind [IJ05, Hea09], getroffen. Demnach ist die zielgerichtete Suche als Realisierung des in diesem Bereich postulierten kognitiven Modells der Informationssuche bzw. -beschaffung zu verstehen, bei dem ein relativ statischer Informationsbedarf mit Hilfe von möglichst konkreten Anfragen befriedigt werden soll [Hea09]. Dies kann u. a. mit Hilfe einer FS oder einer MPS erfolgen. Demgegenüber steht das dynamische Modell der Informationssuche, welches u. a. von Bates als *Berrypicking* [Bat89], von O'Day und Jeffries als *Orienteering* [OJ93] oder von White und Roth als explorative Suche [WR09] bezeichnet wird. Dieses Modell geht von der Annahme aus, dass der Informationsbedarf eines Nutzers nicht statisch ist, sondern sich durch Interaktionen des Nutzers mit dem Suchsystem kontinuierlich verändert [Hea09]. Diese Interaktionen finden dabei typischerweise in Form von Browsing oder Facettierter Navigation statt, so dass diese

konkrete Ausprägungen des Paradigmas der explorativen Suche darstellen. Da man sowohl zielgerichtet als auch explorativ nach ähnlichen Artefakten suchen kann, lässt sich das dritte Paradigma der Ähnlichkeitssuche genau genommen in jede der beiden vorher genannten Gruppen einordnen. Aufgrund seines Fokus auf den Vergleich von Ähnlichkeiten nimmt es allerdings eine besondere Bedeutung im IR ein und wird daher auch in der vorliegenden Arbeit als gesondertes Suchparadigma mit den beiden prinzipiellen Ausprägungen der QbT- und der QbE-Suche betrachtet.

Tabelle 2.2. — Übersicht der von den bereits existierenden Konzepten unterstützten Suchparadigmen mit Charakterisierung des Retrievalprozesses: S $\hat{=}$ Set Retrieval, R $\hat{=}$ Ranked Retrieval und SR $\hat{=}$ Set oder Ranked Retrieval (je nach Retrievalmodell)

	<i>Zielgerichtete Suche</i>		<i>Explorative Suche</i>		<i>Ähnlichkeitssuche</i>	
	FS	MPS	B	FN	QbT	QbE
Ordnungsschemata		\sqrt{S}	\sqrt{S}			
Klassifikationssysteme		\sqrt{S}	\sqrt{S}	\sqrt{S}		
DBS	\sqrt{S}	\sqrt{S}				\sqrt{S}
DMS		\sqrt{S}	\sqrt{S}		\sqrt{SR}	
ERP-Systeme	\sqrt{S}	\sqrt{S}	$(\sqrt{S})^a$		$(\sqrt{SR})^a$	$(\sqrt{S})^b$
PDMS		\sqrt{S}	\sqrt{S}		\sqrt{SR}	
ESS		\sqrt{S}	\sqrt{S}	\sqrt{S}	\sqrt{SR}	
CAD-IR-Systeme		\sqrt{S}				\sqrt{SR}
Lsg. zur Produktsuche			\sqrt{S}			\sqrt{R}
DR-Systeme			\sqrt{S}		\sqrt{SR}	
EDCMS				\sqrt{S}	\sqrt{S}	
CBR-Systeme						\sqrt{SR}
KOS-orientierte Konzepte		\sqrt{S}	\sqrt{S}			
Design Navigator		\sqrt{S}	\sqrt{S}			\sqrt{R}

^awenn DMS integriert

^bwenn von DB bereitgestellt

Insgesamt verdeutlicht die Tabelle 2.2, dass bereits viele Ansätze zur Verbesserung der Informationsversorgung im Bereich der technischen PE existieren. Allerdings adressieren diese jeweils nur bestimmte Aspekte oder unterstützen nur bestimmte Suchparadigmen. Eine Kombination von Suchparadigmen im Sinne einer multikriteriellen Suche, die — wie anhand der beispielhaften Suchszenarios in Kapitel 1.1.2 auf Seite 13 ff. verdeutlicht wurde — gerade für den Bereich der technischen PE benötigt wird, ist nur in Ausnahmefällen möglich. So sind beispielsweise alle Konzepte, die eine Kombination einer QbE-Suche mit einer MPS erlauben, auf einen bestimmten Artefakttyp

— meist das Produkt — beschränkt. Konzepte, die gerade diese enge Fokussierung auf einen bestimmten Artefakttyp vermeiden, weisen hingegen einen Mangel an anderen Aspekten auf. Dies wird am Beispiel von ESS ersichtlich, welche zwar allgemein, aber auch branchenübergreifend ausgerichtet sind, so dass Informationen aus entwicklungsspezifischen Dokumenttypen hier nicht berücksichtigt werden. Andere wiederum, wie das Design Navigator System, erfordern zu viel manuelles Eingreifen von den Produktentwicklern und betrachten zu wenig die im Unternehmen bereits vorhandenen Informationsquellen. Demzufolge werden zwar viele interessante Lösungsmöglichkeiten aufgezeigt; eine umfassende und übergreifende Suchfunktionalität, die diverse Artefakttypen einbezieht und verschiedene Suchmöglichkeiten anbietet, existiert bisher jedoch nicht. Aus diesem Grund wird im folgenden Kapitel 3 ein interaktives Retrievalmodell für komplexe Suchsituationen vorgestellt, das diese Aspekte berücksichtigt.

3. Interaktives Retrievalmodell für komplexe Suchsituationen

In den folgenden Kapiteln wird das im Rahmen des Forschungsverbundes FORFLOW entwickelte interaktive Retrievalmodell zur Unterstützung von Produktentwicklern in komplexen Suchsituationen vorgestellt. Dieses basiert auf der Annahme, dass sämtliche, für Produktentwickler hilfreiche und relevante Informationen während des PEP erstellt oder erfasst werden und somit in diversen IT-Systemen verfügbar sind. Damit ist es nicht Ziel des Konzeptes, Mechanismen zur Eingabe von neuem Wissen zur Verfügung zu stellen, sondern die in einem Unternehmen bereits existierenden Informationen über geeignete Suchmechanismen auffindbar zu machen. Nach einer Darstellung der an ein derartiges System gestellten Anforderungen (Kapitel 3.1), werden in Kapitel 3.2 die zu ihrer Realisierung verwendeten Konzepte erläutert. Deren Integration in ein Gesamtsystem wird schließlich in Kapitel 3.3 veranschaulicht, wobei zwar beide Teilprozesse eines IRS — der Retrieval- und der Indexierungsprozess — beschrieben werden, letzterer jedoch im Mittelpunkt der Betrachtungen liegt.

3.1. Anforderungen an ein Information Retrieval-System für die Produktentwicklung

Die aus den vorhergehenden Betrachtungen und aus der Zusammenarbeit mit Industriepartnern im Rahmen des Forschungsverbundes FORFLOW resultierenden Anforderungen an ein übergreifendes IRS lassen sich prinzipiell in vier Kategorien einordnen. Zwei dieser Kategorien umfassen dabei funktionale Anforderungen sowohl an den Retrieval- als auch an den Indexierungsprozess, wobei letztere sozusagen implizit in den Retrievalanforderungen enthalten sind und damit direkt aus diesen abgeleitet werden können. Hierzu gehören zum einen Vorstellungen über die zu unterstützenden Suchfunktionalitäten und zum anderen Aspekte, die im Hinblick auf die Suchergebnisse bzw. deren Präsentation zu berücksichtigen sind. Beide Kategorien werden in Tabelle 3.1 konkretisiert, wobei den Retrievalanforderungen die jeweils damit einhergehenden Anforderungen an den Indexierungsprozess gegenübergestellt werden.

Tabelle 3.1. — Anforderungen an den Retrievalprozess hinsichtlich der zu unterstützenden Suchfunktionalitäten sowie der Ergebnispräsentation und die damit einhergehenden Anforderungen an die Indexierung

Anforderungen Retrieval	Anforderungen Indexierung
<i>Suchfunktionalitäten</i>	
Finden verschiedener Artefakttypen	Differenzieren & Erzeugen typspezifischer Repräsentationen
Exploratives Suchen, d.h. Navigieren im Datenbestand	Verwenden einer Repräsentationsform, die Navigation erlaubt
Parameterbasiertes Suchen	Indexieren von artefakttypspezifischen Suchparametern
Suchen in Inhalten von Textdokumenten	Extrahieren & Aufbereiten von Inhalten aus Textdokumenten diverser Dateiformate
Explizites Berücksichtigen entwicklungsspezifischer Dokumente beim Suchen	Differenzieren entwicklungsspez. Dokumenttypen & Dateiformate; Extraktion/Repräsentation ihrer Dokumentinhalte
Suchen ähnlicher Artefakte anhand diverser Ähnlichkeitsaspekte (Text, Geometrie, ...)	Erstellen multipler Repräsentationen zur Ähnlichkeitssuche
Wechseln zwischen Artefakttypen beim Suchen ($\hat{=}$ Suchen anhand von Assoziationen)	Erfassen, Speichern & Bereitstellen von Artefaktbeziehungen
Suchen in allen Informationsbeständen des Unternehmens ($\hat{=}$ single point of access)	Erfassen & Indexieren von Informationen aus bestehenden IT-Systemen
Situationsabhängiges Bereitstellen von Informationen	Einbeziehen von Kontextinformationen aus z. B. PMS
<i>Ergebnispräsentation</i>	
Schnelle Ergebnisdarstellung ($\hat{=}$ kurze Antwortzeiten)	Einsatz effizienter Indexstrukturen
Übersichtliche & aussagekräftige Präsentation großer Ergebnismengen inkl. Vorschau	Erzeugen & Bereitstellen von Thumbnails/Snippets und identifizierender Merkmale
Anzeigen von Details zu ausgewählten Ergebnissen	
Zugriff auf Originalinformationen	Erfassen & Bereitstellen von Zugriffspfaden (z. B. zum Öffnen von Dokumenten)
Berücksichtigen von Zugriffsrechten bei der Ergebnisdarstellung	Gewährleistung von Identifikation und Zuordnung zu Quellsystemen
Präsentation aktueller Ergebnisse	Option zweier Indexierungsmodi (Pushmodus für Aktualisierungen & Pullmodus zu festen Zeitpunkten)
Erklärungskomponente zur besseren Nachvollziehbarkeit & Akzeptanzerhöhung	Erfassen & Bereitstellen von Informationen zur näheren Erklärung

Daneben ziehen Retrievalanforderungen der beiden anderen Kategorien nicht notwendigerweise Indexierungsanforderungen nach sich oder umgekehrt, weshalb sie im Folgenden kurz erläutert werden. Anforderungen der dritten Kategorie sind ebenfalls funktionaler Natur und befassen sich mit den erforderlichen Optionen der Anfrageformulierung. Folglich beziehen sie sich primär auf den Retrievalprozess, in dem der Nutzer neben einer Schlagworteingabe, wie sie von Websuchmaschinen bekannt ist, die Möglichkeit haben sollte, ein Beispielobjekt als Anfrage für eine Ähnlichkeitssuche anzugeben. Letzteres sollte sowohl in 2D als auch in 3D sowie unter Verwendung eines Elementes aus der Ergebnismenge im Sinne von *More Like This*-Anfragen [Nud10] definiert werden können. Des Weiteren besteht gerade für komplexere Suchsituationen die Notwendigkeit, mehrere Suchkriterien mittels einer multikriteriellen Suche beliebig kombinieren zu können. Schließlich ist für die Befriedigung des im Suchszenario D in Kapitel 1.1.2 auf Seite 15 beispielhaft beschriebenen Informationsbedarfs eine Suche nach Dichtungen allein mit Hilfe einer Ähnlichkeitssuche nach ringförmigen Produkten nicht ausreichend. Vielmehr sollte die Menge der gefundenen ähnlichen Produkte unter Hinzunahme weiterer Kriterien, wie beispielsweise dem Außen- und Innenradius oder auch der Dicke des Ringes, verfeinert werden können. Allerdings ist bei einer Kombination mehrerer Suchkriterien auch immer zu berücksichtigen, dass den einzelnen Kriterien abhängig von der Situation, in der sich ein Nutzer befindet, eine unterschiedliche Bedeutung im Suchprozess zukommt. So wird beispielsweise beim Entwerfen, d. h. beim geometrischen Modellieren eines Produktes mehr Wert auf die Geometrieähnlichkeit von Produkten gelegt, weshalb eine situationsspezifische Gewichtung der Suchkriterien erforderlich ist. Außerdem können, gerade bei ungenauen Vorstellungen über das Suchergebnis, nicht immer alle benötigten Suchkriterien sofort angegeben werden. Aus diesem Grund muss es möglich sein, die Anfrage schrittweise in eine bestimmte Richtung zu konkretisieren, wobei der Nutzer zwar in gewisser Weise geleitet bzw. gelenkt wird, er aber immer noch das Gefühl haben muss, die einzelnen Schritte der Suche selbst zu kontrollieren [Ahm06b, Bat90]. Damit wird eine Anfragesprache benötigt, die einerseits zwar einfach und leicht verständlich, aber andererseits auch mächtig genug ist, um — trotz Vagheit — komplexe Anfragen formulieren zu können.

Schließlich lassen sich auch nicht-funktionale Anforderungen identifizieren, die sich entweder auf den Retrieval- oder den Indexierungsprozess oder auch auf beide beziehen. Allgemein sollte ein IRS für mehrere Nutzer anwendbar sein und diesen eine einfache Handhabung bieten, indem es verständlich, klar strukturiert und einfach bzw. mit minimalem Schulungsaufwand zu bedienen ist. Schließlich soll der Nutzer nicht durch eine zu komplexe Anwendung abgeschreckt, sondern die Akzeptanz und damit die Nutzungshäufigkeit des Systems erhöht werden. Um dies zu erreichen, muss das System aber auch wirtschaftlich sein. Das bedeutet, dass der Aufwand für die Indexierung der Informationen und der Nutzen, der durch eine Bereitstellung relevanter Informationen erreicht werden kann (kürzere Such- und damit Entwicklungszeit, höherer

Wiederverwendungsgrad, niedrigere Kosten, ...), in einem angemessenen Verhältnis zueinander stehen müssen. Folglich ist gerade im Hinblick auf die Indexierung der Informationen ein möglichst hoher Automatisierungsgrad anzustreben. Dabei ist u. a. auf eine hohe Indexierungstiefe zu achten, die dann erreicht ist, wenn die Indexierung eines Artefakts durch eine oder auch mehrere Repräsentationen dessen Information bzw. Inhalt zu einem hohen Grad an Genauigkeit abdeckt [ABB⁺04]. Dies wiederum bedingt weitere Aspekte wie die Berücksichtigung verschiedener Sprachen v. a. in global agierenden Unternehmen, den Umgang mit mangelnder Datenqualität (häufig in Form nicht-gepflegter Metadaten) oder den Einsatz einer terminologischen Kontrolle unter Verwendung eines einheitlichen Vokabulars, um eine gemeinsame Sprache zu gewährleisten. Zusätzlich sollte das System sowohl unstrukturierte als auch semi- bzw. strukturierte Daten handhaben können, anpassungs- und erweiterungsfähig sowie plattformunabhängig anwendbar sein.

3.2. Basiskonzepte zur Erfüllung der Anforderungen

Um die aufgezeigten, vielfältigen Anforderungen bewältigen zu können, wird auf vier bestehende und bewährte Konzepte aus den Bereichen des IR, der Visualisierung und der Modellierung zurückgegriffen. Sie werden im Weiteren zunächst separat erläutert.

3.2.1. Facettierte Suche

Für die Realisierung sowohl einer zielgerichteten als auch einer explorativen Suche bietet sich das Konzept der facettierten Suche an. Hierbei handelt es sich um einen kombinierten Ansatz, der zusätzlich zu einer Schlagwort- bzw. Volltextsuche auch eine Facettierte Navigation im Datenbestand ermöglicht [Den03, Tun09]. Dazu werden dem Nutzer sogenannte Facetten zur Anfrageformulierung zur Verfügung gestellt. Diese repräsentieren Elemente einer polyhierarchischen Klassifikation (vgl. Kapitel 2.1.2), welche eindeutig definierte und sich gegenseitig ausschließende Aspekte oder Eigenschaften einer Objektmenge charakterisieren und dabei für jedes Objekt dieser Menge bestimmte Ausprägungen besitzen [WR09]. Beispiele hierfür sind die Facetten *Hersteller*, *Farbe* und *Baujahr* einer Objektmenge *Fahrzeuge*, die jedes einzelne Fahrzeug durch bestimmte, zugehörige Ausprägungen beschreiben. Dies ermöglicht dem Nutzer, ähnlich zu einer Metadaten- bzw. Parametersuche, eine Menge von Suchkriterien (z. B. *Farbe*=(schwarz OR rot) AND *Baujahr*>=2008) zu definieren, um bestimmte Objekte (hier Fahrzeuge) zu finden. Der Unterschied jedoch besteht darin, dass diese nicht alle gleichzeitig im ersten Suchschritt festgelegt, sondern iterativ zu einer Anfrage kombiniert werden. Dies erfolgt durch schrittweises Auswählen von Facetten und entsprechenden Ausprägungen, was den Vorteil einer sofortigen Sicht-

barkeit der Auswirkungen einer Selektion mit sich bringt [Tun09]. Hierzu wird dem Nutzer für jede Facette bzw. Facettenausprägung eine sogenannte Anfragevorschau⁴⁷ in Form eines numerischen Wertes angezeigt, die verdeutlicht, wie viele Ergebnisse bei einer Selektion dieser Facette bzw. Ausprägung im Ergebnis verbleiben würden. Zusätzlich zu dieser Vorschaufunktion wird damit auch ein Einblick in die Struktur eines Datenbestandes geboten, indem der Nutzer einen Überblick über die Verteilung der Facetten bzw. ihrer Ausprägungen auf die Daten erhält [GMPS00]. Somit kann der Nutzer durch jede Selektion sein Wissen erweitern und folglich seinen Informationsbedarf und damit seine Anfrage über die Zeit weiterentwickeln, konkretisieren und entsprechend gestalten. Außerdem ermöglicht die Ermittlung dieser Anfragevorschauen, dass Facetten aus dem weiteren Suchprozess ausgeschlossen werden können, falls deren Selektion in einer leeren Ergebnismenge ($\hat{=}$ Anfragevorschau ist 0) resultiert. Somit lässt sich ein navigierender Ansatz realisieren, der den Nutzer durch die Bereitstellung von nur zielführenden Selektionskriterien in gewisser Weise durch den Suchprozess leitet und mittels schrittweisem Filtern zu einer erfolgsversprechenden Ergebnismenge führt. Letztere ist dadurch charakterisiert, dass sie einerseits aus einer unsortierten Menge von Suchergebnissen besteht, weshalb die facettierte Suche zur Gruppe der Set Retrieval-Ansätze zu zählen ist. Zum anderen kann die darin enthaltene, benötigte Information auf unterschiedlichen Wegen erreicht werden [Den03], was die Tatsache von verschiedenen Vorgehensweisen der Nutzer beim Suchen in Betracht zieht.

Charakteristische Merkmale von Facetten

Wie bereits erwähnt, stellen Facetten Aspekte bzw. Eigenschaften einer Objektmenge dar und charakterisieren diese anhand verschiedener Ausprägungen. Dabei können letztere prinzipiell entweder in textueller Form, wie beispielsweise bei den Facetten **Hersteller** und **Farbe**, oder als Zahlenwerte (z. B. Facette **Baujahr**) vorliegen. Abhängig davon ergeben sich weitere Anforderungen, die insbesondere hinsichtlich der Visualisierung sowie der Selektionsmöglichkeiten auf der Anfrageseite und damit auch bei der Indexierung von Facetteninformationen zu berücksichtigen sind. Folglich ist zwischen verschiedenen Facettentypen zu unterscheiden, welche im Weiteren gemäß des einer Facette bzw. ihrer Semantik zugrunde liegenden Skalenniveaus [Vog05] als nominale, ordinale oder metrische Facetten bezeichnet werden. Dabei handelt es sich bei nominalen Facetten um Aspekte bzw. Eigenschaften, die durch textuelle Bezeichnungen oder Namen einer Nominalskala näher beschrieben werden. Diese dienen der reinen Kennzeichnung und Differenzierung der Ausprägungen, wobei keinerlei Rangfolge zwischen ihnen definiert werden kann. Als Beispiel kann die Facette **Hersteller** betrachtet werden, deren Ausprägungen **Audi**, **BMW**, **Mercedes** usw., gleichberechtigt nebeneinander stehen. Die Ausprägungen ordinaler Facetten, welche entweder textu-

⁴⁷Im Englischen entweder als *query preview* oder *facet count value* bezeichnet [Tun09, WR09].

ell oder als Zahlenwert vorliegen können, weisen hingegen eine Ordnung auf und sind somit untereinander anhand einer 'größer – kleiner', 'schlechter – besser' oder ähnlichen Relation vergleichbar (z. B. Ausprägungen **Euro6**, **Euro5**, **Euro4** usw. der Facette **Schadstoffklasse**). Folglich sollte eine derartige Reihenfolge auch bei der Anzeige der Facetten in der GUI mit berücksichtigt werden, um die Bedeutung der Facette zu verdeutlichen und damit die Verständlichkeit für den Nutzer zu erhöhen. Zusätzlich muss bei beiden Facettentypen gewährleistet sein, dass der Nutzer die Möglichkeit hat, eine oder auch mehrere Ausprägungen aus- bzw. auch abwählen zu können. Im Gegensatz dazu erfordern metrische Facetten nicht nur die Auswahlmöglichkeit bestimmter Ausprägungen. Da hier die Ausprägungen, wie beispielsweise bei der Facette **Preis**, aus reellen Zahlen bestehen, muss der Nutzer einerseits seine Anfrage auch auf Wertebereiche beziehen können. Aus diesem Grund sollten Intervallangaben bzw. -abfragen auf dieser Facette möglich sein. Zum anderen lassen sich metrische Ausprägungen ebenfalls in eine Rangfolge bringen und ermöglichen darüber hinaus die Ermittlung von Differenzen bzw. Abständen zwischen den Ausprägungen, die es bei deren Visualisierung mit zu berücksichtigen gilt. Zusätzlich zu dieser Typdefinition ist eine Facette außerdem durch weitere Merkmale charakterisiert, welche insbesondere für den Indexierungsprozess von Bedeutung sind. So ist abhängig von der Wertigkeit zwischen einwertigen und mehrwertigen Facetten zu differenzieren [HEE⁺02]. Während eine einwertige Facette für ein Objekt genau eine Ausprägung besitzt, kann eine mehrwertige Facette mehrere Ausprägungen zur Beschreibung des Objektes aufweisen. So ist beispielsweise ein Fahrzeug genau einem **Hersteller** zuzuordnen (= einwertige Facette), kann aber über mehrere **Ausstattungsmerkmale** verfügen (= mehrwertige Facette). Letztere ist zudem als hierarchische Facette gemäß Hearst et al. zu betrachten [HEE⁺02], da hier in Merkmale der **Innenausstattung** und Merkmale der **Außenausstattung** gruppiert werden kann. Flache Facetten hingegen weisen keine hierarchische Unterteilung ihrer Ausprägungen auf.

Beispielanwendungen und ihre Grenzen

Die im Rahmen der facettierten Suche ablaufende dynamische Anfrageformulierung und die damit verbundene hohe Interaktivität zwischen System und Nutzer wurden bereits 1994 von Ben Shneiderman in seiner Veröffentlichung „Dynamic Queries for Visual Information Seeking“ [Shn94] als notwendige Kriterien zukünftiger IRS im Umgang mit komplexen Suchsituationen gesehen. In dem von ihm mitentwickelten System FilmFinder konnten Nutzer daher visuell einen Datenbestand von Filmen durchstöbern, indem sie schrittweise mittels Schieberegler oder Markierungsfeldern nach Titel, Schauspieler, Schauspielerin, Regisseur, Länge und Bewertung filtern konnten [AS94]. Da diese Filterkriterien als flache Facetten einer Objektmenge *Filme* betrachtet werden können, wurde hier sozusagen ein erster Ansatz eines facettenbasierten Vorgehens beim Suchen realisiert. Allerdings war dieser noch sehr rudimentär ausgeprägt, da weder eine Schlagwortsuche unterstützt, noch Anfragevorschauen

für den Nutzer bereit gestellt wurden. Erst etwa zehn Jahre später wurden diese Aspekte in diversen Forschungsprojekten betrachtet, wobei das an der Universität von Berkeley von Hearst et al. entwickelte FLAMENCO-System⁴⁸ als Vorreiter der oben beschriebenen facettierten Suche gilt. Dabei handelt es sich bei FLAMENCO ($\hat{=}$ **FL**exible **in**formation **A**ccess using **ME**tadata in **NO**vel **CO**mbinations) um ein Open Source-System, welches beispielsweise zur Suche nach Bildern eingesetzt werden kann. Dazu werden diese auf Basis ihrer Metadaten hierarchischen Facetten zugeordnet und sind sowohl über eine Facettierte Navigation als auch eine Schlagwortsuche auffindbar [HEE⁺02, YSLH03]. Weitere Beispiele sind der von Marchionini et al. in [Mar06, CMO⁺07] vorgestellte und in [CM08] erweiterte Relation Browser zur Suche in Webseiten des US Bureau of Labor Statistics, welcher u. a. Anfragevorschauen zusätzlich in Form graphischer Balken und Facetten ähnlich zu Tag Clouds in einer Facet Cloud visualisiert, oder das von Bartolini in [Bar09] beschriebene SCENIQUE⁴⁹-System, das sich anstatt auf Metadaten auf die Inhalte von Bildern konzentriert und dazu neben auf Basis von Tags ermittelten, inhaltsbeschreibenden Facetten eine Facette zur Ähnlichkeitssuche anhand visueller Merkmale mit einbezieht. Des Weiteren findet man das Konzept einer facettierten Suche auch immer häufiger bei kommerziellen Anwendungen, wie beispielsweise bei der vom Unternehmen Endeca angebotenen ES-Lösung (vgl. Abbildung 2.7 in Kapitel 2.4) oder bei vielen Online-Plattformen wie eBay.com, Amazon.com oder mobile.de. Auch der Bereich der Softwareentwicklung beschäftigt sich zunehmend mit diesem Thema, so dass mittlerweile neben dem kommerziellen Softwarepaket Facetmap⁵⁰ mit Apache Solr⁵¹ auch ein Java-basierter Open Source Enterprise Search Server auf Basis des Open Source-Suchframeworks Apache Lucene⁵² zur Realisierung dieses Suchkonzeptes am Markt verfügbar ist.

Insgesamt betrachtet sind alle bisherigen Anwendungen für eine Suche im Bereich der technischen PE nicht ausreichend. Dies liegt zum einen darin begründet, dass traditionelle Systeme sich nur auf einen bestimmten Artefakttyp (z. B. Filme, Bilder, Autos, ...) konzentrieren und diesen hauptsächlich anhand beschreibender Daten charakterisieren. Zum anderen werden die zur Anfrageformulierung verwendbaren Facetten in textueller Form aufgelistet, was gerade bei einer großen Anzahl von Facetten mit eventuell vielen Ausprägungen schnell unübersichtlich wird und damit zu einer Informationsüberflutung des Nutzers — welche man eigentlich vermeiden wollte — führt. Auch die durch Facettenkombination gebildete Anfrage selbst wird primär textuell in Form sogenannter *history trails* [HEE⁺02] angezeigt, was insbesondere bei komplexen Anfragen ihre Nachvollziehbarkeit und nachträgliche Veränderung für den Nutzer erschwert. Zur Optimierung insbesondere der beiden letztgenannten Aspek-

⁴⁸<http://flamenco.berkeley.edu>

⁴⁹Semantic and ContEnt-based Image **Q**UErying

⁵⁰<http://www.facetmap.com>

⁵¹<http://lucene.apache.org/solr>

⁵²<http://lucene.apache.org/java/docs/index.html>

te wird daher auf das im Folgenden erläuterte Konzept der parallelen Koordinaten zurückgegriffen.

3.2.2. Parallele Koordinaten

Um hochdimensionale geometrische Daten darstellen zu können, begann Inselberg bereits 1977 mit der Entwicklung eines neuartigen Visualisierungskonzeptes, welches unter dem Begriff der parallelen Koordinaten bekannt wurde [Ins05]. Die Grundidee dieses Konzeptes besteht darin, eine Komplexitätsreduktion vorzunehmen, indem N -dimensionale Daten in den 2D-Raum projiziert werden [Ins85]. Dazu wird zunächst jede Dimension D_i als eine vertikale Koordinatenachse X_i fixer Höhe mit den Dimensionswerten als Achsenpunkte veranschaulicht. Die daraus resultierenden N Koordi-

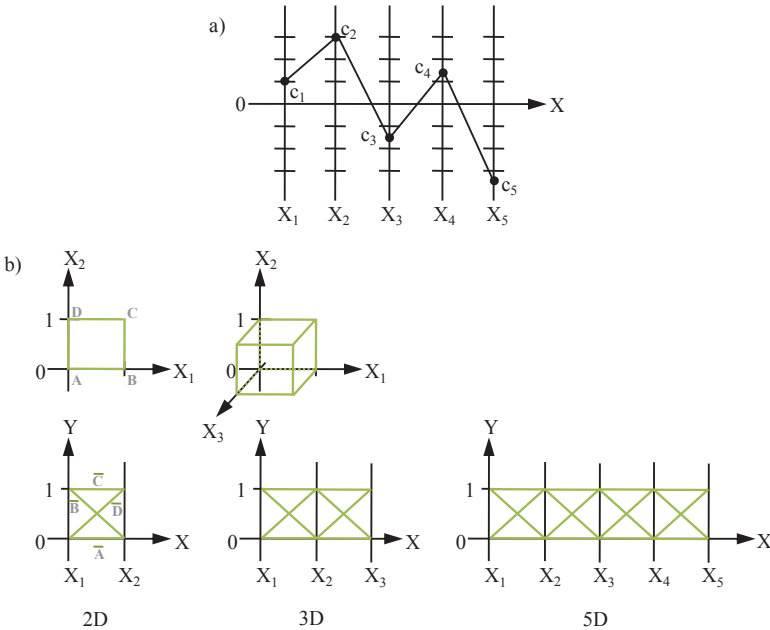


Abbildung 3.1. — a) Darstellung des 5D-Punktes $C = (c_1, c_2, c_3, c_4, c_5)$ in parallelen Koordinaten; b) Kartesische (oben) und parallele Koordinatendarstellung (unten) eines Quadrats (links), eines Würfels in 3D (mittig) und eines Würfels in 5D (rechts), wobei letzterer in kartesischen Koordinaten nicht mehr darstellbar ist (nach [Ins05, S. 153] und [Ins85, S. 90])

natenachsen werden anschließend in äquidistantem Abstand parallel zueinander angeordnet, wodurch ein paralleles Koordinatensystem entsteht. Darin ist es möglich, jeden N -dimensionalen Punkt $P = (p_1, p_2, \dots, p_n)$ durch einen Polygonzug abzubilden, welcher die den Dimensionswerten p_1 bis p_n entsprechenden Punkte P_1 bis P_n auf den Koordinatenachsen X_1 bis X_n miteinander verbindet. Abbildung 3.1 a) zeigt dies für einen fünfdimensionalen Punkt $C = (c_1, c_2, c_3, c_4, c_5)$, wobei zu erkennen ist, dass auf diese Art und Weise relativ einfach jede beliebig große Anzahl von Dimensionen visualisiert werden kann. Des Weiteren macht diese Form der Visualisierung auch eine Analyse mehrdimensionaler Daten möglich, da aus den Polygonzügen Rückschlüsse auf die Daten und ihre Strukturen gezogen werden können. Wie Abbildung 3.1 b) veranschaulicht, kann die rechte untere Darstellung aufgrund der Strukturen auf der linken Seite und in der Mitte als 5D-Würfel interpretiert werden.

3.2.3. Konzept der multiplen Ebenen

Die während des Suchprozesses erforderliche Berücksichtigung von Artefakten unterschiedlichen Typs (Produkte, Dokumente, Personen, ...) sowie deren Beziehungen zueinander kann durch Anwendung des allgemein aus dem Bereich der Modellierung bekannten Konzeptes der multiplen Ebenen realisiert werden. Dazu sind mehrere Ebenen zu definieren, die gegebenenfalls über Beziehungen miteinander verbunden sind. Für den in der vorliegenden Arbeit betrachteten Anwendungsfall der technischen PE wird folglich jeder Artefakttyp durch eine Ebene repräsentiert, die im Weiteren als Artefakttypenebene bezeichnet wird und alle zu diesem Typ gehörigen Artefakte enthält. Beziehungen zwischen einzelnen Artefakten können dabei sowohl zwischen als auch innerhalb von Artefakttypenebenen bestehen und stellen v. a. für die Realisierung einer Suche anhand von Assoziationen nützliche Hilfsmittel dar. Abbildung 3.2 zeigt dieses Ebenenkonzept exemplarisch anhand von vier Artefakttypenebenen. Dabei wird zwischen der Personen-, der Dokument-, der Produkt- und der Werkstoffebene differenziert, welche die zugehörigen Artefakte, dargestellt durch farbige Kreise auf den Ebenen, enthalten. Die gerichteten Pfeile in Abbildung 3.2 illustrieren dabei beispielhafte Artefaktbeziehungen, wobei ein durchgezogener Pfeil eine Verbindung innerhalb einer Ebene kennzeichnet. Letztere sind beispielsweise auf der Produktebene zu finden, da Produkte über 'besteht-aus'-Beziehungen mit anderen Produkten im Sinne von Unterkomponenten verbunden sein können. Folglich kann diese Relation im Rahmen einer Suche genutzt werden, um Fragen wie *Aus welchen Unterkomponenten besteht ein Produkt?* zu befriedigen. Ist ein Pfeil hingegen gestrichelt, symbolisiert er eine Artefaktbeziehung, die von einer Artefakttypenebene auf eine andere Artefakttypenebene verweist. Gemäß Abbildung 3.2 lassen sich somit Suchanfragen wie *Wer hat das Dokument erstellt?*, *Welches Produkt wird im Dokument beschrieben?* oder *Aus welchem Werkstoff wurde das Produkt hergestellt?* bearbeiten, indem generell von den Ausgangsartefakten einer Beziehung zu dessen Zielartefakten gewechselt wird.

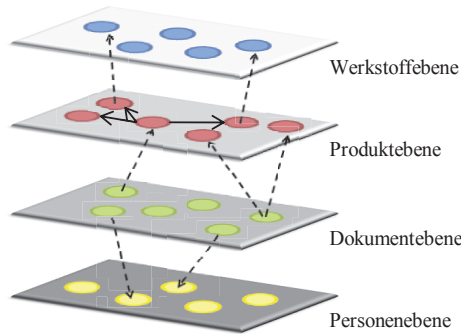


Abbildung 3.2. — Konzept der multiplen Ebenen für die PE mit vier Artefakttypenebenen und exemplarischen Artefaktbeziehungen sowohl innerhalb einer Ebene (durchgezogene Pfeile) als auch zwischen Ebenen (gestrichelte Pfeile)

3.2.4. Konzept des Ranking

Ranked Retrieval-Verfahren (vgl. Kapitel 1.2.2) tragen gemäß Salton wesentlich zur Erhöhung der Benutzerverständlichkeit und -zufriedenheit bei [SM87], weshalb diese insbesondere in mengenorientierte Ansätze, wie dem der facettierten Suche, eingebunden werden sollten. Da eine der in Kapitel 3.1 geschilderten Anforderungen in der Unterstützung einer Ähnlichkeitssuche besteht, kann dies implizit durch deren Integration erreicht werden. Aus diesem Grund wird im folgenden Abschnitt das Suchparadigma der Ähnlichkeitssuche näher betrachtet. Darüber hinaus sollten jedoch auch bei einer reinen Filterung über Facetten Rangfolgen erzeugt werden können, was durch die Bereitstellung sogenannter Präferenzfunktionen erreicht wird. Hierbei handelt es sich allgemein um Funktionen, mit denen der Nutzer seine Präferenz für bestimmte Ausprägungen oder Wertebereiche einer Facette definieren und dadurch ein Ranking der Ergebnisse erzeugen oder beeinflussen kann. In Kapitel 3.3.2 wird diese Funktionalität am Beispiel der GUI noch konkreter illustriert.

Paradigma der Ähnlichkeitssuche

Wie bereits in den vorhergehenden Ausführungen verdeutlicht, werden Ähnlichkeitssuchen zum Auffinden ähnlicher Artefakte eingesetzt. Dabei beschreibt der Begriff der Ähnlichkeit prinzipiell eine Relation, die den Grad der Übereinstimmung von Artefakten beschreibt [Hub84, Wik10a]. Während allerdings bei einer vollständigen Übereinstimmung von einer Gleichheit bzw. Identität der Artefakte zu sprechen ist, sind ähnliche Artefakte durch eine nur teilweise Übereinstimmung charakterisiert. Das

bedeutet, dass ähnliche Artefakte bei einer vergleichenden Betrachtung lediglich gemeinsame Eigenschaften aufweisen. An dieser Stelle ist zu berücksichtigen, dass ein einziger, allgemeingültiger Ähnlichkeitsbegriff nicht existiert; vielmehr bezieht sich Ähnlichkeit immer auf einen bestimmten, der Situation bzw. dem Anwendungsgebiet entsprechenden Aspekt, so dass genau genommen zahlreiche Ähnlichkeiten zu differenzieren sind. Als Beispiele hinsichtlich der in der vorliegenden Arbeit betrachteten Problemstellung sind die textuelle Ähnlichkeit von Dokumenten, die funktionelle, geometrische oder fertigungstechnische Ähnlichkeit von Produkten oder auch die physikalische oder chemische Ähnlichkeit von Werkstoffen zu nennen. Aber auch innerhalb jeder einzelnen dieser Ähnlichkeiten finden sich wiederum diverse Ausprägungen. So kann sich eine geometrische Ähnlichkeit z. B. auf geometrische Abmessungen (Länge, Breite, Höhe, ...), auf die globale Gestalt oder auch nur partiell auf einzelne geometrische Bereiche beziehen [Ber97]. Demzufolge bedingt eine Ähnlichkeitsermittlung zwischen Artefakten immer erst eine Definition von eindeutigen, den jeweils betrachteten Aspekt beschreibenden Eigenschaften. Diese werden im IR als Features bezeichnet [Sch06] und dienen schließlich dem tatsächlichen Vergleich zwischen den Artefakten. Dazu sind sowohl das Vergleichsartefakt, welches üblicherweise als QbE oder QbT an das IRS übergeben wird (vgl. Kapitel 2.2.1 und 2.2.2), als auch die zu vergleichenden Artefakte anhand dieser Features zu beschreiben. Mittels der daraus resultierenden Feature-Repräsentationen kann anschließend ein Ranking von Ähnlichkeitswerten (auch als *retrieval status values* bezeichnet) und damit eine absteigend nach diesen Werten geordnete Liste der Artefakte ermittelt werden. Hierzu wird eine formale Definition des Ähnlichkeitsbegriffes in Form einer mathematischen Funktion $f : X \times X \mapsto \mathbb{R}$ benötigt, die jeweils zwei Feature-Repräsentationen miteinander vergleicht und auf eine reelle Zahl, welche üblicherweise im Intervall $[0; 1]$ liegt, abbildet [Ber97, Sch06]. Diese Funktion kann dabei entweder als Ähnlichkeitsmaß, bei der ein Wert von 0 die minimale und ein Wert von 1 die maximale Ähnlichkeit kennzeichnet (z. B. Kosinusmaß), oder als Distanzmaß mit umgekehrter Zuordnung — d. h. Distanz von 0 bedeutet maximale Ähnlichkeit (z. B. Minkowski-Distanzen) — definiert werden. Da die Beantwortung der Frage, welches Maß für welchen Anwendungsfall geeignet ist, u. a. von den in den Feature-Repräsentationen verwendeten Datentypen sowie den geforderten, zu unterstützenden Eigenschaften (Positivität, Identität, Symmetrie, Dreiecksungleichung, Transformationsinvarianz, ...) abhängt und eine ausführliche Betrachtung dieser den Rahmen der vorliegenden Arbeit sprengen würde, wird für weitere Informationen hierzu auf die entsprechende Literatur (z. B. [Vel01, Sch06, ZADB06]) verwiesen.

Zusammenfassend lässt sich das prinzipielle Vorgehen einer Ähnlichkeitssuche in vier Schritten beschreiben:

1. Definition der Features
2. Erstellung der Feature-Repräsentationen

3. Festlegung der zu verwendenden Ähnlichkeits- bzw. Distanzfunktion
4. Ermittlung des Rankings

Obwohl Letzteres eine Reihenfolge von Artefakten sortiert nach ihren Ähnlichkeitswerten ermittelt, ist zu berücksichtigen, dass eine Ähnlichkeitsbewertung letztendlich immer auch durch einen subjektiven Aspekt geprägt ist und dabei von der Situation, den Erfahrungen und dem Hintergrundwissen des Nutzers abhängt.

3.3. Integration der Basiskonzepte in ein übergreifendes Gesamtsystem — das LFRP-Framework

Die Kombination der vorgestellten Basiskonzepte zu einem übergreifenden IRS resultiert in der Konzeption und prototypischen Umsetzung des sogenannten LFRP-Frameworks. LFRP steht dabei als Akronym für *Multi-Layer Faceted Search with Ranking using Parallel Coordinates* und repräsentiert somit die vier Grundbausteine, auf die sich das System stützt. Für eine ganzheitliche Betrachtung wird im folgenden Kapitel 3.3.1 vorerst ein allgemeiner Überblick über die Architektur des Gesamtsystems gegeben. Obwohl der Retrievalprozess des Systems nicht im Fokus der vorliegenden Arbeit liegt, wird dieser aus Gründen der Vollständigkeit und zum besseren Verständnis in Kapitel 3.3.2 betrachtet. Dazu werden sowohl die GUI als auch die Suchkomponente kurz erläutert⁵³. Im Anschluss widmet sich Kapitel 3.3.3 dem eigentlichen Schwerpunkt dieser Arbeit und stellt das für den Indexierungsprozess entwickelte Indexierungsframework vor.

3.3.1. Architektur des Gesamtsystems

Der Begriff der Architektur im Bereich des IM beschreibt gemäß Heinrich und Burgholzer „(...) die Art und Anzahl der Komponenten der Informations-Infrastruktur oder von Teilen der Informations-Infrastruktur (z. B. eines Anwendungssystems) und die Beziehungszusammenhänge zwischen den Komponenten.“ [HB88, S. 95] Demzufolge stellt Abbildung 3.3 die für das LFRP-Framework notwendigen Subsysteme dar und veranschaulicht deren Zusammenarbeit anhand des ADK-Strukturmodells [FS08]. Aus Gründen der Übersichtlichkeit sind sämtliche Subsysteme zunächst als *Black Boxes* dargestellt; sie werden in den weiteren Kapiteln detaillierter betrachtet.

⁵³Detaillierte Informationen hierzu können der Dissertation von Raiko Eckstein am gleichnamigen Lehrstuhl entnommen werden.

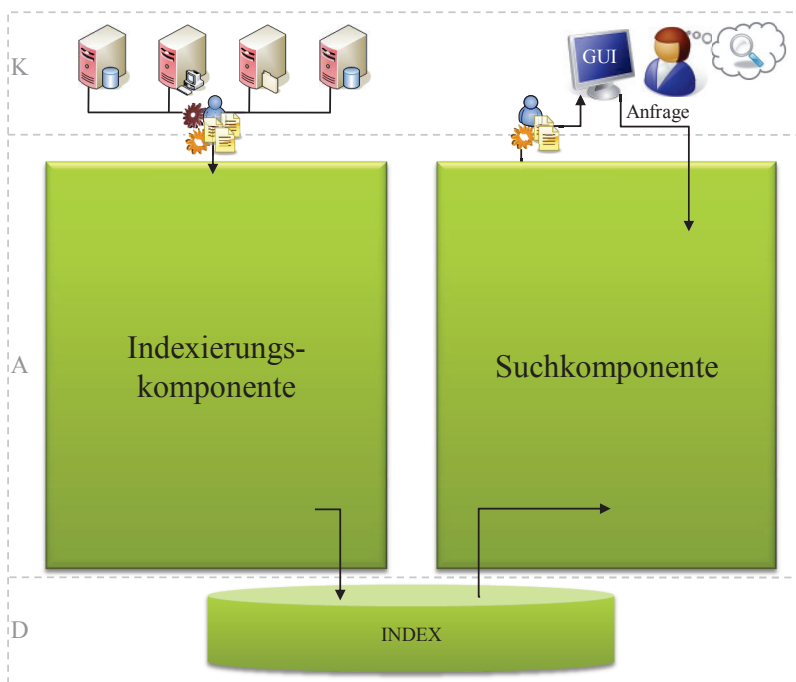


Abbildung 3.3. — Allgemeine Architektur des LFRP-Frameworks nach dem ADK-Strukturmodell

Insgesamt folgt die hier illustrierte Architektur dem allgemeinen Aufbau von IRS, welche auf der obersten Schicht — der Kommunikationsschicht (vgl. den mit *K* markierten grauen Bereich in Abbildung 3.3) — eine GUI zur Kommunikation mit personellen AT erfordern. Über diese können Suchanfragen an das System gestellt und Suchergebnisse präsentiert werden. Darüber hinaus sind die existierenden Informationen (Artefakte) zu erfassen, welche im vorliegenden Fall in unterschiedlichen IT-Systemen eines Unternehmens enthalten sind. Hierfür sind Kommunikationsmechanismen zu diesen maschinellen AT auf dieser obersten Schicht zu realisieren, welche die erfassten Artefakte an eine auf der Schicht der Anwendungsfunktionen (vgl. den mit *A* markierten grauen Bereich in Abbildung 3.3) vorhandene Indexierungskomponente für eine geeignete Aufbereitung übergeben. Zusätzlich wird auf der Anwendungsschicht eine Suchkomponente benötigt, welche die zu einer, vom Nutzer über die GUI eingegebene Anfrage relevanten Suchergebnisse ermittelt und dem Nutzer in der GUI

präsentiert. Die unterste Schicht zur Datenverwaltung (vgl. den mit *D* markierten grauen Bereich in Abbildung 3.3) muss schließlich einen Index enthalten, der die von der Indexierungskomponente entsprechend aufbereiteten Artefakte speichert und für die Suchkomponente zur Verfügung stellt.

3.3.2. Retrievalprozess des LFRP-Frameworks

Zur Suche nach Informationen wird dem Nutzer im LFRP-Framework die in Abbildung 3.4 dargestellte GUI zur Verfügung gestellt. Diese ist insgesamt in drei horizontale Bereiche unterteilt, welche jeweils unterschiedliche Funktionalitäten des Systems übernehmen.

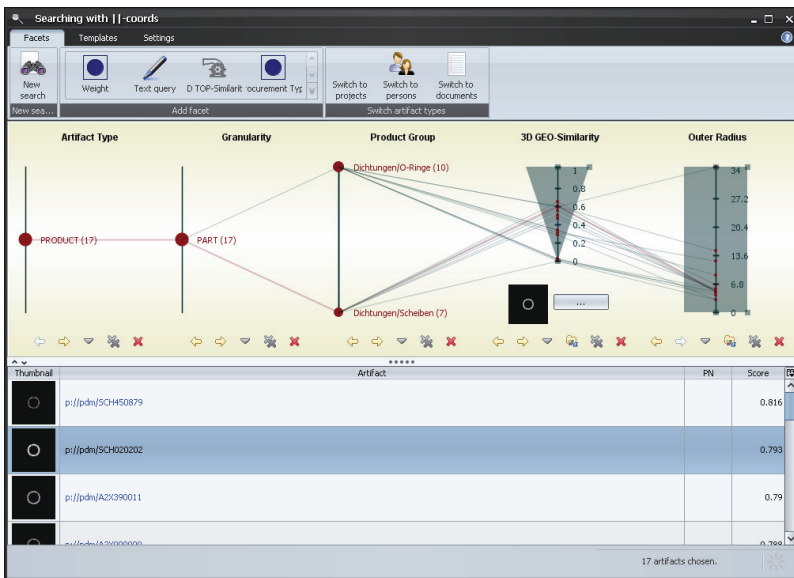


Abbildung 3.4. — Graphische Benutzeroberfläche des LFRP-Frameworks

Die im oberen Bereich enthaltenen Menüs realisieren dabei die Steuerung des Systems, wobei insbesondere das Menü *Facets* zur Charakterisierung von Suchanfragen dient. Hier kann der Nutzer mittels des Menüpunktes *New search* eine neue Suche starten oder über *Switch artifact type* seine Suche auf eine andere Artefakttypenebene (vgl. Kapitel 3.2.3) verlagern. Des Weiteren wird dem Nutzer unter *Add facet* eine

dynamisch aktualisierte Liste von wählbaren Facetten angeboten, aus welcher er die für sein Informationsbedürfnis relevanten Facetten schrittweise auswählen und zu einer Suchanfrage kombinieren kann. Die ausgewählten Facetten werden schließlich im mittleren Bereich der GUI als parallele Koordinaten visualisiert, wobei jede Facette als eine vertikale Achse mit ihren Ausprägungen und zugehörigen Anfragevorschauen als Achsenpunkte dargestellt wird. Auf diese Weise lässt sich eine beliebige Anzahl von Facetten im Zusammenhang abbilden, was die Veranschaulichung einer komplexen, aus mehreren Einzelanfragen bestehenden Anfrage in einer übersichtlichen Form ermöglicht und wesentlich zu einer verbesserten Nachvollziehbarkeit für den Nutzer beiträgt.

Abbildung 3.4 zeigt eine aus mehreren Einzelanfragen zusammen gesetzte Beispielanfrage nach Produkten unter Verwendung der Facetten **Artifact Type**, **Granularity**, **Product Group**, **3D GEO-Similarity** und **Outer Radius**. Diese Anfrage besteht einerseits aus sogenannten *Attributfacetten*, welche die Auswahl einer oder mehrerer Facettenausprägungen zulassen und folglich als reine Filterkriterien dienen (z. B. Facette **Artifact Type**). Da sie jedoch kein Ranking der Ergebnismenge, sondern nur eine Einschränkung dieser bewirken, können sie durch die bereits in Kapitel 3.2.4 erwähnten Präferenzfunktionen ergänzt werden. Abbildung 3.4 zeigt eine solche Präferenzfunktion für die Facette **Outer Radius**, bei der im Beispiel jede Facettenausprägung mit dem gleichen Präferenzwert vom Nutzer belegt ist.

Andererseits kann eine Suchanfrage neben diesen Attributfacetten aber auch sogenannte *Ähnlichkeitsfacetten* enthalten. Diese ermöglichen dem Nutzer die Durchführung einer Ähnlichkeitssuche, wie es in Abbildung 3.4 anhand der Facette **3D GEO-Similarity** gezeigt wird. Hier kann der Nutzer durch Hochladen einer 3D-Geometriebeschreibung eines Produktes nach Produkten suchen, die hinsichtlich ihrer dreidimensionalen Geometrie ähnlich zum Anfrageprodukt sind. Äquivalent dazu lassen sich Ähnlichkeitsfacetten beispielsweise für eine 2D-Geometrieähnlichkeit oder für eine Textähnlichkeit (hier mit der Möglichkeit, Text in ein Eingabefeld einzugeben) anwenden. Im Allgemeinen gilt dabei, dass Ähnlichkeitsfacetten ebenfalls als Achsen dargestellt werden. Im Unterschied zu Attributfacetten jedoch, repräsentieren ihre Achsenpunkte die aus dem Ähnlichkeitsvergleich resultierenden Ähnlichkeitswerte, weshalb sie defaultmäßig mit einer Präferenzfunktion überlagert werden. Insgesamt betrachtet führt jede Aktion des Nutzers, d. h. jedes Hinzufügen oder Entfernen einer Facette bzw. Ausprägung, zu einer Änderung der Ergebnismenge, welche dem Nutzer auf zwei Arten präsentiert wird. Zum einen wird jedes einzelne Ergebnis in Form eines Polygonzuges zwischen den parallelen Koordinaten visualisiert, welcher durch Auswahl des entsprechenden Ergebnisses hervorgehoben wird (siehe rot markierter Polygonzug in Abbildung 3.4). Dies eröffnet nicht nur Einblicke in die Charakteristika der Ergebnisse, sondern ermöglicht auch eine detailliertere Analyse des Datenbestandes. Durch Steuerelemente unter den Achsen können diese u. a. in ihrer Reihenfolge verschoben werden, was Abhängigkeiten bzw. Beziehungen zwischen den Achsen und

damit den Facetten verdeutlicht. Des Weiteren kann der Nutzer über diese Steuerelemente die einzelnen Facetten individuell gewichten, was eine Anpassung der Anfrage an die jeweils konkrete Situation des Nutzers ermöglicht. Zwar könnte diese situationsspezifische Facettengewichtung prinzipiell auch über lernende Verfahren oder mit Hilfe von Expertenwissen sozusagen a priori ermittelt werden; diese automatische Gewichtung jedoch würde nicht nur große Mengen an Trainingsdaten erfordern, sondern würde auch niemals der Vielfalt an Informationsbedürfnissen in diesem Anwendungsbereich gerecht werden. Aus diesem Grund überlässt das LFRP-Framework bewusst dem Nutzer die Gewichtung von Facetten, um so sein konkretes Informationsbedürfnis möglichst gut verstehen und befriedigen zu können.

Zusätzlich zur Visualisierung der Ergebnisse in Form von Polygonzügen, werden diese im unteren Bereich der GUI zusammen mit einer etwas detaillierteren Beschreibung aufgelistet, von wo der Nutzer schließlich direkt oder über Drittsysteme auf die tatsächlichen Informationen zugreifen kann. Um jedoch die Akzeptanz des Systems beim Nutzer zu erhöhen, sollte hierbei stets darauf geachtet werden, dass ein IRS immer auch erklären sollte, warum und wie die Ergebnisse zur Anfrage passen [MK97]. Dies wird im LFRP-Framework prinzipiell durch die bereits beschriebenen Polygonzüge erreicht, die für jedes Ergebnis die zugehörigen Facettenausprägungen verdeutlichen. Um die Relevanz von Ergebnissen zur Anfrage allerdings noch konkreter darzustellen, sollte man zusätzliche Hilfsmittel integrieren, von denen im Folgenden einige beispielhaft und ohne Anspruch auf Vollständigkeit vorgestellt werden.

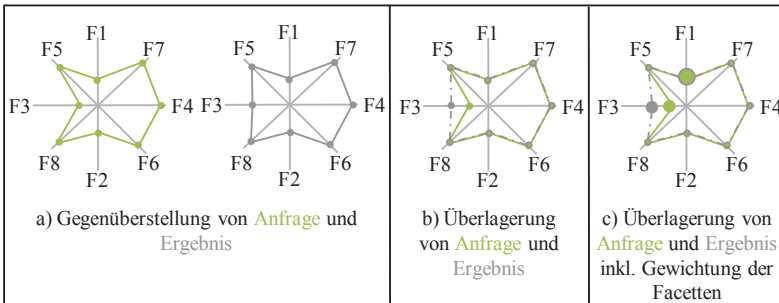


Abbildung 3.5. — Relevanzklärung durch a) Gegenüberstellung oder b) Überlagerung von Sternendiagrammen, in welche zusätzlich c) die Gewichtung einzelner Facetten integriert werden kann

Ausgehend vom Konzept der parallelen Koordinaten besteht eine Möglichkeit der Relevanzklärung im Einsatz sogenannter Sternendiagramme [Spe01]. Wie Abbildung 3.5 zeigt, werden dabei die Facetten nicht parallel sondern sternförmig zueinander angeordnet und die einzelnen Facettenausprägungen miteinander verbunden. Durch eine

Gegenüberstellung (siehe Abbildung 3.5 links) oder eine Überlagerung (siehe Abbildung 3.5 in der Mitte) der Sterndiagramme von Anfrage und Ergebnissen könnte so die Relevanzbeurteilung des Systems für den Nutzer nachvollziehbar visualisiert werden. Von Vorteil hierbei ist, dass keine zusätzlichen Informationen benötigt und bei der Indexierung zu erfassen sind; vielmehr sind die bereits indexierten Daten lediglich in einer anderen Form in der GUI aufzubereiten. Des Weiteren lassen sich sowohl bei Sterndiagrammen als auch bei Polygonzügen die vom Nutzer individuell festgelegten Gewichtungen der einzelnen Facetten gut berücksichtigen, indem man die Flächen der eine Facettenausprägung kennzeichnenden Kreise proportional zum jeweiligen Facettengewicht setzt. Damit lassen sich folglich höher gewichtete Facetten mit größeren Kreisflächen und niedriger gewichtete Facetten mit kleineren Kreisflächen darstellen (vgl. Abbildung 3.5 rechts). Ein weiteres, häufig genutztes Mittel zur Erklärung ist die explizite Darstellung des Relevanzwertes eines Ergebnisses, wie er unter der Spalte *Score* in der Ergebnisliste von Abbildung 3.4 enthalten ist. Dieser Relevanzwert wird im Rahmen des Retrievalprozesses von der Suchkomponente ermittelt und als Dezimalzahl oder Prozentwert angezeigt. Allerdings ist diese Wertangabe für den Nutzer nicht unbedingt hilfreich, da sich der Relevanzwert in den meisten Fällen (u. a. bei der Verwendung mehrerer Ähnlichkeitsfacetten) aus mehreren Relevanzwerten zusammensetzt. Um diesbezüglich die Transparenz für den Nutzer zu erhöhen, können beispielsweise Balkendiagramme in die Ergebnisliste integriert werden, die, wie in Abbildung 3.6 für die Ergebnisse zwei und drei gezeigt, nicht nur die einzelnen Relevanzwerte, sondern auch deren Gewichtung in der Relevanzberechnung beispielsweise anhand entsprechender Balkenbreiten oder spezieller Farben für den Nutzer verdeutlichen. Eine andere Möglichkeit zur Relevanzklärung ist die Verwendung von Textausschnitten, sogenannten *Snippets*, wie sie aus dem Bereich der Web-suchmaschinen bekannt sind (vgl. Ergebnisse eins, vier und fünf in Abbildung 3.6). Sie sind im vorliegenden Anwendungsfall speziell für den Artefakttyp *Dokument* von Bedeutung, da mit Hilfe von Ausschnitten aus den Dokumentinhalten und Hervorhebung von Anfragetermen dem Nutzer hier plausibel aufgezeigt werden kann, weshalb ein Dokument als relevant für die jeweilige Anfrage erachtet wird.

Neben diesen eher relativ unkompliziert umzusetzenden Erklärungskomponenten gibt es aber auch noch andere Hilfsmittel, die explizit gesonderte Indexierungsanforderungen mit sich bringen und demnach bereits bei der Konzeption des Indexierungsframeworks mit zu berücksichtigen sind. Diesbezüglich ist insbesondere die Verwendung von Bildern erwähnenswert, welche prinzipiell in zwei Formen zur Erklärung relevanter Artefakte genutzt werden können. Die erste Form ist die Darstellung von Vorschaubildern, sogenannten *Thumbnails*, welche v. a. bei Produkten und Personen sehr hilfreich sind (vgl. hierzu die Spalte *Thumbnail* in der Ergebnisliste von Abbildung 3.4). Bei der zweiten Form handelt es sich beispielsweise um Grauwert- oder Silhouettebilder, deren Basisdaten bei Nutzung spezieller Konzepte für 2D- oder 3D-Geometrien im Rahmen des Repräsentationserstellungsprozesses benötigt und erzeugt

Thumbnail	Artifact			Score
	D:\Lastenheft.doc	... O-Ringe müssen folgende Eigenschaften besitzen ... sind O-Ringe von XYZ GmbH zu verwenden ...		1
	P:\PDM\A2X342000		Text Similarity Outer Radius	0.974
	P:\PDM\SCH001001		Text Similarity Outer Radius	0.974
	D:\Montageanleitung.doc	... den O-Ring zwischen den beiden Subsystemen anbringen ... mit einem O-Ring abdichten ...		0.974
	D:\Betriebsdaten.doc	... Außendurchmesser O-Ring liegt zwischen ... Zylinderkopf inklusive fünf O-Ringen ...		0.951

Abbildung 3.6. — Erweiterung der Ergebnisliste mit Snippets (Ergebnis 1, 4 und 5) und Balkendiagrammen (Ergebnis 2 und 3) zur besseren Erklärung der Relevanz dieser Ergebnisse

werden (vgl. hierzu beispielhaft die Veranschaulichung der ersten beiden Konzepte in Tabelle 5.3). Unabhängig von der konkreten Form jedoch erfordert die Nutzung von Bildern immer einen zusätzlichen Zwischenschritt im Indexierungsprozess, in dem die Bilder — entweder aus den ohnehin ermittelten Basisdaten oder aus den gegebenen Geometriedaten — zuerst generiert und anschließend als zusätzliche Information für das jeweilige Artefakt gespeichert werden.

Für die Realisierung sämtlicher Funktionalitäten im Rahmen des Retrievalprozesses verfügt das LFRP-Framework schließlich über eine Suchkomponente, deren Arbeitsweise aus Gründen der Vollständigkeit zwar in Abbildung 3.7 grob skizziert ist, die jedoch nicht im Detail erläutert werden soll. Für konkrete Informationen diesbezüglich wird u. a. auf Veröffentlichungen wie [EHW09, EHW10] verwiesen. An dieser Stelle soll lediglich darauf aufmerksam gemacht werden, dass bei der Konzeption der Indexierungskomponente beispielsweise auch das Modul zur Rechteprüfung Berücksichtigung finden muss. Schließlich muss anhand der Berechtigungen eines Nutzers geprüft werden, welche Ergebnisse dieser sehen darf und welche nicht. Diese Prüfung könnte grundsätzlich dadurch realisiert werden, dass zu jedem Artefakt alle Nutzer mit ihren zugehörigen Rechten gespeichert werden. Da dieses Vorgehen allerdings sehr aufwändig und im Fall von Rechteänderungen nicht trivial ist, nutzt das LFRP-Framework eine andere Variante: Für jedes Artefakt sind dessen Informationsquelle sowie seine zugehörige ID aus dieser Informationsquelle zu indexieren (siehe auch Kapitel 4.4). Damit können nach der Ermittlung der Ergebnisse anhand dieser Informationen für jedes Ergebnis die Berechtigungen für den jeweiligen Nutzer überprüft werden und eventuell Ergebnisse, die nicht sichtbar sein dürfen, aus der Ergebnismenge entfernt bzw. Ergebnisse, die nicht zugreifbar sein dürfen, in der Ergebnismenge eingeschränkt dargestellt werden.

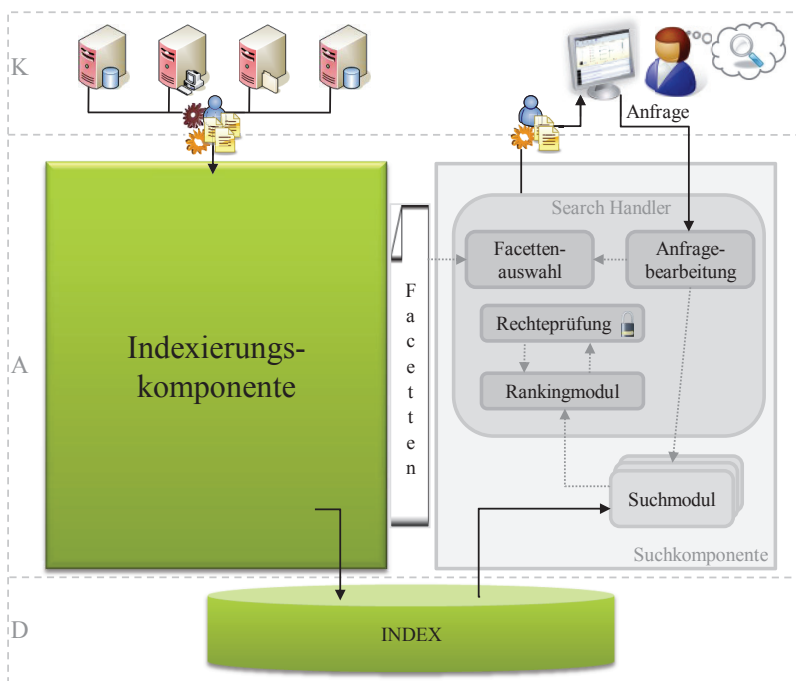


Abbildung 3.7. — Architektur der Suchkomponente des LFRP-Frameworks

3.3.3. Indexierungsprozess des LFRP-Frameworks

Voraussetzung für die Realisierung der im vorhergehenden Kapitel vorgestellten Suchfunktionalitäten ist eine geeignete Aufbereitung der bereits im Unternehmen vorhandenen Informationen. Um dies zu gewährleisten, wird ein Framework für die Indexierungskomponente vorgestellt, welches den Indexierungsvorgang gemäß Abbildung 3.8 in vier Schritte unterteilt. Als erstes sind dabei die im Unternehmen vorhandenen und zu indexierenden Informationen in das IRS zu importieren (Schritt *Datenimport*), wobei abhängig von der Art der Informationsquelle verschiedene Daten zu berücksichtigen sind. Zum einen handelt es sich hierbei um Artefakte wie Produkte, Werkstoffe und andere, welche mittels beschreibender Daten (= Metadaten) in den IT-Systemen (z. B. PDM, ERP, SRM, ...) verwaltet werden. Zum anderen finden sich auch in diesen, aber v. a. in Dateiverwaltungssystemen archivierte Dokumente, die ebenfalls

indexiert werden sollten. Demzufolge muss es möglich sein, sowohl Artefakte mit ihren zugehörigen Metadaten als auch Dokumente an die Indexierungskomponente zu übergeben.

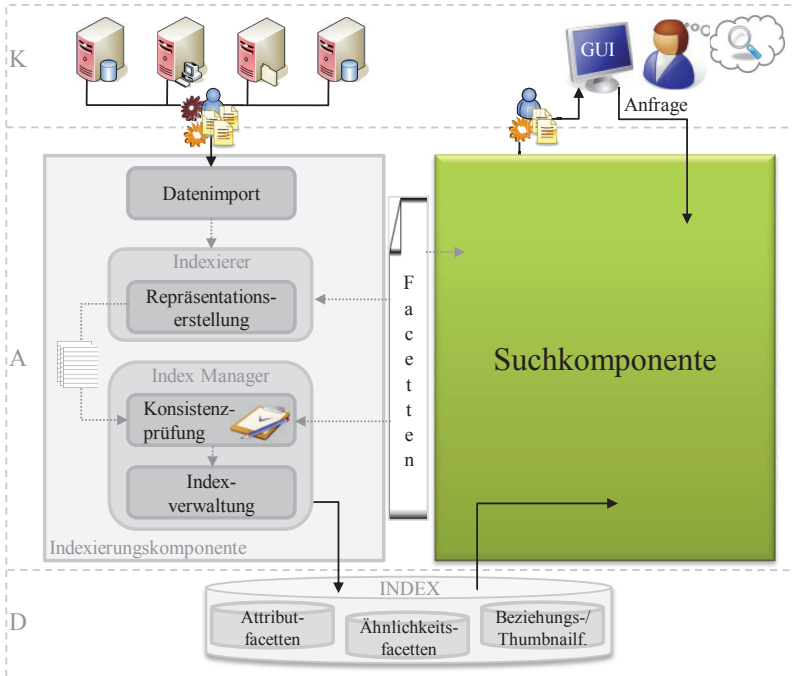


Abbildung 3.8. — Indexierungskomponente und Index des LFRP-Frameworks: Vereinfachte Darstellung des Indexierungsvorgangs

Hierzu stellt die Indexierungskomponente zwei Schnittstellenmethoden zur Verfügung, die zur Realisierung zweier Indexierungsmodi dienen. Mit der ersten Methode wird aufgrund der in Unternehmen üblicherweise großen Datenmengen ein sogenannter *Pullmodus* angeboten, über den sich die Indexierungskomponente sozusagen selbst die Daten aus den Drittsystemen holt. Dazu wird entweder in festen Zeitabständen (z. B. alle 24 Stunden) oder zu fixen Zeitpunkten (z. B. jede Nacht) — was je nach Verarbeitungs- und Kommunikationsauslastung der Drittsysteme im Unternehmen mittels eines Schedulers konfiguriert werden kann — ein Indexierungsvorgang gestartet, der alle seit der letzten Indexierung neu hinzugefügten oder aktualisierten Daten importiert und verarbeitet. Die zweite Methode wird zur Berücksichtigung wichtiger

Änderungen benötigt, indem über einen sogenannten *Pushmodus* ein Drittsystem von sich aus Aktualisierungen, wie beispielsweise wichtige Änderungen an Produktstammdaten, zeitnah an die Indexierungskomponente melden kann. In beiden Fällen wird die Übergabe einer Containerdatei vom Drittsystem gefordert, deren Aufbau in Kapitel 4.4 genauer erläutert wird. Die Übergabe und damit der Aufruf der jeweiligen Methode bei der Indexierungskomponente bzw. beim Drittsystem kann dabei über eine lose Kopplung mittels geeigneter Middleware-Technologien⁵⁴ hergestellt werden. So liefern beispielsweise ERP-Systeme gemäß Gronau von Haus aus eine Middleware mit aus, über die andere Programme mittels Remote Procedure Calls aufgerufen werden können [Gro04]. Welche Technologien jedoch im konkreten Anwendungsfall einzusetzen sind, hängt von den unternehmerischen Gegebenheiten, d. h. von den anzubindenden Drittsystemen und den von ihnen unterstützten Kommunikationswegen ab. Folglich kann und soll in der vorliegenden Arbeit keine allgemeingültige Lösung für diese Kommunikation gegeben werden.

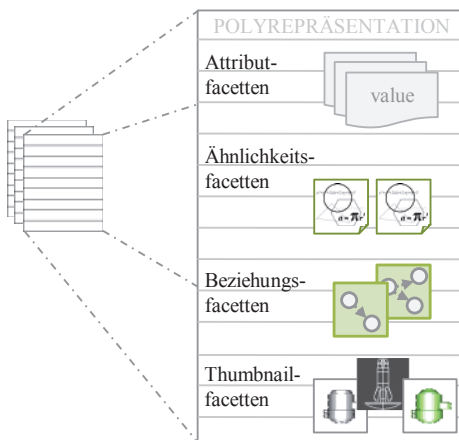


Abbildung 3.9. — Artefaktbeschreibung als Polyrepräsentation

Sobald die Informationen im IRS verfügbar sind, sind sie für die Suche von einem *Indexierer* entsprechend aufzubereiten (Schritt *Repräsentationserstellung*). Dazu werden die eingehenden Informationen zunächst anhand ihres Artefakttyps differenziert, so dass anschließend jedes einzelne Artefakt separat verarbeitet und unter Verwendung

⁵⁴Der Begriff Middleware bezeichnet gemäß Leszinski eine Softwareschicht, die auf Basis standardisierter Protokolle und Schnittstellen Dienste für die Integration heterogener IT-Systeme bereitstellt [Les01]. Aufgrund der Vielfalt an Realisierungsmöglichkeiten für diese Schicht wird für eine ausführliche Betrachtung auf [Ham05] verwiesen.

eines Facettenschemas (in Abbildung 3.8 durch das Dokument mit der Inschrift *Facetten* dargestellt) geeignet beschrieben werden kann. Da diese Artefaktbeschreibung jedoch sowohl die für Attribut- und Ähnlichkeitsfacetten als auch die zur Durchführung von Ebenenwechsel (vgl. Kapitel 3.2.3) benötigten Informationen enthalten muss, wird für jedes Artefakt mindestens eine sogenannte *Polyrepräsentation* erzeugt. Unter einer Polyrepräsentation wird — ausgehend von dem von Ingwersen entwickelten Prinzip der Polyrepräsentation zur Verbesserung der Precisionwerte eines IRS [IJ05] — in der vorliegenden Arbeit eine Repräsentation verstanden, die selbst aus einer oder mehreren Repräsentationen unterschiedlicher Form besteht und somit als Container für mehrere Repräsentationen des gleichen Artefakts fungiert. Wie in Abbildung 3.9 dargestellt, können dabei generell vier Arten von Repräsentationen in einer Polyrepräsentation transportiert werden, die jeweils unterschiedliche charakteristische Eigenschaften eines Artefakts beschreiben. Um diesbezüglich das Verständnis im Zusammenspiel mit dem Retrievalprozess zu verbessern, werden diese als *Attributfacetten*, *Ähnlichkeitsfacetten*, *Beziehungsfacetten* und *Thumbnailfacetten* bezeichnet, wobei es sich bei der Ausprägung einer Facette in diesem Fall um eine Repräsentation der jeweiligen Artefaktinformation handelt. Wie diese (Poly-) Repräsentationserstellung im Detail abläuft, welche Komponenten hierzu eingesetzt werden und warum es in manchen Fällen für ein Artefakt auch mehrere Polyrepräsentationen geben kann, wird in Kapitel 5 ausführlich erläutert.

Bevor eine Polyrepräsentation allerdings letztendlich einem geeigneten Index, welcher in Form von Subindizes von einem *Index Manager* verwaltet wird, hinzugefügt werden kann (Schritt *Indexverwaltung*), ist sie einer *Konsistenzprüfung* zu unterziehen. Die Notwendigkeit einer solchen Prüfung liegt v. a. darin begründet, dass eine Artefaktinformation häufig zwar in Systemen verwaltet, aber zusätzlich auch — eventuell modifiziert — in Dokumenten enthalten ist. Folglich muss insbesondere bei Nichtübereinstimmung dieser Informationen eine Entscheidung getroffen werden, bei welcher Information es sich um die tatsächlich richtige und damit für die Indexierung gültige Information handelt. In Kapitel 6 werden daher die beiden Schritte Konsistenzprüfung und Indexverwaltung näher betrachtet.

4. Datenimport — Entwicklungsspezifische Artefakte und ihre Besonderheiten

Um während des Retrievalprozesses die richtigen Informationen in der richtigen Art und Weise bereitstellen zu können, ist nicht nur zu analysieren, welche Artefakte für Produktentwickler für eine Suche von Bedeutung sind, sondern v. a. welche Suchkriterien für diese Artefakte berücksichtigt werden müssen (vgl. Kapitel 4.1 und 4.2). Des Weiteren ist die Frage nach den Quellen für diese Informationen zu betrachten (Kapitel 4.3) und wie aus diesen die Informationen in das IRS importiert werden können. Die im Rahmen des LFRP-Frameworks hierzu vorgeschlagene Verwendung einer Containerdatei bzw. deren Aufbau wird in Kapitel 4.4 im Detail präsentiert.

4.1. Betrachtung ausgewählter entwicklungsspezifischer Artefakte

Wie in Kapitel 1.1.2 erläutert, ist der objektive Informationsbedarf von Produktentwicklern durch eine Vielzahl unterschiedlicher Artefakte (Produkte, Personen, Dokumente, ...) charakterisiert, welche sich jeweils einem bestimmten Artefakttyp zuordnen lassen. Gemäß Abbildung 4.1 sind dabei insbesondere die fünf Artefakttypen *Produkt*, *Dokument*, *Werkstoff*, *Person* und *Projekt* für die technische PE von Bedeutung. Auf sie wird zwar in der vorliegenden Arbeit der Fokus der Betrachtungen gelegt; eine Erweiterung um zusätzliche Artefakttypen (z. B. um einen Artefakttyp *Methode* zur Abbildung von zusätzlichem Prozesswissen) ist aber jederzeit möglich. Somit erfolgt in den nachfolgenden Kapiteln eine separate Analyse dieser fünf Artefakttypen hinsichtlich der für sie relevanten Suchkriterien. Dazu wird mit Hilfe der von der Object Management Group⁵⁵ entwickelten und standardisierten Modellierungssprache Unified Modeling Language (UML)⁵⁶ für jeden Artefakttyp eine eigene *Artefakttyp-hierarchie* — ähnlich zu der in Abbildung 4.1 illustrierten allgemeinen Hierarchie —

⁵⁵<http://www.omg.org>

⁵⁶<http://www.uml.org>

erstellt. Ohne Anspruch auf Vollständigkeit enthält diese für jeden Artefakttyp bzw. seine durch Spezialisierung abgeleiteten Subartefakttypen diejenigen Informationen als Attribute mit zugehörigen Datentypen, welche als Suchkriterien und damit als Facetten während des Retrievalprozesses für Produktentwickler zur Verfügung stehen sollten⁵⁷. Die einzige Ausnahme hiervon stellen die Attribute *ArtefaktID* und *ArtefaktURI*⁵⁸ der Superklasse *Artefakt* dar, da diese zur eindeutigen Identifizierbarkeit sowie für den Zugriff auf die Originalartefakte benötigt werden. Die Spezialisierung in Subartefakttypen, d. h. die Zuordnung der Instanzen der Superklasse zu den Subklassen, erfolgt gemäß UML anhand eines Diskriminators, welcher als Unterscheidungsmerkmal fungiert und dessen Werte sich aus den Namen der Subklassen ergeben. Demnach ist der in Abbildung 4.1 abgebildete Diskriminator *Artefakttyp* ein für jedes Artefakt zwingend erforderliche Attribut, das in Form einer Attributfacette zur Filterung verfügbar sein muss.

Da während der weiteren Betrachtung lediglich ein beispielhafter Überblick über wichtige Attribute der Artefakttypen gegeben werden soll, würde eine ausführliche Betrachtung jedes einzelnen, in den folgenden Artefakttyphierarchien enthaltenen Attributs den Rahmen der vorliegenden Arbeit sprengen. Zur Vollständigkeit und besseren Verständlichkeit werden daher alle in den Kapiteln 4.1.1 bis 4.1.4 als Attribute dargestellte Suchkriterien im Anhang definiert und gegebenenfalls durch Beispielausprägungen näher beschrieben.

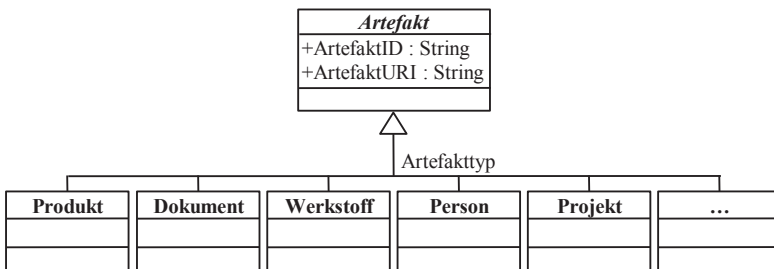


Abbildung 4.1. — Für den Bereich der technischen PE zu differenzierende Artefakttypen, deren zugehörige Artefakte die drei Attribute *ArtefaktID*, *ArtefaktURI* und *Artefakttyp* (= Diskriminator) besitzen müssen

⁵⁷Zur übersichtlicheren Darstellung wird grundsätzlich auf die Angabe von Einheiten oder Wertebereichen verzichtet.

⁵⁸Ein Uniform Resource Identifier (URI), der den Verzeichnispfad zur Originalquelle, die das Artefakt enthält, beschreibt.

4.1.1. Artefakttyp Produkt

Im Allgemeinen handelt es sich bei einem Produkt um ein Erzeugnis, welches als „(...) ein durch Produktion entstandener gebrauchsfähiger bzw. verkaufsfähiger materieller oder immaterieller Gegenstand.“ [Deu77, S. 54] definiert ist. Obwohl immaterielle Gegenstände in Form von Software aufgrund der stetig zunehmenden Mechatronisierung immer mehr Einzug in den Bereich der technischen PE halten, werden in der vorliegenden Arbeit vorrangig materielle Gegenstände im Sinne technischer Systeme betrachtet. Dabei ist unter einem technischen System gemäß Hubka „(...) ein, aus einer endlichen Menge von Elementen nach bestimmten Regeln geordnetes Ganzes.“ [Hub84, S. 11] zu verstehen, so dass dieses in Teilsysteme zerlegt werden kann. Diese Teilsysteme stellen selbst wieder Produkte dar und können auf verschiedene Art und Weise differenziert werden. So unterscheidet beispielsweise Koller technische Systeme eher allgemein anhand ihres Fachgebietes in mechanische, hydraulische, pneumatische und elektrische Systeme [Kol98]. Hubka hingegen unterteilt u. a. nach der Operandenart, d. h. nach dem zu verarbeitenden Objekt, in stoff-, energie- und informationsverarbeitende Systeme [Hub84]. Da sich neben diesen noch zahlreiche andere Kriterien für eine Differenzierung von Produkten finden lassen, werden im Folgenden sowie in Abbildung 4.2 drei allgemein bekannte Möglichkeiten exemplarisch verdeutlicht.

Differenzierung anhand der Produktgranularität

Zum einen kann je nach Granularität zwischen sehr einfach und komplex strukturierten Produkten unterschieden werden, was eine Differenzierung in die beiden Subklassen *Einzelteil* und *Baugruppe* erfordert. Da diese Unterteilung auch für eine Suche relevant ist, sollte sie mittels einer Facette **Produktgranularität**, welche als Diskriminator fungiert und damit für alle Produkte verfügbar sein muss, unterstützt werden (vgl. Abbildung 4.2 links). Hierbei ist unter einem Einzelteil, welches häufig auch als Bauteil bezeichnet wird, ein Element zu verstehen, das nicht zerstörungsfrei weiter zerlegt werden kann [Hub84, Kol98]. Typische Beispiele für derartige Einzelteile sind Maschinenelemente wie die in Abbildung 4.2 stellvertretend dargestellten Schrauben und Dichtungen, welche in gleicher oder ähnlicher Form immer wieder verwendet werden [MWJV07]. Diese sind zwar als Spezialisierungen der zweiten Differenzierungsmöglichkeit, nämlich anhand der *Produktgruppe*, visualisiert (vgl. Abbildung 4.2 unten), stellen jedoch gleichzeitig auch Subklassen der Superklasse *Einzelteil* dar. Somit verdeutlicht bereits dieses einfache Beispiel die Begrenztheit und damit die Nicht-Eignung von Monohierarchien zur Klassifizierung von Artefakten⁵⁹. Im Gegensatz zu Einzelteilen zeichnen sich Baugruppen dadurch aus, dass sie aus mehreren, fest oder beweglich miteinander verbundenen Elementen bestehen, die selbst wieder Einzelteile oder Baugruppen sind [Sch87, Kol98]. Demzufolge verfügen Baugruppen über

⁵⁹Aus Gründen einer übersichtlicheren Darstellung und besseren Verständlichkeit wird im Weiteren die Verwendung von Monohierarchien zur graphischen Erklärung allerdings beibehalten.

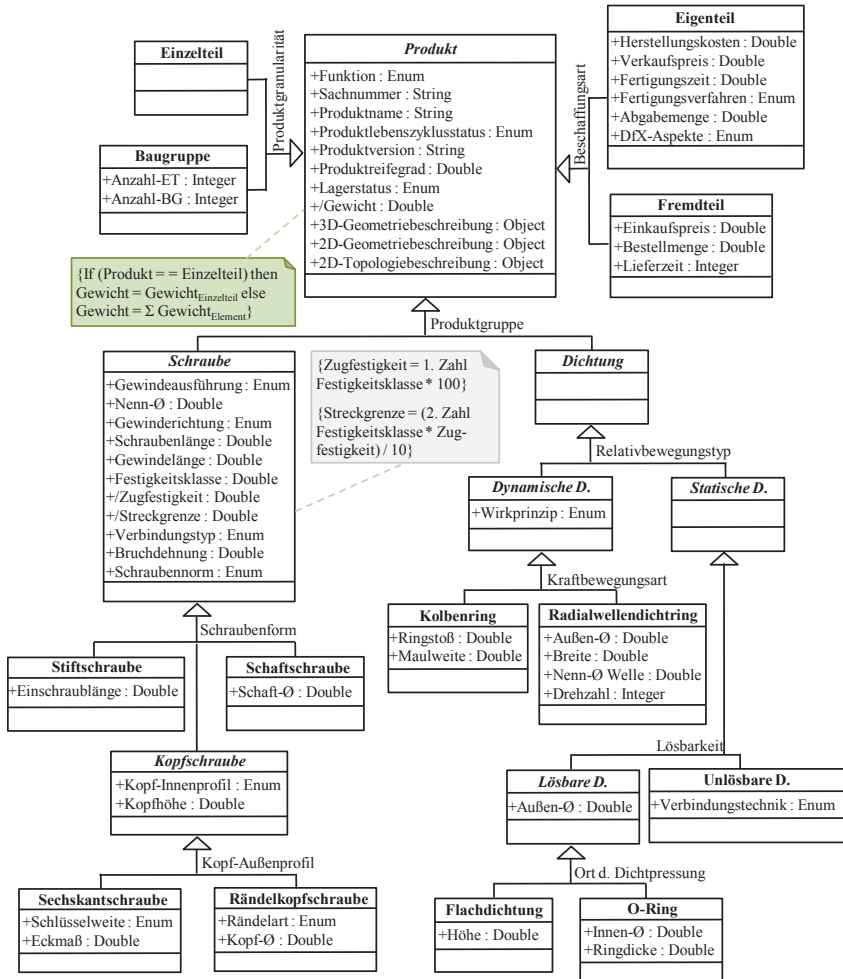


Abbildung 4.2. — Exemplarische Artefakttyphierarchie für Produkte in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute und Fokus auf Spezialisierungsbeziehungen

eine durch diese Elemente und ihre Beziehungen festgelegte Produktstruktur, die eine Bandbreite von eher einfachen (z. B. Kupplung, Zylinderkopf) über komplexe (z. B. Motor, Auto) bis hin zu hochkomplexen (z. B. Schiff, Flugzeug) Beziehungsgeflechten aufweisen kann [Les01, Ver02]. Damit stellen nicht nur die Anzahl der enthaltenen Einzelteile (*Anzahl-ET*) und Baugruppen (*Anzahl-BG*) interessante Suchkriterien für derartige Artefakte dar; vielmehr interessieren hier eigentlich die Elemente der Baugruppe an sich. Da diese selbst wiederum Produkte darstellen, sind sie folglich über *Produkt* \leftrightarrow *Produkt*-Beziehungen in Form von Assoziations- oder Aggregationsbeziehungen mit anderen Produkten verbunden (vgl. Kapitel 4.2 und Abbildung 4.3).

Differenzierung anhand der Produktgruppe

Eine zweite, bereits erwähnte Möglichkeit zur Differenzierung ist die Verwendung der *Produktgruppe*, wie sie Abbildung 4.2 in der nach unten ausgerichteten Vererbungsstruktur zeigt. Hierbei wird beispielhaft zwischen den beiden Subklassen *Schraube* und *Dichtung* unterschieden, für die jeweils eigene charakteristische Suchkriterien erforderlich sind. Während beispielsweise bei Schrauben die *Gewindeausführung*, die *Schraubenlänge* und die *Festigkeitsklasse* notwendige Kriterien zum Auffinden der benötigten Artefakte darstellen, ist bei Dichtungen zunächst die Frage nach dem *Relativbewegungstyp*, d. h. ob es sich um eine dynamische oder eine statische Dichtung handeln soll, von Interesse. Abhängig von der Beantwortung dieser Frage werden schließlich weitere Suchkriterien, wie beispielsweise die *Lösbarkeit* von statischen Dichtungen, für eine weitere Filterung benötigt. Während demzufolge Suchkriterien der Superklasse im Rahmen einer Suche gleichermaßen für alle Artefakte der zugehörigen Subklassen relevant sind, sind sie an die Subklassen zu vererben. Suchkriterien einer Klasse, die selbst keine weiteren Subklassen besitzt, sind hingegen auf die jeweilige Klasse beschränkt und daher nur für die zu dieser Klasse gehörigen Artefakte relevant. So sind im Beispiel von Abbildung 4.2 die *Schlüsselweite* und das *Eckmaß* für Sechskantschrauben oder der *Innendurchmesser* und die *Ringdicke* für O-Ringe spezifische Kriterien. Die *Produktgruppe* hingegen ist zusammen mit anderen in der Superklasse *Produkt* enthaltenen und im noch folgenden Abschnitt „Generelle Produktattribute“ beschriebenen Suchkriterien für alle Produkte — unabhängig von deren weiteren Differenzierung — gültig und damit als allgemeines Suchkriterium für Produkte relevant.

Differenzierung anhand der Beschaffungsart

Eine dritte mögliche Differenzierung basiert schließlich auf der *Beschaffungsart* der Produkte. Wie Abbildung 4.2 auf der rechten Seite zeigt, lassen sich sämtliche Produkte entweder als *Eigenteil* oder als *Fremdteil* charakterisieren. Unter der Subklasse *Eigenteil* werden dabei alle Produkte zusammengefasst, die im eigenen Unternehmen produziert werden und damit mögliche Suchkriterien wie *Herstellungskosten*, *Verkaufspreis*, *Fertigungszeit* oder das zugehörige *Fertigungsverfahren* erfordern. Auch

die bei der Entwicklung und Herstellung der Eigenteile berücksichtigten Gerechtigkeiten hinsichtlich eines Kriteriums oder eines Bereiches X (z. B. Kosten, Fertigung), welche auch als *DfX-Aspekte* bezeichnet werden [And05a, Stö09], stellen hier wichtige Suchkriterien dar und sollten z. B. unter Einsatz eines Prozessnavigators, dessen Konzept als Ergebnis des FORFLOW-Forschungsverbundes entstanden ist⁶⁰, erfasst werden. Demgegenüber bündelt die Subklasse *Fremdteil* alle Produkte, die von außen, d. h. von Lieferanten, bezogen werden und für deren Auffinden folglich Kriterien wie der *Einkaufspreis*, die *Bestellmenge* oder die *Lieferzeit* benötigt werden. Insbesondere bei diesen letzten Kriterien ist ein wichtiger Aspekt bei deren Erfassung zu berücksichtigen. Wird nämlich ein Fremdteil (z. B. eine Schraube) nicht nur von einem, sondern von mehreren Lieferanten bezogen, so besitzt dieses üblicherweise mehrere verschiedene Werte für das Attribut *Einkaufspreis*. Folglich ist für eine Anwendung der Attribute als Facetten für jedes Attribut dessen Wertigkeit festzulegen, so dass dieses entweder als ein- oder als mehrwertige Facette zu definieren ist.

Generelle Produktattribute

Grundsätzlich ist allen Produkten gemeinsam, dass sie aufgrund ihrer Struktur, d. h. aufgrund ihres Aufbaus aus Elementen und diese Elemente verbindenden Beziehungen, eine funktionierende Lösung für eine Problemstellung darstellen und dabei der Erfüllung eines bestimmten Zwecks dienen [Hub84, Hab08]. Letzteres wird durch ein bestimmtes Verhalten des Systems erreicht, welches aus der Umwelt stammende Eingangsgrößen in wiederum für die Umwelt benötigte Ausgangsgrößen unter Einhaltung definierter Randbedingungen überführt [PBF07, Hab08]. Dieses zweckgebundene Verhalten wird als *Produktfunktion* bezeichnet, die üblicherweise lösungsneutral beschrieben wird und Ausprägungen wie beispielsweise halten, speichern, leiten, trennen, dichten, antreiben oder andere besitzen kann [Hub84, AB⁺07]. Dabei sind Produkte natürlich nicht notwendigerweise auf die Erfüllung einer Funktion beschränkt, sondern können durchaus auch mehrere Funktionen (in der Regel eine Hauptfunktion und mehrere Nebenfunktionen) erfüllen. Weiterhin besitzt jedes Produkt gemäß Hubka Attribute, die es genauer definieren und folglich ebenfalls als Suchkriterien berücksichtigt werden sollten [Hub84]. Beispiele hierfür sind:

- die *Sachnummer*, die als auftragsunabhängige Nummer zur eindeutigen Identifikation und Differenzierung von Produkten in einem Unternehmen dient [PBF07, Wie08] und somit gleichzeitig die *ArtefaktID* dieses Artefakttyps repräsentiert,
- der *Produktname*,
- der *Produktlebenszyklusstatus*, der mittels Ausprägungen wie Entwurf, Geprüft, in Freigabe, Freigegeben usw. beschreibt, in welcher Prozessphase sich das Produkt befindet,

⁶⁰Siehe [EFH⁺09] für nähere Informationen.

- die *Produktversion*, die den Änderungsstand eines Produktes dokumentiert,
- der *Produktreifegrad*, der den zu einem bestimmten Zeitpunkt erreichten Zustand eines Produktes durch die Gegenüberstellung von zu erfüllenden Soll- und bereits erfüllten Ist-Anforderungen beschreibt [KMW09],
- der *Lagerstatus* und
- das *Gewicht*.

Das Attribut *Gewicht* nimmt dabei im Vergleich zu den anderen Attributen eine Sonderstellung ein, was in Abbildung 4.2 gemäß UML-Notation durch einen vorangestellten Schrägstrich gekennzeichnet ist. Derartig markierte Attribute stellen sogenannte abgeleitete Attribute dar, da sie aus anderen Attributen anhand einer vordefinierten Berechnungsvorschrift ermittelt werden können. Bei dieser Berechnungsvorschrift kann es sich im einfachsten Fall um eine einfache Rechenregel handeln, wie es am Beispiel der *Zugfestigkeit* und *Streckgrenze* von Schrauben in Abbildung 4.2 veranschaulicht ist (siehe grau gefärbte Einschränkung der Klasse *Schraube*). Beide Größen werden mit Hilfe des Attributs *Festigkeitsklasse* ermittelt, welche durch zwei Zahlen der Form $X.Y$ (z. B. 4.6) beschrieben wird. Hierbei ergibt sich die Zugfestigkeit als das 100fache der ersten Zahl X ($4 * 100 = 400$) und die Streckgrenze als ein Zehntel des Produktes aus der zweiten Zahl Y und der Zugfestigkeit ($\frac{1}{10}(6 * 400) = 240$). Daneben sind aber durchaus auch komplexer gestaltete Berechnungsvorschriften beispielsweise in Form von *if-then-else*-Regeln anwendbar, wie dies bei der Ermittlung des Gewichts (siehe grün gefärbte Einschränkung der Klasse *Produkt* in Abbildung 4.2) der Fall ist. Abhängig davon, ob es sich bei dem betrachteten Produkt um ein Einzelteil oder eine Baugruppe handelt, errechnet sich das Gewicht entweder als Gewicht des Einzelteils oder als Summe der Gewichte der in der Baugruppe enthaltenen Elemente.

Schließlich werden Produkte heutzutage mittels computerunterstützter Techniken im zwei- und dreidimensionalen Raum entwickelt, wodurch sowohl zwei- als auch dreidimensionale Beschreibungen in Form technischer Zeichnungen und CAD-Modelle für diese Produkte existieren. Diese dienen zwar primär zur Darstellung geometrischer Informationen, können aber auch topologische, d. h. strukturelle Informationen enthalten, so dass folglich diese *Geometrie-* und *Topologiebeschreibungen* ebenfalls als Suchkriterien für Produkte zu berücksichtigen sind. Allerdings sind diese nur aus den entsprechenden Dokumenten zu gewinnen, weshalb die Beziehung *Produkt* \leftrightarrow *Dokument* eine wichtige Rolle sowohl für die Indexierung als auch das Retrieval spielt. Gleiches gilt im Übrigen für die Beziehung *Produkt* \leftrightarrow *Werkstoff*, da jedes Produkt aus einem bestimmten Werkstoff gefertigt wird. Wie bereits erwähnt, werden diese Beziehungen in Kapitel 4.2 separat betrachtet.

Insgesamt zeigt sich, dass bei der Erfassung von Suchkriterien für Produkte zahlreiche Aspekte hinsichtlich artefaktspezifischer Attribute, Attributwertigkeit, Berechnungsregeln usw. zu berücksichtigen sind. Obwohl das zur Veranschaulichung in Abbil-

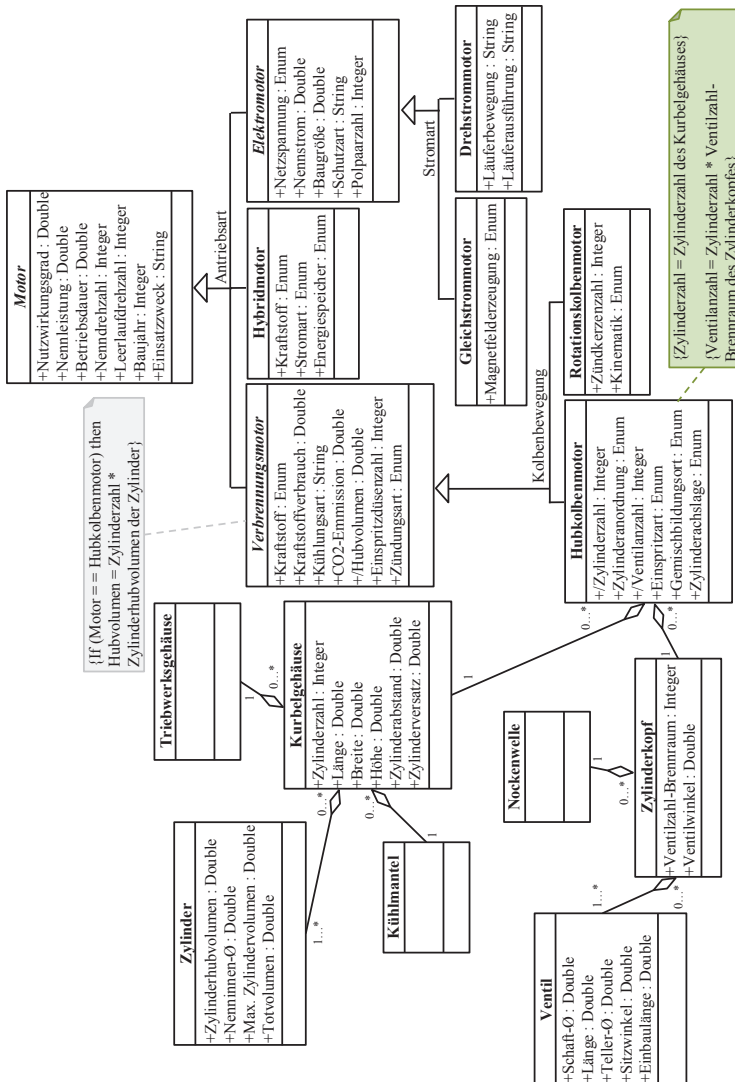


Abbildung 4.3. — Exemplarische Artefakttyphierarchie für Motoren als UML-Klassendiagramm mit relevanten Suchkriterien als Attribute und Fokus auf Spezialisierungs- und Aggregationsbeziehungen

dung 4.2 verwendete Beispiel einer Produkthierarchie relativ einfach und keinesfalls vollständig ist, verdeutlicht es bereits sehr gut die mit einer Indexierung verbundene Komplexität. Letztere wird umso höher, je umfangreicher und detaillierter die Produkthierarchie modelliert ist. Dies wird am Beispiel einer weiteren in Abbildung 4.3 dargestellten Produkthierarchie deutlich, welche mögliche Suchkriterien für die Baugruppe *Motor* zeigt. Im Vergleich zur Produkthierarchie aus Abbildung 4.2 sind hier zusätzlich Aggregationsbeziehungen zwischen den einzelnen Elementen der Baugruppe berücksichtigt, da diese für die Ermittlung bestimmter Attribute eine Rolle spielen. So benötigt man beispielsweise zur Ermittlung der *Ventilanzahl* eines Hubkolbenmotors zum einen dessen *Zylinderzahl*, welche der *Zylinderzahl* des im Motor enthaltenen Kurbelgehäuses entspricht. Zum anderen wird die Ventilanzahl je Brennraum (Attribut *Ventilzahl-Brennraum*) des zugehörigen Zylinderkopfes benötigt (siehe grün gefärbte Einschränkung der Klasse *Hubkolbenmotor* in Abbildung 4.3).

4.1.2. Artefakttyp Dokument

Neben Produkten bilden Dokumente einen weiteren wichtigen Artefakttyp im Bereich der technischen PE. Aus diesem Grund soll im Folgenden der Begriff des Dokuments zunächst allgemein erläutert und anschließend im betrachteten Anwendungsbereich genauer spezifiziert werden.

Allgemeine Definition

Während Andermann et al. ein Dokument als „(...) die materielle Einheit eines Trägers dokumentarischer Daten.“ [ABB⁺04, S. 28] verstehen, fasst die VDI-Richtlinie 2219 diesen Begriff etwas enger und definiert ihn als „(...) eine als Einheit behandelte Zusammenstellung von Informationen, die nichtflüchtig auf einem Informationsträger gespeichert sind.“ [Ver02, S. 82f] Dabei ist generell zwischen dem Originaldokument (= Primärdokument) und den sogenannten Sekundär- bzw. Tertiärdokumenten, welche durch eine Dokumentation von Dokumenten der jeweils vorhergehenden Stufe entstehen, zu unterscheiden [ABB⁺04]. Sowohl für die Dokumentation als auch die Erstellung der Dokumente selbst werden dabei heutzutage vielfältige Techniken eingesetzt, welche vom klassischen Druck (Zeitschrift, Buch, Zeitung) bis hin zur multimedialen Erzeugung (Audio, Video, Bild) in unterschiedlichsten Dimensionen (2D, 3D) und Kombinationen reichen. Damit verbunden ist schließlich auch die Art des zur Speicherung verwendeten Informationsträgers, woraus die zwei grundsätzlichen Erscheinungsformen der Papierdokumente und der elektronischen Dokumente resultieren. Welchen Inhalt diese jedoch konkret tragen, hängt vom jeweiligen Einsatzzweck des Dokumentes sowie von der Umgebung, in der das Dokument benötigt wird, ab. So dienen Dokumente im unternehmerischen Umfeld u. a. zum Belegen von Geschäftsprozessen und den damit verbundenen geschäftlichen Transaktionen und Entscheidungen

[Sut96]. Die Inhalte sind dabei allerdings so vielfältig, dass unterschiedliche Dokumententypen, welche sich jeweils auf eine bestimmte Art von Inhalt konzentrieren, zu differenzieren sind. Zudem kann die logische Organisation des Inhaltes prinzipiell in den drei Formen strukturiert, semi-strukturiert und unstrukturiert erfolgen, von denen erstere beispielsweise in Form von XML-Dokumenten im Unternehmensumfeld zwar immer häufiger, aber im Vergleich doch eher selten zum Einsatz kommt. In den meisten Fällen sind Dokumente unstrukturiert oder bestenfalls semi-strukturiert, wobei letztere zumindest implizit eine, wenn auch schwache und innerhalb des gleichen Dokumenttyps variierende Struktur enthalten [TS06, Sto07].

Dokumente der technischen PE

Da aufgrund der oben genannten allgemeinen Faktoren die Vielfalt existierender Dokumente enorm ist, sind die für den Anwendungsbereich der technischen PE relevanten Dokumente zu identifizieren. Da jedoch auch ihre Anzahl und Mannigfaltigkeit nicht vollständig erfassbar sind, soll mit Abbildung 4.4 ein Einblick in das mögliche Spektrum der während des PEP erstellten und benötigten Dokumente gegeben werden.

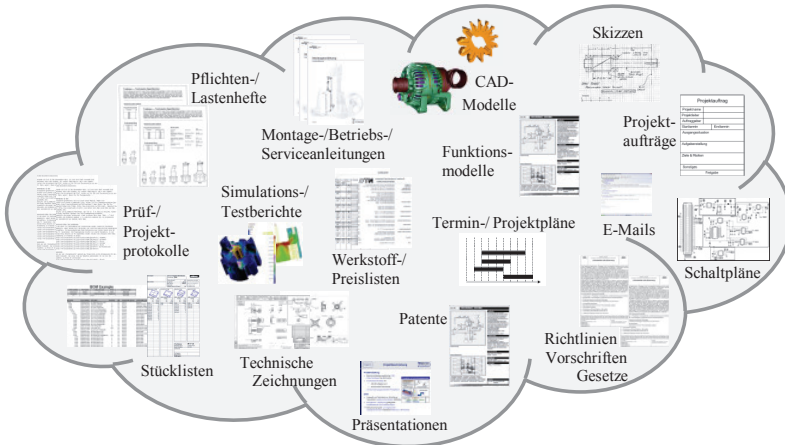


Abbildung 4.4. — Sammlung von Dokumententypen, die in der PE von Bedeutung sind

Wie man in dieser Darstellung erkennen kann, ist die PE durch zahlreiche unterschiedliche Dokumententypen geprägt. So finden sich neben allgemein gängigen Dokumenten wie Projektplänen, Protokollen, Präsentationen und E-Mails auch spezielle, nur in der PE vorkommende Dokumententypen wie beispielsweise CAD-Modelle, Stücklisten, Funktionsmodelle oder Simulationsergebnisse. Alle diese Dokumententypen sind auf un-

verschiedliche Inhalte fokussiert, weshalb der *Dokumenttyp* ein wesentliches Suchkriterium für diesen Artefakttyp darstellt. Dies gilt auch für Metadaten, die Dokumente unabhängig von ihrem Dokumenttyp näher charakterisieren. Da es sich hierbei um Daten über Daten bzw. Sachverhalte handelt, die im Fall von Dokumenten primär organisatorische und damit inhaltsunabhängige, aber auch inhaltsabhängige Aspekte (z. B. Titel, Kurzfassung) in strukturierter Form beschreiben, tragen sie wesentlich zu einer verbesserten Auffindbarkeit der Dokumente bei [ABB⁺04, Sto07]. Obwohl Metadatenstandards wie beispielsweise Dublin Core⁶¹ für Internetressourcen oder MPEG-7⁶² für Multimediaobjekte existieren und eine vordefinierte Menge beschreibender Terme festlegen, sind diese für eine Anwendung im LFRP-Framework nicht ohne Weiteres zu übernehmen. Einer der Gründe hierfür ist, dass der im Rahmen von Dublin Core als Metadatum geführte Autor eines Dokumentes ein eigenständiges Artefakt darstellt, zu dem eine *Dokument* ↔ *Person*-Beziehung besteht (vgl. Kapitel 4.2). Somit muss im Detail analysiert werden, welche Metadaten tatsächlich als Suchkriterien für Dokumente zu verwenden sind. Wie Abbildung 4.5 in der Superklasse *Dokument* zeigt, gehören hierzu u. a.

- die bei der Verwaltung in Drittsystemen zur Identifikation verwendete und die *ArtefaktID* dieses Artefakttyps repräsentierende *DokumentID*⁶³,
- das Datum der Dokumenterstellung (= *Erstellungsdatum*),
- das Datum des letzten Zugriffs auf das Dokument (= *Zugriffsdatum*),
- die *Dokumentbezeichnung* in Form eines Titels oder Namens (z. B. Dateiname),
- die im Rahmen einer Versionierung den Änderungszustand dokumentierende *Dokumentversion*,
- das *Archivierungssystem*, welches den Ort der Dokumentaufbewahrung (z. B. Aktenschränk, IT-System oder andere) beschreibt und damit einen Zugriff auf das Dokument ermöglicht, sowie
- die Sprache, in welcher der Dokumentinhalt verfasst ist (= *Dokumentsprache*).

Des Weiteren ist zu berücksichtigen, dass Papierdokumente in der technischen PE trotz zunehmender Digitalisierung noch immer einen hohen Stellenwert einnehmen. Aufgrund des kreativen Charakters von Produktentwicklern und der damit verbundenen Kopf-Hand-Assoziation, werden insbesondere bei der Entwicklung neuer und innovativer Produkte häufig zunächst Skizzen zum Festhalten und Kommunizieren von

⁶¹<http://www.dublincore.org>

⁶²<http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>

⁶³Fehlt eine derartige DokumentID, z. B. aufgrund der Verwaltung in einfachen Dateiverwaltungssystemen, so ist diese im Rahmen der Indexierung in Form eines Universally Unique Identifier (UUID) — einem von der Open Software Foundation entwickelten und im RFC 4122 beschriebenen Standard für Identifikatoren (<http://tools.ietf.org/html/rfc4122>) — zu erzeugen.

Ideen erstellt oder einfache Auslegungsrechnungen zur überschlägigen Dimensionierung des Produktes durchgeführt [MHC09]. Damit ist die *Erscheinungsform*, welche ein Dokument entweder als Papier- oder als elektronisches Dokument klassifiziert, ein weiteres wichtiges Attribut für Dokumente. Allerdings sind diese Papierdokumente oft sehr unstrukturiert, so dass sich eine Extraktion der Inhalte trotz einer für die weitere Verarbeitung notwendigen Digitalisierung mittels OCR sehr schwierig und aufwändig gestalten würde. Aus diesem Grund wurde der Fokus für die weitere Betrachtung auf die Gruppe der elektronischen Dokumente gelegt, was die Hinzunahme zusätzlicher organisatorischer Metadaten in Form der *Dateigröße*, des zur Erstellung verwendeten *AwS* oder des zur Speicherung genutzten *Dateiformates* als Suchkriterien der Superklasse *Dokument* hier rechtfertigt.

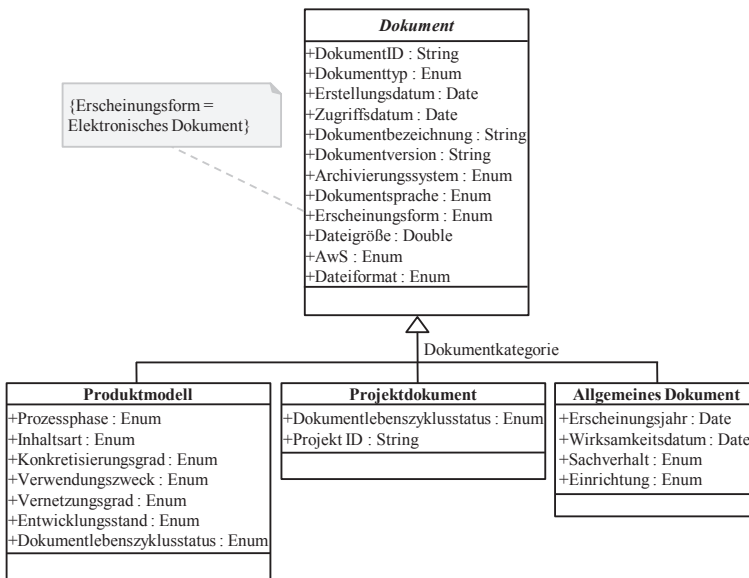


Abbildung 4.5. — Exemplarische Artefakttyphierarchie für Dokumente in Form eines UML-Klassendiagramms mit der *Dokumentkategorie* als Diskriminator und relevanten Suchkriterien als Attribute

Differenzierung anhand der Dokumentkategorie

Wie Produkte lassen sich auch Dokumente anhand verschiedener Kriterien differenzieren. Eines dieser Kriterien ist die *Dokumentkategorie*, anhand welcher Dokumente ge-

maß Abbildung 4.5 in die drei Subklassen *Produktmodell*, *Projektdokument* und *Allgemeines Dokument* eingeordnet werden. Diese Unterteilung resultiert aus der Tatsache, dass Dokumente im Bereich der technischen PE drei wesentliche Themen unterschiedlich stark adressieren. So dienen diese zum einen der Beschreibung der verschiedenen Aspekte eines Produktes (Funktion, Gestalt, Qualität, Produktion, ...) [Sch99], weshalb ihr Inhalt hauptsächlich aus Produktinformationen besteht. Diese von Günzler und auch von Lauer als Produktmodelle bezeichneten Dokumente resultieren folglich als (Zwischen-) Ergebnisse aus den einzelnen Phasen des PEP, wobei die aus einer bestimmten Sichtweise, d. h. im Hinblick auf eine gegebene Zielsetzung, relevanten Eigenschaften eines Produktes erfasst und festgelegt werden [Gün05, Lau09]. Somit zeichnen sich Produktmodelle neben den verschiedenen Zielsetzungen auch durch unterschiedliche Konkretisierungsstufen eines Produktes aus, weshalb sie sich nicht nur durch die *Prozessphase*, in der sie entstanden sind, sondern gemäß Lauer auch anhand weiterer prozessorientierter Parameter (*Inhaltsart*, *Konkretisierungsgrad*, *Verwendungszweck*, *Vernetzungsgrad* und *Entwicklungsstand*) charakterisieren lassen [Lau09]. Werden diese Attribute folglich von einem PMS oder unter Nutzung des bereits zuvor erwähnten Prozessnavigators erfasst, können sie als wertvolle Suchkriterien für Produktmodelle dienen. Daneben sind bei der Erstellung dieser Produktmodelle sowie der zugrunde liegenden Entwicklungstätigkeit Dokumente zu beachten, die bestimmte, einzuhalten- de Vorgaben definieren. Diese sind nicht auf ein bestimmtes, zu entwickelndes Produkt zugeschnitten, sondern weisen einen allgemeinen Charakter und damit eine generelle Gültigkeit auf. Dokumente wie Richtlinien, Normen, Vorschriften oder Gesetze gehören zu dieser Kategorie der allgemeinen Dokumente, für die folgende Attribute als zusätzliche Suchkriterien denkbar sind:

- das *Erscheinungsjahr* des Dokumentes,
- das Datum, ab dem das Dokument wirksam ist (Attribut *Wirksamkeitsdatum*),
- der im Dokument adressierte *Sachverhalt* (z. B. Sicherheit, Umweltschutz, Ergonomie, ...) und
- die *Einrichtung*, die das Dokument verabschiedet hat (DIN, ISO, VDI, VDA, ...).

Neben diesen beiden Kategorien von Dokumenten wird schließlich noch eine dritte Gruppe benötigt, welche die Art der Geschäftsabwicklung im Unternehmen adressiert. Da die Entwicklung von Produkten üblicherweise in Form von Projekten erfolgt, werden deren Planung, Abwicklung und Kontrolle ebenfalls in Dokumenten festgehalten. Aus diesem Grund finden sich auch hier zahlreiche für das Projektgeschäft typische Dokumenttypen wie Projektpläne, Terminpläne oder Kostenkalkulationen. Für diese sollte u. a. eine Suche anhand des zugehörigen Projektes (vgl. *Dokument* ↔ *Projekt*-Beziehung in Kapitel 4.2) sowie über den aus WfMS oder DMS verfügbaren *Dokumentlebenszyklusstatus* möglich sein. Letzterer wird durch Ausprägungen wie in

Tabelle 4.1. — Zuordnung einer Auswahl von Dokumenttypen zu den drei Dokumentkategorien *Projektdokument*, *Produktmodell* und *Allgemeines Dokument*

Projektdokument	Produktmodell	Allgemeines Dokument
Terminplan	Lastenheft	Umweltrichtlinie
Projektplan	Pflichtenheft	Sicherheitsrichtlinie
Meilensteinmatrix	Anforderungsliste	Konstruktionsrichtlinie
Meilensteindokument	Skizze	Konstruktionskatalog
Kostenkalkulation	Technische Zeichnung	Norm
Ressourcenplan	CAD-Modell	Gesetz
Lessons Learned	Schaltplan	Arbeitsanweisung
Projektauftrag	Stückliste	E-Mail
Änderungsprotokoll	Funktionsmodell	Präsentation
Besprechungsprotokoll	Einflussmatrix	Marktanalyse
Änderungsmitteilung	Datenblatt	Checkliste
Statusbericht	Patent	Materialliste
	Prüfbericht	Unternehmensrichtlinie
	Bewertungsprotokoll	
	Simulationsbericht	
	Montageanleitung	
	Betriebsanleitung	
	Recyclinganleitung	
	Machbarkeitsstudie	
	Preisliste	
	Produktionsplan	
	Testbericht	

Bearbeitung, in Änderung, in Freigabe usw. beschrieben und stellt auch für Produktmodelle ein wichtiges Suchkriterium dar.

Zusammenfassend zeigt Tabelle 4.1 exemplarisch eine Zuordnung von Dokumenttypen zu den einzelnen Kategorien, von denen jedoch jeder einzelne Dokumenttyp aufgrund seines Inhaltes wiederum durch spezifische Attribute charakterisiert ist. Folglich sind für eine Suche im Rahmen des LFRP-Frameworks sämtliche relevanten Dokumenttypen genauer zu analysieren und um die für eine Suche benötigten inhaltsbezogenen Attribute zu erweitern. Da allerdings eine Betrachtung aller aufgezeigten Dokumenttypen den Rahmen der vorliegenden Arbeit sprengen würde und bisherige Suchansätze

kaum entwicklungsspezifische Dokumenttypen, d. h. Produktmodelle, berücksichtigen, werden mögliche inhaltsbezogene Suchkriterien für diese bei der Schilderung des zweiten Indexierungsschrittes — der Repräsentationserstellung — im Detail analysiert (vgl. Kapitel 5.4).

4.1.3. Artefakttyp Werkstoff

Werkstoffe sind in der technischen PE von entscheidender Bedeutung, da ohne sie keine Produktherstellung möglich wäre. Folglich beeinflussen diese bzw. ihre Eigenschaften die zur Erfüllung der Produktfunktion benötigten Merkmale eines Produktes, wie beispielsweise sein Gewicht oder seine Hitzebeständigkeit, und tragen darüber hinaus auch zur Festlegung der Produktgestalt bei. Letzteres resultiert aus der Tatsache, dass Werkstoffe nicht beliebig verformbar sind und demzufolge auch nur den Einsatz bestimmter Fertigungsverfahren erlauben, welche selbst wiederum nur bestimmte Gestaltsformen erzeugen und nicht alle Werkstoffe verarbeiten können. Als Konsequenz aus diesen Wechselwirkungen zwischen Funktion, Material, Fertigung und Gestalt resultiert eine hohe Komplexität, die eine Suche nach dem richtigen Werkstoff zu einer nicht zu unterschätzenden Aufgabe macht [Ash07, MRB⁺08].

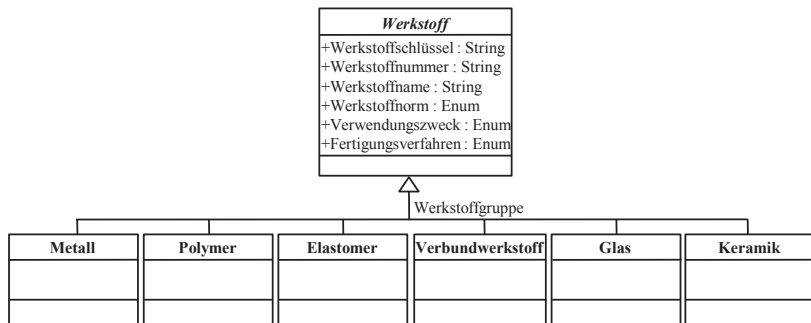


Abbildung 4.6. — Exemplarische Artefakttyphierarchie für Werkstoffe in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute

Verstärkt wird diese Komplexität durch die zunehmende Zahl an neu entwickelten Werkstoffen, deren Anzahl laut Ashby bereits im Jahr 2007 bei etwas mehr als 120000 verschiedenen Werkstoffarten lag [Ash07]. Zur besseren Strukturierung differenziert man diese abhängig von ihren Eigenschaften und dem damit verbundenen Verhalten in sechs Werkstoffgruppen, so dass jeder Werkstoff, wie in Abbildung 4.6 anhand des Diskriminators *Werkstoffgruppe* visualisiert, über seine jeweilige Subklasse — *Metall*, *Polymer*, *Elastomer*, *Verbundwerkstoff*, *Glas* oder *Keramik* — gefunden werden

kann [Ash07]. Zusätzlich wird jeder Werkstoff sowohl durch identifizierende als auch klassifizierende Attribute näher beschrieben [MRB⁺08]. Die Identifikation erfolgt dabei primär über den auch unter dem Begriff des Werkstoffkürzzeichens geführten *Werkstoffschlüssel*, welcher u. a. die chemische Zusammensetzung eines Werkstoffes definiert und einer kurzen, aber dennoch eindeutigen Verwendung von Werkstoffen in Dokumenten wie Stücklisten oder technischen Zeichnungen dient. Er stellt folglich die *ArtefaktID* dieses Artefakttyps dar. Zum anderen wird mit Hilfe der *Werkstoffnorm* bzw. deren Normenbezeichnung, eine einheitliche und allgemein gültige Definition des Werkstoffes ermöglicht. Als klassifizierende Attribute werden der *Werkstoffname*, welcher die Einordnung des Werkstoffes in eine konkrete Subklasse der entsprechenden Werkstoffgruppe beschreibt, und eine *Werkstoffnummer*, die im Sinne einer Klassifikationsnotation eine genauere Orientierung ermöglicht, verwendet. So verdeutlicht beispielsweise die Werkstoffnummer 1.5023, dass es sich bei dem jeweiligen Werkstoff um den Federstahl ($\hat{=}$ Werkstoffname) 38Si7 ($\hat{=}$ Werkstoffschlüssel) handelt, welcher als 23. Element (4. und 5. Ziffer der Werkstoffnummer) der Gruppe Bau-, Maschinenbau-, Behälterstahl (2. und 3. Ziffer der Werkstoffnummer) der Hauptgruppe Stahl (1. Ziffer der Werkstoffnummer) klassifiziert ist [MT08, Wik10e].

Obwohl diese Attribute wichtige Kriterien für das Auffinden eines geeigneten Werkstoffes darstellen, kann bei der Entwicklung eines Produktes nicht davon ausgegangen werden, dass per definitionem bekannt ist, welche Werkstoffgruppe am besten für dieses Produkt geeignet ist. Vielmehr sind aus den Produktanforderungen entsprechende, zu erfüllende Anforderungen an den Werkstoff abzuleiten und in Werkstoffeigenschaften zu übersetzen [MRB⁺08]. So sind neben den in Abbildung 4.6 enthaltenen Attributen *Verwendungszweck* und *Fertigungsverfahren*, welche die möglichen Einsatzfelder des Werkstoffes (z. B. als Federring oder Schraubensicherung) und die anwendbaren Fertigungsverfahren (z. B. Schweißen, Löten, ...) beschreiben, noch andere Eigenschaften für eine Auswahl des richtigen Werkstoffes zu berücksichtigen. Allerdings beziehen sich diese nicht nur auf mechanische, sondern auch physikalische, technologische, gebrauchsspezifische und wirtschaftliche Gesichtspunkte, so dass ihre Zahl und damit die Menge potentieller Suchkriterien sehr umfangreich ist. Aus diesem Grund soll Abbildung 4.7, welche für jeden der fünf Aspekte eine exemplarische Auflistung zugehöriger Werkstoffeigenschaften zeigt, einen Eindruck der verfügbaren Eigenschaften vermitteln. Je nachdem welche dieser Eigenschaften im konkreten Anwendungsfall benötigt werden, ist dabei eindeutig festzulegen, ob eine Eigenschaft als Attribut für alle Werkstoffe oder nur für einzelne Subklassen gelten soll. Der Grund hierfür liegt in den unterschiedlichen Begriffsdefinitionen, die mit einer Eigenschaft in Verbindung gebracht werden. Als Beispiel hierfür soll kurz die *Härte* von Werkstoffen betrachtet werden. Obwohl prinzipiell jeder Werkstoff über seine Härte (z. B. hart, mittelhart, weich) charakterisiert werden kann, werden je nach Werkstoffgruppe zur Messung der tatsächlichen Härtewerte unterschiedliche Verfahren eingesetzt [MRB⁺08]. So erfolgt die Härteprüfung bei Elastomeren beispielsweise nach dem Shore-Verfahren, weshalb

man von der sogenannten Shore-Härte spricht; bei weichen bis mittelharten Metallen hingegen wird das Brinell-Verfahren verwendet, das die sogenannte Brinell-Härte ermittelt. Sofern erforderlich, ist es also auch möglich, die einzelnen Härteskalen als spezifische Attribute der jeweiligen Subklassen zu betrachten.

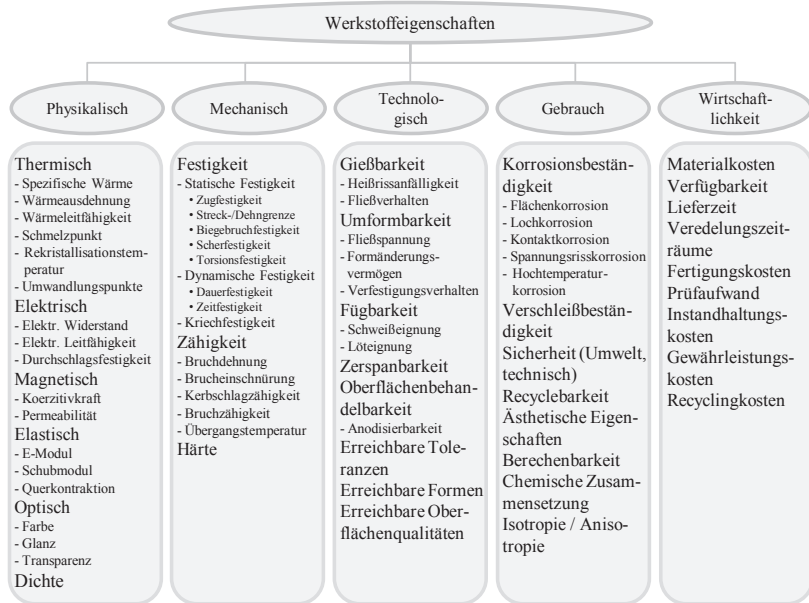


Abbildung 4.7. — Überblick über wichtige Werkstoffeigenschaften, welche unterschiedliche, bei der Entwicklung zu berücksichtigende Aspekte adressieren (nach [MRB⁺08, S. 30])

4.1.4. Artefakttyp Person

Wie in Kapitel 1.1.2 erläutert, sind Ansprechpartner für Produktentwickler oftmals die erste Quelle, um an Informationen zu spezifischen Produktfragen oder Problemen bzw. deren Lösungen zu gelangen. Bei der Analyse dieser Ansprechpartner zeigt sich, dass unter Berücksichtigung der verschiedenen Thematiken grundsätzlich drei wesentliche Personengruppen von Interesse sind. Als Differenzierungskriterium hierfür kann die Rolle herangezogen werden, die eine Person generell im Umfeld der Lieferkette (supply chain) einnimmt. Daraus resultieren die drei in Abbildung 4.8 unter Ver-

wendung des Diskriminators *SC-Rolle* dargestellten Subklassen *Lieferant*, *Kunde* und *Unternehmensmitarbeiter*.

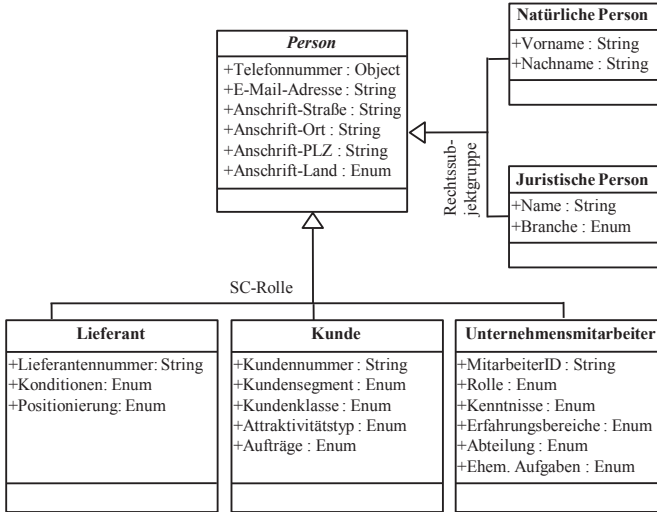


Abbildung 4.8. — Exemplarische Artefakttyphierarchie für Personen in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute

Die erste Gruppe der Lieferanten ist dabei v. a. zur Klärung von Fragen in Bezug auf die Beschaffung von benötigten Werkstoffen und Fremtteilen erforderlich. Attribute wie die *Lieferantenummer* zur eindeutigen Identifikation, die im Rahmen der Beschaffung gewährleisteten *Konditionen* sowie die Position innerhalb des typischerweise mehrstufigen Zulieferernetzwerks (Attribut *Positionierung*) können hier als brauchbare beschreibende Attribute betrachtet werden.

Die zweite Gruppe umfasst die Empfänger bzw. Käufer des entwickelten Produktes, die entweder als Auftraggeber die Entwicklung eines bestimmten Produktes fordern oder als potentielle Abnehmer für ein neues, innovatives Produkt in Frage kommen. Um ihre Daten eindeutig zuordnen zu können, verfügen auch sie über eine *Kundennummer* zur Identifikation ($\hat{=}$ *ArtefaktID*). Darüber hinaus spielen — insbesondere bei der Entwicklung neuer Produkte — die im Rahmen des CRM erfassten Kundenmerkmale eine wichtige Rolle, da aus ihnen notwendige Anforderungen abgeleitet werden können, wie beispielsweise ein hoher Exklusivitäts- oder ein niedriger Kostenfaktor. Hierzu gehören u. a. Informationen darüber, welchem *Kundensegment* (Geschäfts- oder Privatkunde), welcher *Kundenklasse* (Stamm- oder Neukunde) sowie welchem

im Rahmen einer ABC-Analyse betrachteten *Attraktivitätstyp* ein Kunde angehört. Aber auch die Kundenhistorie in Form bisheriger *Aufträge* sowie persönliche (z. B. Alter, Familienstand, Interessen, Bildung) oder finanzielle (z. B. Einkommen, Bonität, Versicherungen) Informationen können hier unter Umständen als hilfreiche Attribute betrachtet werden, um zukünftige Kundenwünsche zu prognostizieren [Pul04].

Die dritte und für Produktentwickler wohl wichtigste Informationsquelle ist der eigene Kollegenkreis, also die Gruppe der *Unternehmensmitarbeiter*, deren Daten üblicherweise unter Verwendung einer *MitarbeiterID* im Unternehmen gespeichert werden. Sie übernehmen nicht nur unterschiedliche *Rollen* innerhalb eines Unternehmens, sondern auch innerhalb des PEP (z. B. Konstrukteur, Tester, Fertigungsplaner, ...) und haben damit unterschiedliche Aufgaben zu erfüllen. Außerdem verfügt jeder Mitarbeiter über individuelle *Kenntnisse* und *Erfahrungsbereiche* und kann somit sozusagen als Experte andere Mitarbeiter bei Problemen oder Fragen in diesen Gebieten hilfreich unterstützen. Folglich ist das Auffinden von Kollegen, insbesondere wenn diese nicht im unmittelbaren Arbeitsumfeld tätig sind, von besonderer Bedeutung. Neben den bereits erwähnten Attributen sollten daher weitere Kontextfaktoren wie beispielsweise die *Abteilung*, der ein Mitarbeiter angehört, oder die von ihm bereits in der Vergangenheit erfüllten Aufgaben (Attribut *Ehem. Aufgaben*) als Suchkriterien Berücksichtigung finden. Gleiches gilt im Übrigen auch für die Kontakt- und Adressdaten des Mitarbeiters. Diese besitzen generell für jede Person Gültigkeit, was allerdings nicht für den Namen einer Person gilt. Diesbezüglich ist zwischen den beiden Rechtssubjekten *Natürliche Person* und *Juristische Person* zu unterscheiden (siehe Abbildung 4.8 rechts), wobei jede natürliche Person (mindestens) einen *Vornamen* und einen *Nachnamen* besitzt. Insbesondere bei Instanzen der Subklassen *Lieferant* und *Kunde* kann eine Person aber durchaus auch eine juristische Person sein, die sich durch Attribute wie den *Namen* der Organisation oder die *Branche* näher beschreiben lässt.

4.1.5. Artefakttyp Projekt

Da die Durchführung von Projekten nicht nur auf den Bereich der technischen PE beschränkt ist, bilden diese einen für die Suche wichtigen und für viele Anwendungsbereiche relevanten Artefakttyp. Dabei ist ein Projekt allgemein als ein komplexes Vorhaben definiert, „(...) welches zeitlich durch einen definierten Anfangs- und Endtermin begrenzt (...) sowie durch die Einmaligkeit seiner Bedingungen (...) gekennzeichnet ist.“ [HHMS06, S. 7] Dieser Definition folgend kann somit grundsätzlich jedes Projekt zunächst durch einen fest definierten *Starttermin* sowie seinen *Endtermin* beschrieben werden, welche darüber hinaus eine Ermittlung der zwischen ihnen liegenden Zeitspanne als *Zeitdauer* des Projektes erlauben (vgl. Abbildung 4.9). Weiterhin stellen die einmaligen Bedingungen wichtige beschreibende und damit für eine Suche zu berücksichtigende Kriterien dar. Zu ihnen zählen u. a. die *Ziele*, die mit der

Durchführung des Projektes angestrebt werden, sowie die für die Projektabwicklung zur Verfügung stehenden Ressourcen. Vor allem das maximal ausschöpfbare *Budget*, aber auch die tatsächlich im Projekt entstandenen *Kosten* sind hier von entscheidender Bedeutung. Betrachtet man neben dieser definitorischen Sichtweise zusätzlich den organisatorischen Aspekt, so lassen sich Projekte üblicherweise noch durch folgende Attribute typisieren:

- eine *Kurzbezeichnung* bzw. ProjektID zur eindeutigen Identifikation,
- einen *Projektnamen*,
- eine das Projekt bzw. seinen Inhalt näher erläuternde *Beschreibung* und
- eine Kennzahl, die der Überwachung und dem gegebenenfalls rechtzeitigen Eingreifen bei Planabweichungen dient, indem sie angibt, wie weit ein Projekt bereits fortgeschritten ist (= *Projektfortschritt*) [HHMS06].

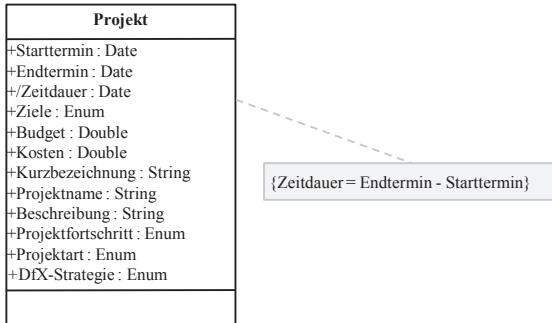


Abbildung 4.9. — Artefakttyp Projekt als UML-Klasse mit relevanten Suchkriterien als Attribute

Im Hinblick auf das vorliegende Anwendungsgebiet der technischen PE sind Projekte nicht nur durch diese eher allgemein gültigen Projektattribute, sondern durch zwei weitere, eher entwicklungsspezifische Kriterien charakterisierbar. So kann zum einen eine Selektion anhand der *Projektart* erfolgen, welche eine Differenzierung von Projekten anhand der drei in Kapitel 1.1.2 erläuterten Konstruktionsarten der Neu-, Anpassungs- und Variantenkonstruktion vornimmt. Zum anderen wird in jedem Projekt, d. h. mit jeder PE, eine bestimmte DfX-Strategie verfolgt (z. B. Leichtbau, geringe Kosten, hohe Qualität). Diese wird jeweils zu Beginn eines Projektes von den Verantwortlichen festgelegt und beeinflusst demzufolge die für Produkte zu berücksichtigenden DfX-Aspekte (vgl. Kapitel 4.1.1) [Stö09]. Damit stellt sie ein für Produktentwickler hilfreiches projektbeschreibendes und -differenzierendes Attribut dar.

4.2. Beziehungen zwischen Artefakten

Die Analyse der in der PE benötigten Artefakttypen hat gezeigt, dass jeder einzelne Artefakttyp über spezifische Attribute verfügt, welche hauptsächlich als Attribut- und teilweise als Ähnlichkeitsfacetten für das vorgeschlagene LFRP-Framework genutzt werden können. Allerdings reichen diese fokussierten Informationen alleine nicht aus; vielmehr interessieren oftmals Eigenschaften, die nur durch die Verbindung zu einem

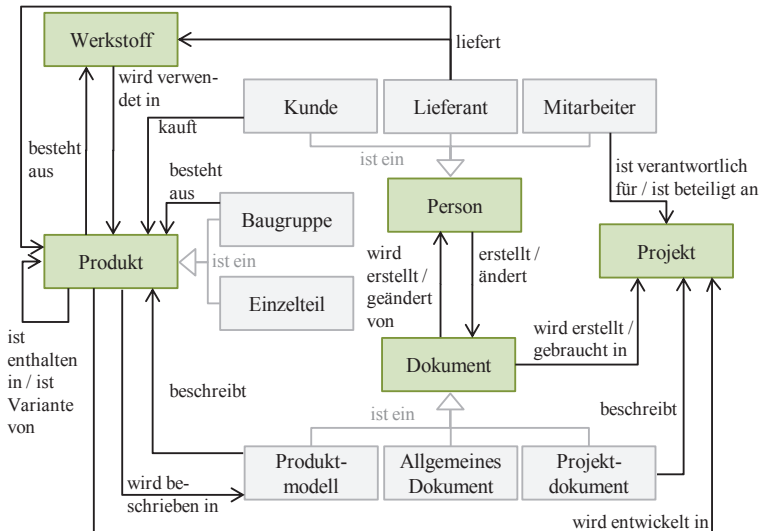


Abbildung 4.10. — Exemplarisches Beziehungsnetz bei Unterscheidung von fünf Artefakttypen (grün markiert): Produkt, Person, Dokument, Werkstoff und Projekt

anderen Artefakt ersichtlich sind. So werden beispielsweise Dokumente häufig anhand ihres Autors gesucht, da lediglich bekannt ist, dass Herr X bzw. Frau Y das Dokument erstellt hat. Dabei ist der Autor eines Dokumentes allerdings als selbstständiges Artefakt zu betrachten, zu dem, wie Abbildung 4.10 zeigt, eine Verbindung über die Beziehung *Dokument 'wird erstellt von' Person* besteht. Gleiches gilt auch für Artefakte, die dem gleichen Artefakttyp angehören. So bestehen Produkte häufig aus mehreren Komponenten (*Produkt 'besteht aus' Produkt*) oder werden in mehreren Varianten gefertigt (*Produkt 'ist Variante von' Produkt*), wobei es sich bei diesen Komponenten bzw. Varianten selbst wiederum um eigenständige Produkte mit spezifischen Attributen handelt. Folglich kann eine Suche zwar nur auf einen Artefakttyp

beschränkt sein, was jedoch nicht zwangsläufig der Fall sein muss. Aus diesem Grund ist es erforderlich, sowohl Beziehungen innerhalb eines Artefakttyps als auch zwischen unterschiedlichen Artefakttypen im Rahmen einer Suche verfolgen und somit gegebenenfalls zwischen Artefakttypebenen wechseln zu können (vgl. Ausführungen zum Ebenenkonzept in Kapitel 3.2.3). Dies bedingt seinerseits eine Indexierung von Beziehungen als Beziehungsfacetten, wobei zu berücksichtigen ist, dass die Zahl der zu erfassenden Artefaktbeziehungen mit der Zahl der differenzierten Artefakttypen steigt. Daraus resultiert unter Umständen nicht nur ein umfangreiches Beziehungsnetz, das es im Rahmen des Retrievalprozesses zu berücksichtigen gilt, sondern auch eine weitere Erhöhung der Komplexität bei der Indexierung. Zur Verdeutlichung zeigt Abbildung 4.10 ein Beispiel für ein Beziehungsnetz, welches lediglich die in Kapitel 4.1 erläuterten Artefakttypen als Knoten und die zwischen ihnen existierenden Beziehungen als gerichtete Kanten enthält. Obwohl hier nur ein überschaubarer Ausschnitt der möglichen, zwischen den Artefakttypen bestehenden Beziehungen dargestellt ist, wird deutlich, dass die Erfassung von Beziehungsfacetten im Indexierungsprozess und ihre Bereitstellung zur Nutzung im Retrievalprozess bereits bei einer geringen Anzahl von Artefakttypen einen deutlichen Mehraufwand erfordern.

4.3. Quellen für Artefaktinformationen

Nachdem die für eine Suche relevanten Artefakttypen identifiziert wurden, ist zu analysieren, aus welchen Quellen Informationen zu den Artefakten sowie ihren Beziehungen gewonnen werden können. Im Folgenden wird diese Frage für jede Informationsart einzeln beantwortet.

Produktdaten Informationen zur Beschreibung von Produkten finden sich zum einen in den bereits zuvor erwähnten Produktmodellen, die gerade für diesen Zweck erstellt werden. Demzufolge nehmen insbesondere Dokumente dieser Gruppe, wie Pflichtenhefte, CAD-Modelle, technische Zeichnungen oder Stücklisten, im Hinblick auf eine Suche einen besonderen Stellenwert ein und dürfen nicht unbeachtet bleiben. Zum anderen werden in immer mehr Unternehmen PDMS zur Verwaltung von Produktdaten eingesetzt. Darin werden für jedes Produkt die relevanten Attribute nicht nur verwaltet, sondern auch eindeutig festgelegt. Letzteres erfolgt durch die Definition von Bedingungen, welche die für ein Attribut möglichen Ausprägungen einschränken. Abbildung 4.11 verdeutlicht dieses Vorgehen am Beispiel der beiden Attribute *Area* (Einsatzort) und *Position* (konkrete Position) für Kabeltüllen (Abbildung 4.11 links) im Automobilbau. Diese dienen beim Verlegen von Kabelleitungen oder Schläuchen im Innenraum eines Fahrzeugs zum Schutz bei scharfkantigen Durchbrüchen sowie zum Abdichten von Durchbrüchen. Dabei werden Kabel bzw. Schläuche nur in bestimmten Bereichen eines Fahrzeugs verlegt, was die möglichen Ausprägungen für den Ein-

satzort derartiger Kabeltüllen auf die Werte Motorraum (engine compartment), Fahrgastzelle (passenger compartment), Laderaum (trunk), Fahrgestell (chassis) und Motorblock (engine block) (vgl. Abbildung 4.11 oben) und ihre konkrete Position entsprechend auf die Ausprägungen A-Säule (A-pillar), B-Säule (B-pillar), Sitz (Seat), Radkasten (Wheel house) usw. (vgl. Abbildung 4.11 unten) beschränkt.

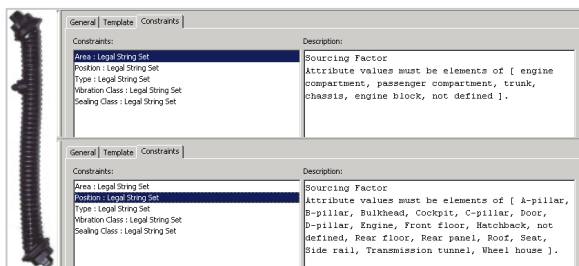


Abbildung 4.11. — Definition von Ausprägungen für die beiden Attribute *Area* (oben) und *Position* (unten) einer Kabeltülle (links) zur Verwendung in einem PDMS

Dokumentinformationen Primäre Quelle für Dokumentinformationen ist natürlich das Dokument selbst, dessen Verwaltung abhängig von der Softwarelandschaft eines Unternehmens und den gebräuchlichen Strukturen im Allgemeinen auf zwei Arten erfolgt. Zwar werden durchaus IT-Systeme wie DMS, PMS oder auch Versions- bzw. Konfigurationsmanagementsysteme eingesetzt, in denen die Dokumente oftmals zusätzlich mit Hilfe manuell gepflegter Metadaten beschrieben werden; der jedoch viel häufigere Anwendungsfall ist die Speicherung von Dokumenten in Verzeichnisstrukturen auf Netzlaufwerken, die entweder nur für den Autor selbst ($\hat{=}$ privates Netzlaufwerk) oder die an einem Projekt beteiligten Mitarbeiter ($\hat{=}$ Projektlaufwerk) zugänglich sind. Obwohl die Ablage der Dokumente dabei meist in vordefinierten Ordnerstrukturen erfolgt, ist ihre (Wieder-) Auffindbarkeit gerade im unternehmerischen Umfeld mit einer großen Anzahl von Projekten nicht immer garantiert.

Werkstoffdaten Die Verwaltung von Informationen über Werkstoffe findet heutzutage überwiegend in Datenbanken oder ERP-Systemen statt. Während letztere v. a. auf die Speicherung und Bereitstellung von beschaffungsrelevanten Daten wie Preise, Lieferzeiten, Mengen usw. fokussiert sind, geben DB eher Auskunft über die physikalischen und mechanischen Eigenschaften, indem sie deren Werte für Dichte, Schmelzpunkt, Festigkeit oder Härte enthalten. Aber auch Dokumente, insbesondere Simulations- und Testberichte sowie CAD-Modelle oder

Stücklisten, können abhängig von ihrem Inhalt als mögliche Informationsquellen für Werkstoffe in Frage kommen.

Personendaten Um Informationen über Personen zu erhalten, die im Umfeld eines Unternehmens agieren, sind diverse Ressourcen von Bedeutung, die jedoch davon abhängen, welcher Personengruppe eine Person zuzurechnen ist. So enthalten beispielsweise ERP-Systeme Daten zu Lieferanten und Kunden, die zur Abwicklung der Beschaffung und des Vertriebs von Produkten benötigt werden. Darüber hinaus werden für diese beiden Gruppen häufig eigene IT-Systeme zur Verwaltung und zur Pflege der zu ihnen bestehenden Beziehungen eingesetzt, so dass durch eine Anbindung dieser SRM- und CRM-Systeme relevante Daten daraus für eine Suche verwendet werden können. Informationen zu Unternehmensmitarbeitern hingegen findet man vorrangig in WfMS oder PMS, wie auch dem FORFLOW-Prozessnavigator, in denen die Verantwortlichkeiten bzw. Rollen für bestimmte Projekte und Aufgaben festgelegt werden. Allerdings handelt es sich hierbei eher um aufgabenspezifische Daten. Daher sollten zusätzlich Organigramme, welche die Einordnung eines Mitarbeiters in die Aufbauorganisation des Unternehmens veranschaulichen, und sogenannte Skill-Datenbanken, in denen sowohl harte (Kenntnisse, Sprachen, Abschlüsse, Zertifikate) als auch weiche Faktoren (Aussagen zu Teamfähigkeit, Belastbarkeit, ...) ein Profil des Mitarbeiters zeichnen [HZ00], zur Erfassung weiterer Kontextfaktoren herangezogen werden.

Projektinformationen Projekte werden primär mit Hilfe von PMS in Unternehmen verwaltet, so dass diese die Hauptquelle für eine Erfassung von Projektinformationen darstellen. Ferner werden zu jedem Projekt gemäß Kapitel 4.1.2 verschiedene Projektdokumente wie beispielsweise Projekt-, Termin- oder Ressourcenpläne angefertigt, die ebenfalls relevante und spezifische Informationen für ein Projekt enthalten.

Artefaktbeziehungen Welche Artefakte über welche Beziehungen miteinander verbunden sind, kann einerseits aus den diversen im Unternehmen verfügbaren Drittsystemen ermittelt werden. In PDMS werden beispielsweise nicht nur Produkte verwaltet; vielmehr können mit Hilfe von Relationsobjekten auch zugehörige Produktmodelle bei diesen hinterlegt werden (vgl. Kapitel 2.2.4). Ein weiteres Beispiel stellen ERP-Systeme dar, welche die Beziehung zwischen einem Produkt und dessen Lieferant(en) abbilden und somit die Erfassung von *Produkt* \leftrightarrow *Person*-Beziehungen ermöglichen. Andererseits ist aber auch zu berücksichtigen, dass Dokumente aufgrund ihres Inhaltes ebenfalls Aufschluss über existierende Beziehungen geben. So beschreibt beispielsweise eine Stückliste den Aufbau eines Produktes durch Auflistung sämtlicher Komponenten. Da es sich bei der Stückliste um ein Produktmodell handelt, können bereits implizit die beiden *Dokument* \leftrightarrow *Produkt*-Beziehungen (*Dokument* 'beschreibt' *Produkt* und

Produkt 'wird beschrieben in' Dokument gemäß Abbildung 4.10) erstellt werden. Zusätzlich lassen sich in diesem speziellen Fall aus der Produktstruktur auch noch verschiedene *Produkt* \leftrightarrow *Produkt* - Beziehungen extrahieren (siehe Kapitel 5.4.5).

Insgesamt zeigt sich somit, dass für die Erfassung von für eine Suche relevanten Informationen zwei wesentliche Quellen zu berücksichtigen sind. Zum einen handelt es sich hierbei um die im Unternehmen vorhandenen Drittsysteme, aus denen die Daten in strukturierter Form ausgelesen werden können. Zum anderen müssen aber auch Dokumente, welche in den meisten Fällen eher unstrukturierte Daten enthalten, indexiert werden, da gerade ihre Inhalte wichtige Informationen für Produktentwickler aufweisen. Aus diesem Grund wird für das Importieren der zu indexierenden Daten die Verwendung einer Containerdatei empfohlen, die beide Informationsquellen berücksichtigt und deren Aufbau im folgenden Kapitel erläutert wird.

4.4. Importieren von Artefaktinformationen

Obwohl mit XFML⁶⁴ seit dem Jahr 2002 ein von Peter van Dijck entwickeltes, offenes XML-basiertes Format zum Austausch von facettierten Daten existiert, ist dieses für eine Anwendung im LFRP-Framework nicht geeignet. Die Gründe hierfür sind zum einen in der Begrenztheit des Formates zu sehen, welches lediglich für eine Zuordnung von Facettenausprägungen zu einem einzigen Artefakttyp — speziell für die Zuordnung von behandelten Themen in einer Webseite zu dieser Webseite — konzipiert wurde. Zum anderen setzt Dijck voraus, dass die Daten bereits vor dem eigentlichen Import facettiert werden, was im vorliegenden Anwendungsfall gerade als eine der Aufgaben des LFRP-Frameworks gesehen wird.

Um folglich sowohl nicht bereits facettierte, artefaktbeschreibende Attribute in Form von einfachen Name / Wert-Paaren als auch Dokumente dem Indexierungsframework übergeben zu können, wird im Rahmen des LFRP-Frameworks die Erstellung einer Containerdatei von den Drittsystemen gefordert. Hierfür bietet sich die Verwendung des Zip-Dateiformates an, welches sich nicht nur dadurch auszeichnet, dass es eine komprimierte oder auch nicht-komprimierte Zusammenfassung von mehreren, zusammengehörenden Dateien oder Verzeichnisbäumen ermöglicht, sondern darüber hinaus den Vorteil bietet, dass die für eine Implementierung benötigten Klassenbibliotheken frei zur Verfügung stehen [Wik10f]. Allerdings sollte gerade im Unternehmensumfeld darauf geachtet werden, dass unter Umständen auch sehr große Datenmengen zu übertragen sind. Aus diesem Grund wurde im Rahmen der prototypischen Java-Implementierung des Frameworks die von der Enter AG⁶⁵ in Zürich entwickelte und

⁶⁴eXchangeable Faceted Metadata Language (<http://petervandijck.com/xfml>)

⁶⁵<http://www.enterag.ch>

über das Sourceforge-Portal von GeekNet.Inc kostenfrei beziehbare Java-Bibliothek `Zip64File`⁶⁶ genutzt, welche gegenüber dem im Java Development Kit (JDK) standardmäßig enthaltenen *java.util.zip*-Paket eine Übertragung von mehr als 65000 Einzeldateien bzw. einer Dateigröße von mehr als vier Gigabytes erlaubt.

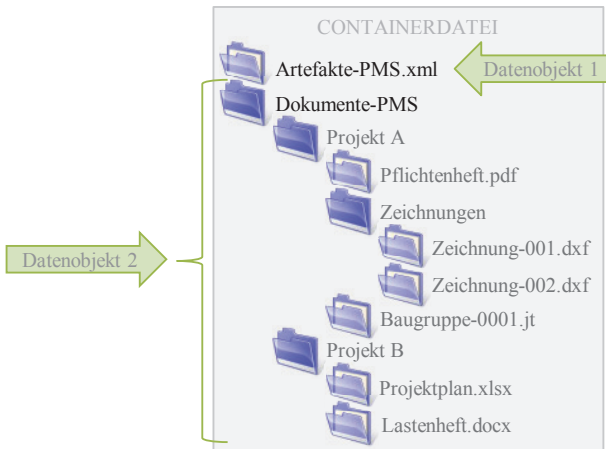


Abbildung 4.12. — Geforderte Struktur der Containerdatei für den Datenimport anhand eines Beispiels

Die Containerdatei selbst sollte dabei so gestaltet sein, dass mit ihr prinzipiell zwei Datenobjekte transportiert werden können, von denen gemäß Abbildung 4.12 eines für die Übertragung der strukturierten Attributdaten (Datenobjekt 1) und eines zur Übermittlung der Dokumente (Datenobjekt 2) verwendet wird. Während sich für Letzteres die Nutzung eines Verzeichnisbaumes anbietet, aus dem unter Umständen bei vordefinierter Strukturierung und Ordnerbenennung zusätzliche Kontextdaten ausgelesen werden können (z. B. Dokumenttyp *Zeichnung* für die Dateien *Zeichnung-001.dxf* und *Zeichnung-002.dxf* in Abbildung 4.12), wird die Definition der Attributdaten in Form einer XML-Datei (*Artefakte-PMS.xml* in Abbildung 4.12) gefordert. Ihre Struktur ist dabei der in Listing 4.1 dargestellten Document Type Definition (DTD) zu entnehmen, welche im vorliegenden Fall die beiden Elemente `<Informationsquelle/>` und `<Artefakte/>` für die Gültigkeit der Datei vorschreibt [NS06].

⁶⁶<http://sourceforge.net/projects/zip64file>

Listing 4.1 — Exemplarische DTD der XML-Importdatei: artefaktimport.dtd

```

<!-- Wurzelement -->
<ELEMENT Artefaktimport (Informationsquelle,Artefakte)>

<!-- Beschreibung der Organisationsdaten -->
<ELEMENT Informationsquelle (Zeitstempel)>
<!ATTLIST Informationsquelle
    quelleID ID #REQUIRED
    quelletyp (PDM|ERP|DMS|DB|PMS|Dokument) #REQUIRED>
<ELEMENT Zeitstempel (#PCDATA)>

<!-- Definition der zu importierenden Artefaktinformationen -->
<ELEMENT Artefakte (Artefakt+)>

<!-- Beschreibung eines Artefakts -->
<ELEMENT Artefakt (Metadatum+|Assoziation*)>
<!ATTLIST Artefakt
    artefakttyp (Produkt|Werkstoff|Dokument|Person|Projekt) #
    REQUIRED
    artefaktID ID #REQUIRED
    ssIdent ID #REQUIRED
    artefaktURI CDATA #REQUIRED>

<!-- Definition eines Metadatums als Name/Wert-Paar -->
<ELEMENT Metadatum (Name,Wert)>
<!ELEMENT Name (#PCDATA)>
<!ELEMENT Wert (#PCDATA)>

<!-- Beschreibung einer Artefaktbeziehung -->
<ELEMENT Assoziation (Assoziationselement+)>
<!ATTLIST Assoziation
    assoziationstyp (Werkstoff-verwendet-in-Produkt|
        Dokument-beschreibt-Produkt|
        Dokument-erstellt-in-Projekt|
        Dokument-erstellt-von-Person) #REQUIRED>
<ELEMENT Assoziationselement EMPTY>
<!ATTLIST Assoziationselement
    artefaktID ID #REQUIRED>

```

Zum einen wird mit Hilfe des Elementes `<Informationsquelle/>` die Angabe von organisatorischen Daten gefordert, die Auskunft darüber geben, aus welcher Quelle⁶⁷ (`quelleID` und `quelleTyp`) wann und zu welchem Zeitpunkt (`<Zeitstempel/>`) eine Information ausgelesen wurde. Diese Daten werden im Rahmen der Indexierung benötigt, um einerseits bei Fehlen von zwingend erforderlichen Daten eine manuelle Bearbeitung durch den Nutzer anhand einer entsprechend aufbereiteten Protokollierung der Daten gewährleisten zu können (vgl. Kapitel 5.4.1). Zum anderen muss bei Vorhandensein der gleichen Information aus unterschiedlichen Quellen deren Konsistenz und Aktualität sichergestellt werden, wofür ebenfalls auf diese Informationen zurückgegriffen wird (siehe Kapitel 6). Weiterhin dienen organisatorische Daten bei einer Indexierung zu festgelegten Zeitpunkten (= Pullmodus) für die Feststellung, welche Daten seit dem letzten Indexierungslauf zu berücksichtigen sind, da hier prinzipiell nur die ab dem letzten Extraktionszeitpunkt geänderten oder neu erstellten Daten indiziert werden müssen.

Das zweite Element `<Artefakte/>` dient schließlich der Beschreibung der im Drittsystem enthaltenen Artefaktinformationen, indem für jedes Artefakt ein Element `<Artefakt/>` mit vier Attributen und mindestens einem weiteren Subelement zu erzeugen ist. Dabei ist hinsichtlich der Attribute zum einen der `artefakttyp` des Artefakts anzugeben, welcher nur einen Wert aus der Gruppe der im Retrievalprozess zu differenzierenden Artefakttypen (hier: Produkt, Werkstoff, Dokument, Person oder Projekt) annehmen kann. Des Weiteren ist jedes Artefakt, wie in Kapitel 4.1 erläutert, durch eine `artefaktURI` und eine `artefaktID` zu charakterisieren. Letztere dient zwar der eindeutigen Identifikation von Artefakten, ist allerdings gerade im Hinblick auf eine erforderliche Prüfung von Zugriffsrechten (vgl. Kapitel 3.3.2) in den meisten Fällen alleine nicht ausreichend. Der Grund hierfür liegt darin, dass in vielen Unternehmen aufgrund der mit der Zeit gewachsenen Strukturen keine Konsolidierung und Vereinheitlichung der Daten über alle existierenden IT-Systeme hinweg stattgefunden hat. Demzufolge wird ein und dasselbe Artefakt, welches in verschiedenen AWS anhand unterschiedlicher Aspekte beschrieben wird, meist nicht zentral, sondern in systemspezifischen Datenbeständen verwaltet, wodurch es jeweils über eine andere systemspezifische Identifikationsnummer (ID) verfügt. Ein typisches Beispiel hierfür ist die Archivierung von Produktdaten. Während in einem PDMS vorrangig die geometrischen Eigenschaften beispielsweise einer Kabeltülle erfasst und gespeichert werden, findet man in einem ERP-System die beschaffungsrelevanten Informationen zu dieser. Dabei bezieht man sich zwar in beiden Fällen auf das gleiche Produkt (gekennzeichnet durch die Sachnummer); aus der internen Systemsicht betrachtet, handelt es sich jedoch um zwei unterschiedliche Produkte, die sich durch ihre systemspezifische ID unterscheiden. Will man also überprüfen, welche Zugriffsrechte ein Nutzer für ein Produkt in einem dieser Systeme besitzt, so ist eine Anfrage an das jeweilige System

⁶⁷Eine Quelle kann im Fall einer notwendigen Nachbearbeitung auch ein Dokument sein (vgl. Kapitel 5.4.1).

mit Hilfe der systemspezifischen ID zu stellen. Somit sollte bei einer dezentralen Datenverwaltung, von welcher in der vorliegenden Arbeit ausgegangen wird, zusätzlich zur **artefaktID** immer auch diese systemspezifische ID **ssIdent** mit angegeben werden. Daneben muss jedes Artefakt mindestens ein **<Metadatum/>**-Element besitzen, welches als Name / Wert-Paar sozusagen das für das Artefakt anwendbare Suchkriterium ($\hat{=}$ Facette) mit seinem zugehörigen Wert ($\hat{=}$ Facettenausprägung) beschreibt. Des Weiteren besteht die Möglichkeit mit Hilfe des Elementes **<Assoziation/>** Beziehungen zu anderen Artefakten zu beschreiben. Wird diese Option genutzt, so ist jede Beziehung durch einen, aus einer fest vordefinierten Menge von Beziehungsarten auszuwählenden **assoziationsstyp** und durch mindestens ein **<Assoziationselement/>** — das mit dem beschriebenen Artefakt in Beziehung stehende Artefakt — zu definieren.

Zur Verdeutlichung des beschriebenen Aufbaus zeigt Anhang C ein Beispiel einer gültigen XML-Datei zum Import von Artefaktinformationen aus einem PDMS.

5. Repräsentationserstellung — Informationsextraktion aus XML-Datei und Dokumenten

Der nächste Schritt in der Überführung der bereits existierenden Informationen in das IRS besteht darin, die mit Hilfe der zuvor beschriebenen Containerdatei importierten Artefakte in geeignete Repräsentationen — die in Kapitel 3.3.3 erläuterten Polyrepräsentationen — zu transformieren. Dazu steht ein *Indexierer* zur Verfügung, der gemäß Abbildung 5.1 die übergebene Containerdatei unter Verwendung verschiedener Module verarbeitet. Welche Module allerdings wann anzuwenden sind, hängt dabei vom Inhalt der Containerdatei ab. Je nachdem ob nur eines der in Kapitel 4.4 beschriebenen Datenobjekte oder beide enthalten sind, können drei Indexierungsszenarios unterschieden werden. Diese werden in den Kapiteln 5.2 bis 5.4 näher erläutert. Grundsätzlich müssen jedoch in allen drei Szenarios genau die Informationen indexiert werden, welche für die Facetten, die einem Nutzer im Retrievalprozess als Suchkriterien zur Verfügung stehen sollen, benötigt werden. Damit der *Indexierer* folglich entscheiden kann, welche Informationen zu berücksichtigen sind bzw. welche Facetten für welche Artefakttypen benötigt werden, sind diese im Vorhinein beispielsweise durch einen Fachvertreter des Unternehmens in Abstimmung mit den Anwendern des LFRP-Frameworks festzulegen. Dies erfolgt durch Definition eines Facettenschemas (Abbildung 5.1 rechts), dessen Aufbau und Inhalt im folgenden Kapitel 5.1 verdeutlicht wird.

5.1. Schema zur Definition der suchbaren Facetten und ihrer Abhängigkeiten

Wie in Kapitel 4 gezeigt wurde, erfordert nicht nur jeder Artefakttyp andere Facetten als Suchkriterien, sondern besitzt unter Umständen auch eine sehr umfangreiche Anzahl von als Facetten verwendbaren Attributen. Daher sollte es nicht das primäre Ziel sein, möglichst alle denkbaren Facetten in eine Suche einzubeziehen. Schließlich darf der Nutzer gemäß Kapitel 3.1 nicht überfordert werden, um eine Akzeptanz und damit eine regelmäßige Anwendung des Systems zu gewährleisten. Aus diesem Grund

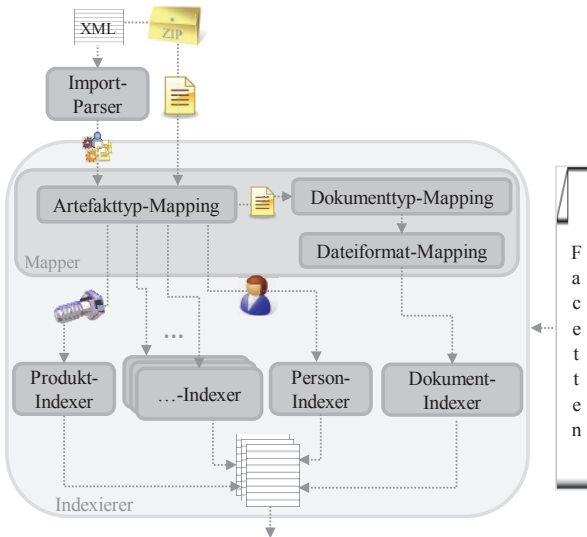


Abbildung 5.1. — Aufbau des Indexierers und vereinfachter Ablauf der darin stattfindenden Repräsentationserstellung unter Nutzung eines Facettenschemas

sollten nur die tatsächlich für einen Anwendungsbereich relevanten Facetten berücksichtigt werden. Allerdings sind diese wiederum nicht für alle Unternehmen gleich, so dass eine allgemeingültige Festlegung nicht sinnvoll ist. Daher wird mit der Definition eines Facettenschemas die Möglichkeit gegeben, das System an die spezifischen Bedürfnisse im Unternehmen und damit an die der konkreten Nutzergruppe anzupassen, indem man einerseits die Facetten selbst und andererseits die Zuordnung der Facetten zu den Artefakttypen in Form einer Artefakttyphierarchie festlegt. Aus diesem Grund setzt sich das zu definierende Facettenschema gemäß seines zugrunde liegenden XML-Schemas (siehe Anhang D) aus zwei miteinander verknüpften Elementen — einem `<facets/>` und einem `<entities/>`-Element — zusammen, auf die im Folgenden genauer eingegangen wird.

Definition der Facetten

Das zu Beginn des Facettenschemas anzugebende Element `<facets/>` dient der allgemeinen Definition der verwendbaren Facetten. Dabei ist jede einzelne Facette grundsätzlich durch ein Element `<facet/>` zu beschreiben, das sowohl für den Indexierungsprozess, den Retrievalprozess als auch für eine Darstellung der Facette auf der GUI

näher zu charakterisieren ist. Wie Listing 5.1 veranschaulicht, setzt sich eine derartige Facettendefinition aus diversen Subelementen und Attributen zusammen, von denen allerdings nur einige für das Indexieren der Artefaktinformationen relevant und damit in der vorliegenden Arbeit von Interesse sind⁶⁸. Diese können prinzipiell in die beiden Gruppen der obligatorischen und der fakultativen Charakteristika differenziert werden, wobei erstere für alle Facetten verpflichtend anzugeben sind (siehe rot markierte Schlüsselwörter in Listing 5.1). Dies gilt zum einen für die Typdeklaration einer Facette mittels des Attributs `facetType`, welches die Art der Ähnlichkeitsfacette als QbE oder QbT bzw. den Typ der Attributfacette als nominal, ordinal oder metrisch beschreibt. Zum anderen muss für jede Facette neben einer ID (Attribut `facetId`) auch deren Wertigkeit definiert sein, die angibt, ob eine Facette auch mehr als eine Ausprägung besitzen darf. Sind mehrere Ausprägungen erlaubt, so ist das Attribut `multiIndexable` auf den Wert `true` zu setzen. Falls nicht, muss es den Wert `false` erhalten.

Listing 5.1 — Inhalt eines `<facet/>`-Elementes gemäß `facetSchema.xsd` in Anhang C

```
<xs:complexType name="FacetDefinitionType">
  <xs:sequence>
    <xs:element name="visual" type="VisualType" minOccurs="0" maxOccurs="1" />
    <xs:element name="synset" type="ComparableTerms" minOccurs="0" maxOccurs="1" />
    <xs:element name="rule" type="CalculationRule" minOccurs="0" maxOccurs="1" />
  </xs:sequence>
  <xs:attribute name="facetId" type="xs:string" use="required" />
  <xs:attribute name="facetType" type="FacetType" use="required" />
  <xs:attribute name="multiIndexable" type="xs:boolean" use="required"/>
  <xs:attribute name="similarityType" type="SimilarityType" use="optional" />
  <xs:attribute name="documentType" type="DocumentType" use="optional" />
  <xs:attribute name="weightingEnabled" type="xs:boolean" />
</xs:complexType>
```

⁶⁸Zur Vollständigkeit und für ein besseres Verständnis zeigt Listing 5.1 auch die für den Retrievalprozess benötigten Attribute (dunkelblau markiert) sowie das für die Facettendarstellung erforderliche Subelement (hellblau markiert). Für weitere Details hierzu wird allerdings auf die Dissertation von Herrn Raiko Eckstein verwiesen.

Fakultative Charakteristika sind hingegen optional und müssen somit nicht notwendigerweise gesetzt werden (siehe orange markierte Schlüsselwörter in Listing 5.1). Allerdings können manche Facetten ohne diese zusätzliche Information nicht erzeugt werden, wie dies beispielsweise bei den in Kapitel 4.1.1 beschriebenen abgeleiteten Facetten *Zugfestigkeit* und *Streckgrenze* der Fall ist. Hier sind konkrete Berechnungsvorschriften festzulegen, weshalb derartige Facetten um ein `<rule/>`-Element erweitert werden können. Dieses Element ist gemäß des in Anhang D enthaltenen XML-Schemas selbst wiederum ein komplexer Datentyp, der durch ein `operator`-Attribut und maximal drei Rechenterme zu beschreiben ist. Letzteres resultiert aus der Tatsache, dass nicht nur die vier Grundrechenarten (Differenz, Summe, Produkt und Quotient) zu unterstützen sind, sondern dass auch die Möglichkeit zur Definition von *if-then-else*-Regeln sowie von Identitätsfunktionen gegeben sein muss. Folglich wird mit dem `operator`-Attribut die Art der Berechnungsregel festgelegt. Die Rechenterme sind im Anschluss jeweils als `<term/>`-Element anzugeben und können entweder eine Berechnungsvorschrift (`<rule/>`), eine Referenz auf den Wert einer anderen Facette (`<reference/>`) — eventuell versehen mit einer Einschränkung (Subelement `<constraint/>`) — oder einen konkreten numerischen (`<argumentDigit/>`) oder textuellen Wert (`<argumentText/>`) darstellen. Neben diesen Berechnungsvorschriften ist für die meisten Facetten schließlich noch eine Angabe notwendig, die dem *Indexierer* mitteilt, welche Begrifflichkeiten bzw. Wörter zur Erzeugung einer bestimmten Facette führen. Diese Festlegung ist notwendig, da oftmals diverse Begrifflichkeiten (unter Umständen auch in verschiedenen Sprachen) zur Beschreibung ein und desselben Sachverhaltes verwendet werden. So kann beispielsweise eine Information über die Produktfunktion sowohl durch die Wörter Funktion, Produktfunktion, Gesamtfunktion oder Hauptfunktion gekennzeichnet werden, welche alle die Bildung der Facette **Product Function** erfordern. Aus diesem Grund wird für manche Facetten ein sogenanntes Synset, d. h. eine Menge von Synonymen, Hyponymen, Hyperonymen usw., benötigt, welches gemäß Stock alle Worte, die den mit der Facette beschriebenen Aspekt verdeutlichen, zusammenfasst und als gleichrangige Elemente dieser Menge betrachtet [Sto07]. Obwohl im WWW u. a. verschiedene Synonymsammlungen teilweise frei (z. B. WordNet⁶⁹ für die englische und OpenThesaurus⁷⁰ für die deutsche Sprache), teilweise kostenpflichtig (z. B. GermaNet⁷¹ als deutsche Komponente des EuroWordNet) verfügbar sind und für diese Problemstellung prinzipiell zur Anwendung kommen könnten, ist zu berücksichtigen, dass sie einen nur sehr allgemeinsprachlichen Charakter aufweisen. Das bedeutet, dass fachspezifische Begrifflichkeiten, wie sie im vorliegenden Anwendungsfall benötigt werden, in diesen Sammlungen meist nicht berücksichtigt sind. Sie sind eher in einem unternehmensspezifischen Glossar zu finden, aus dem — sofern vorhanden — die zu einem Wort gehörende Vorzugsbenennung und damit die `facetId` ermittelt werden könnte. Allerdings hat sich in

⁶⁹<http://wordnet.princeton.edu>

⁷⁰<http://www.openthesaurus.de>

⁷¹<http://www.sfs.uni-tuebingen.de/lsd>

der Zusammenarbeit mit diversen Industriepartnern im Forschungsverbund gezeigt, dass nicht allgemein von der Existenz eines solchen Glossars ausgegangen werden kann. Aus diesem Grund wird im Rahmen des LFRP-Frameworks die Definition der für eine Facette relevanten Begrifflichkeiten im Facettenschema empfohlen, wozu das Subelement `<synset/>` zur Verfügung gestellt wird⁷².

Zur Verdeutlichung zeigt Listing 5.2 vier Beispiele für derartige Facettendefinitionen. So wird mit den unter (a) enthaltenen Angaben die Attributfacette **Artifact Type** beschrieben, deren Ausprägungen als Nominalwerte nicht in eine Rangfolge gebracht werden können und die für jedes Artefakt nur genau eine Ausprägung besitzen darf. Im Vergleich zu dieser stellt die Facette **Product Function** unter (b) ebenfalls eine nominalskalierte Attributfacette dar, die allerdings über mehrere Ausprägungen verfügen kann. Folglich ist ihr Attribut `multiIndexable` auf `true` gesetzt. Des Weiteren besitzt diese Facette ein `<synset/>`-Element, das festlegt, welchen im Rahmen einer Informationsextraktion erfassten Begrifflichkeiten diese Facette zuzuordnen ist. Gleiches gilt für die im Folgenden dargestellten metrischen Facetten **Tensile Strength** (c) und **Project Duration** (d), für die darüber hinaus Berechnungsvorschriften zu ihrer Ermittlung festgelegt sind.

Listing 5.2 — Konkrete Definition von Facetten anhand von vier Beispielen

```
(a) <facet facetId="artifactType" facetType="NOMINAL"
    multiIndexable="false" />

(b) <facet facetId="productFunction" facetType="NOMINAL"
    multiIndexable="true">
    <synset>
        <compTerm>Produktfunktion</compTerm/>
        <compTerm>Funktion</compTerm/>
        <compTerm>Gesamtfunktion</compTerm/>
        <compTerm>Hauptfunktion</compTerm/>
    </synset/>
</facet/>

(c) <facet facetId="tensileStrength" facetType="
    FUNCTION_INTERVAL"
    multiIndexable="false">
    <synset>
        <compTerm>Zugfestigkeit</compTerm/>
    </synset/>
```

⁷²Die in diesen `<synset/>`-Elementen definierten Begriffe können gleichzeitig für eine Erweiterung von Schlagwortanfragen im Retrievalprozess genutzt werden, um den Recall und damit die Ergebnisqualität zu verbessern.

```

<rule operator="PRODUCT">
  <term>
    <reference facetId="propertyClass">
      <constraint constraintOperator="EQUAL">
        <!-- 1.Zahl der Festigkeitsklasse = max. 2
              Ziffern
              vor dem Punkt -->
        <regExpression>\d{1,2}\.<regExpression/>
      <constraint/>
    <reference/>
  </term>
  <term>
    <argumentDigit>100</argumentDigit>
  </term>
</rule>
</facet>

```

```

(d) <facet facetId="projectDuration" facetType="
    FUNCTION_INTERVAL"
    multiIndexable="false">
      <synset>
        <compTerm>Projektdauer</compTerm>
        <compTerm>Zeitdauer</compTerm>
        <compTerm>Geplante Zeitdauer</compTerm>
      </synset>
      <rule operator="DIFFERENCE">
        <term>
          <reference facetId="finishDate" />
        </term>
        <term>
          <reference facetId="startDate" />
        </term>
      </rule>
    </facet>

```

Definition der Artefakttyphierarchie

Da eine alleinige Definition der auswählbaren Facetten für die Umsetzung einer Suche nicht ausreicht, muss im zweiten Teil des Facettenschemas die für das LFRP-Framework gültige Artefakttyphierarchie definiert werden. Sie beschreibt, welche Facetten für welche Artefakttypen gültig sind und folglich bei einer Suche anwendbar

sein sollen. Damit lassen sich einerseits im Rahmen der Indexierung die richtigen Facetten für ein Artefakt erstellen und andererseits während des Retrievalprozesses eine dynamische Facettenbereitstellung, bei der nur die für einen Artefakttyp gültigen Facetten zur weiteren Selektion angeboten werden, gewährleisten. Letzteres erfordert allerdings zusätzlich für jedes Artefakt die Definition und Indexierung einer Facette mit `facetId`=“artifactHierarchyPath“, welche für die Suchkomponente die Position des Artefakts in der Artefakttyphierarchie beschreibt. Dabei ist zu beachten, dass ein Artefakt auch anhand mehrerer Diskriminatoren spezialisiert werden kann, weshalb durchaus mehrere Hierarchiepfade für ein Artefakt existieren können. Ein Beispiel hierfür sind Produkte, die abhängig von ihrer Produktgruppe und ihrer Beschaffungsart unterschiedliche Facetten erfordern (siehe Kapitel 4.1.1). Letztere sind während des Retrievalprozesses dynamisch, durch Bilden der Vereinigungsmenge der zu den Pfaden gehörigen Facettenmengen, zu ermitteln. Folglich ist das Attribut `multiIndexable` dieser Facette mit dem Wert `true` zu versehen. Unabhängig davon erfolgt die Definition der Artefakttyphierarchie selbst unter Nutzung des Elements `<entities/>`, welches für jeden einzelnen Artefakttyp bzw. seine Subklassen wiederum ein `<entity/>`-Element enthält. Gemäß Listing 5.3 sind diese durch drei Attribute zu beschreiben, von denen insbesondere das `root`-Attribut hervor zu heben ist. Dieses Attribut dient zur Kennzeichnung des Wurzelementes der Artefakttyphierarchie und ist folglich bei diesem — üblicherweise der Superklasse *Artefakt* — auf den Wert `true` zu setzen. Bei allen anderen wird es automatisch mit dem Defaultwert `false` belegt.

Listing 5.3 — Inhalt eines `<entity/>`-Elementes gemäß `facetSchema.xsd` in Anhang C

```
<xs:complexType name="EntityType">
  <xs:sequence>
    <xs:element name="facet" type="FacetRefType" minOccurs
      ="0"
      maxOccurs="unbounded" />
    <xs:element name="relationfacet" type="
      RelationFacetType"
      minOccurs="0" maxOccurs="unbounded" />
    <xs:element name="dependency" type="DependencyType"
      minOccurs="0"/>
    <xs:element name="parent-entity">
      <xs:complexType>
        <xs:attribute name="entityId" type="xs:string"
          use="required" />
      </xs:complexType>
    </xs:element>
  </xs:sequence>
```



```

<xs:attribute name="entityId" type="xs:string" use="
  required" />
<xs:attribute name="name" type="xs:string" use="required"
  />
<xs:attribute name="root" type="xs:boolean" default="
  false" />
</xs:complexType>

```

Daneben muss jedes `<entity/>`-Element mindestens ein Subelement zur Beschreibung der Hierarchiebeziehungen besitzen (Subelement `<parent-entity/>`), welches eine Referenz auf die entsprechende Superklasse enthält. Dies ist notwendig, um die Gültigkeit von Facetten einer Superklasse auch für eine Subklasse deutlich machen zu können. Schließlich sollen beispielsweise die in Listing 5.4 für Dokumente allgemein gültigen Facetten (vgl. `facetIds` „filetype“, „degreeOfConcreteness“ usw.) auch für die beiden definierten Subklassen der technischen Zeichnungen und CAD-Modelle bei einer Suche verfügbar sein. Die tatsächliche Festlegung der für ein `<entity/>`-Element gültigen Facetten erfolgt durch die Angabe von `<facet/>`-Elementen, wobei auch hier grundsätzlich zwischen obligatorischen und fakultativen Facetten zu differenzieren ist. Stellt eine Facette einen Diskriminator gemäß Kapitel 4.1 dar, welcher eine Spezialisierung in Subklassen definiert, so handelt es sich hierbei um eine obligatorische Facette (vgl. `facetId`=“documenttype“ in Listing 5.4). Das bedeutet, dass diese Facette für alle Artefakte des jeweiligen Artefakttyps gesetzt werden muss, da sonst keine eindeutige Zuordnung etwaiger weiterer Facetten möglich ist. Aus diesem Grund ist, wie in Listing 5.4 gezeigt, für jede Facette über das Attribut `required` anzugeben, ob diese gesetzt werden muss (Wert = `true`) oder nicht (Wert = `false`). Falls obligatorische Facetten existieren, sind diese im Anschluss durch ein `<dependency/>`-Element näher zu charakterisieren, indem definiert wird, welche Ausprägung dieses Diskriminators zu einer bestimmten Subklasse führt. So sind nach Listing 5.4 alle Dokumente mit der Ausprägung `TECHNICAL_DRAWING` für den Dokumenttyp als technische Zeichnungen weiter zu behandeln, die über zusätzliche Facetten wie beispielsweise die **Zeichnungsnummer** verfügen. Des Weiteren können für jeden Artefakttyp bzw. seine Subklassen auf Basis des in Kapitel 4.2 erläuterten Beziehungsnetzwerks mit Hilfe des Elementes `<relationfacet/>` Beziehungsfacetten definiert werden, die während des Retrievalprozesses eine Einschränkung der Ergebnismenge auf der gleichen Artefakttypenebene ermöglichen (z. B. Dokumentenmenge anhand des Autors verfeinern) oder die Suche auf eine andere Artefakttypenebene verlagern (z. B. die in Produkten verwendeten Materialien finden). Wie diese Beziehungsfacetten jedoch schließlich auf der GUI umgesetzt werden können, ist nicht Gegenstand der vorliegenden Arbeit⁷³.

⁷³Für nähere Informationen wird auf die Dissertation von Herrn Raiko Eckstein am gleichnamigen Lehrstuhl verwiesen.

Listing 5.4 — Ausschnitt einer Artefakttyphierarchie mit exemplarischer Definition des Artefakttyps *Dokument* und seinen beiden Subklassen *Technische Zeichnung* und *CAD-Modell*

```

<entity entityId="document" name="Document">
  <facet required="true" facetId="documenttype" />
  <facet required="false" facetId="file_name" />
  <facet required="false" facetId="filetype" />
  <facet required="false" facetId="degreeOfConcreteness" />
  <facet required="false" facetId="purposeOfUse" />
  <facet required="false" facetId="degreeOfCrossLinking" />
  <facet required="false" facetId="developmentStatus" />
  <facet required="false" facetId="document_lifecycle_state"
    />

  <!-- relations to other layers -->
  <relationfacet required="false" facetId="author"
    toEntityId="person" toEntityFacetId="
      surname"/>

  <!-- hierarchy mapping -->
  <dependency>
    <from facetId="documenttype" />
    <mapping>
      <value>TECHNICALDRAWING</value>
      <to>drawing</to>
    </mapping>
    <mapping>
      <value>CAD.MODEL</value>
      <to>cadModel</to>
    </mapping>
  </dependency>

  <parent-entity entityId="artifact" />
</entity>

<entity entityId="drawing" name="Technical Drawing">
  <facet required="false" facetId="drawing_number" />
  <facet required="false" facetId="2dGeometry"/>
  <facet required="false" facetId="2dTopology"/>
  <parent-entity entityId="document" />
</entity>

```

```
<entity entityId="cadModel" name="CAD Model">
  <facet required="false" facetId="3dGeometry"/>
  <parent-entity entityId="document" />
</entity>
```

5.2. Szenario 1: Import einer XML-Datei

Im Fall des Importszenarios 1 werden nur beschreibende Metadaten der in einem Drittsystem verwalteten Artefakte in strukturierter Form, d. h. mittels einer gemäß Kapitel 4.4 erstellten XML-Datei, an den *Indexierer* übergeben. Demzufolge ist hier in einem ersten Schritt diese XML-Datei zu parsen (Modul *Import-Parser* in Abbildung 5.1), wozu beispielsweise die frei verfügbare XML-Parser-Implementierung Apache Xerces⁷⁴ eingesetzt werden kann. Als Ergebnis dieses Vorgangs entsteht eine Menge von zu indexierenden Artefakten, von denen jedes als eigenständiges Objekt durch die zuvor beschriebenen vier Attribute **artefakttyp**, **artefaktID**, **ssIdent** und **artefaktURI** eindeutig charakterisiert ist und neben einer Menge von Metadaten in Form von Name-Wert-Paaren auch organisatorische Informationen und — sofern in der XML-Datei definiert — Assoziationen besitzt. Diese Charakterisierung ermöglicht im weiteren Verlauf eine Differenzierung der Artefakte. Hierzu wird das in Abbildung 5.1 enthaltene *Mapper*-Modul aufgerufen, welches zunächst eine Unterscheidung anhand des Artefakttyps vornimmt (Schritt *Artefakttyp-Mapping* in Abbildung 5.1). Abhängig vom Artefakttyp wird somit unter Nutzung des Entwurfsmusters *Fabrikmethode* der für ein Artefakt zuständige Indexer ermittelt (vgl. erste Ebene in Abbildung 5.2), welcher für die eigentliche Indexierung zuständig ist. An diesen wird im Anschluss ein für das jeweilige Artefakt erstellte und von der Superklasse **IndexableArtifact** abgeleitete Subklassenobjekt (z. B. ein **IndexableProduct**- oder **IndexablePerson**-Objekt) zur Verarbeitung weitergeleitet. Demzufolge werden Produkte mit ihren Daten an einen **ProductIndexer**, Personen mit ihren Informationen an einen **PersonIndexer** usw. zur Repräsentationserstellung übergeben. Eine Ausnahme hierbei stellen allerdings Artefakte vom Typ *Dokument* dar, welche vor der eigentlichen Verarbeitung gemäß Abbildung 5.1 zwei weitere Differenzierungsschritte — ein *Dokumenttyp-Mapping* und ein *Dateiformat-Mapping* — zu passieren haben. Da allerdings bei dem hier betrachteten Importszenario davon ausgegangen wird, dass das jeweilige Drittsystem keine Dokumente verwaltet, werden die Gründe für die Notwendigkeit dieser beiden Module in Kapitel 5.3 erläutert. Jeder einzelne Indexer überführt schließlich die übergebenen Daten in eine Reihe von Facetten. Diese stellen die in einer Polyrepräsentation enthaltenen Repräsentationen dar und sind beim alleinigen Import einer XML-Datei auf Attribut- und Beziehungsfacetten

⁷⁴<http://xerces.apache.org>

beschränkt. Folglich wird für jede Assoziation eine Beziehungsfacette in Form eines **ArtifactRelationship**-Objektes und für jedes Metadatum eine Attributfacette erzeugt. Ersteres enthält gemäß den importierten Daten die ID des Artefakts, zu dem eine Beziehung besteht, sowie die Angabe, um welche Art von Beziehung es sich handelt und zu welcher Artefakttypebene die Beziehung besteht. Was die Erzeugung der Attributfacetten betrifft, so muss eine Zuordnung der Metadaten zu den entsprechenden im Facettenschema definierten Facetten erfolgen. Unter der in Kapitel 5.1 gemachten Annahme, dass die für eine Facette relevanten Begrifflichkeiten im Facettenschema definiert werden, erfolgt dies unter Nutzung der in den `<synset/>`-Elementen definierten Wörter. Durch einen Vergleich dieser mit dem Namen des jeweils betrachteten Metadatums kann die entsprechende Facette identifiziert und gebildet werden. Diese wird durch ein **ExtractedFacetData**-Objekt repräsentiert, das den aus dem Facettenschema ermittelten Facettennamen ($\hat{=}$ dem Attribut *facetId* des `<facet/>`-Elementes) als Attribut und die zugehörige(n) Ausprägung(en) als **ExtractedFacetValue**-Objekt(e) — jedes gekennzeichnet durch die organisatorischen Daten der Importdatei (= Informationsquelle und Zeitstempel der Extraktion) — besitzt. Für den Fall, dass während des Vergleichs keine Facette für ein Metadatum im Schema gefunden wird, sollte die im Facettenschema die Textsuche (QbT) repräsentierende Facette mit dem Wert des Metadatums als Facettenausprägung hinzugefügt werden. Obwohl während des Retrievalprozesses zwar keine spezielle Facette zur Filterung nach diesem Wert zur Verfügung steht, kann auf diese Weise ein Datenverlust vermieden und eine Auffindbarkeit über eine Textsuche gewährleistet werden.

Weiterhin ist zu berücksichtigen, dass bei manchen Suchkriterien eine Festlegung darüber erfolgen muss, in welcher Form deren Werte zu importieren sind. So sollten beispielsweise geographische Angaben, wie der Sitz eines Zuliefererunternehmens, nicht nur aus einer einzelnen Ortsangabe bestehen. Aufgrund der Globalisierung interessiert hier vielmehr die Zugehörigkeit dieses Ortes zu einem Kontinent, einem Land und /oder einer Region, um beispielsweise im Fall der Suche nach einem Lieferanten für ein Produkt zwischen nahe und weit entfernten Lieferanten differenzieren zu können. Aus diesem Grund werden geographische Angaben meist durch hierarchische Pfadangaben mit vordefinierten Trennzeichen (z. B. „Europa/Deutschland/Bayern/Bamberg“) beschrieben, die sozusagen implizit eine Hierarchie bzw. Abfolge von Unterkategorien definieren. Werden derartige Pfadangaben beim Import der Metadaten verwendet, so ist abhängig von der konkreten Anwendung festzulegen, wie diese im Rahmen der Indexierung zu verarbeiten sind. Dabei sind für eine Filterung anhand einer derartigen Ortsangabe prinzipiell zwei Möglichkeiten denkbar. Während man bei der ersten Möglichkeit für jede Unterkategorie eine eigene Facette auf der GUI zur Verfügung stellt, beschränkt sich die zweite Möglichkeit auf die Anwendung einer einzigen Facette. Letztere wird gemäß Hearst et al. [HEE⁺02] als *hierarchische Facette* bezeichnet, wobei sich mit jeder Selektion die Ausprägungen und damit auch die Bedeutung der Facette verändern. Wie in Abbildung 5.3 unter a) verdeutlicht wird,

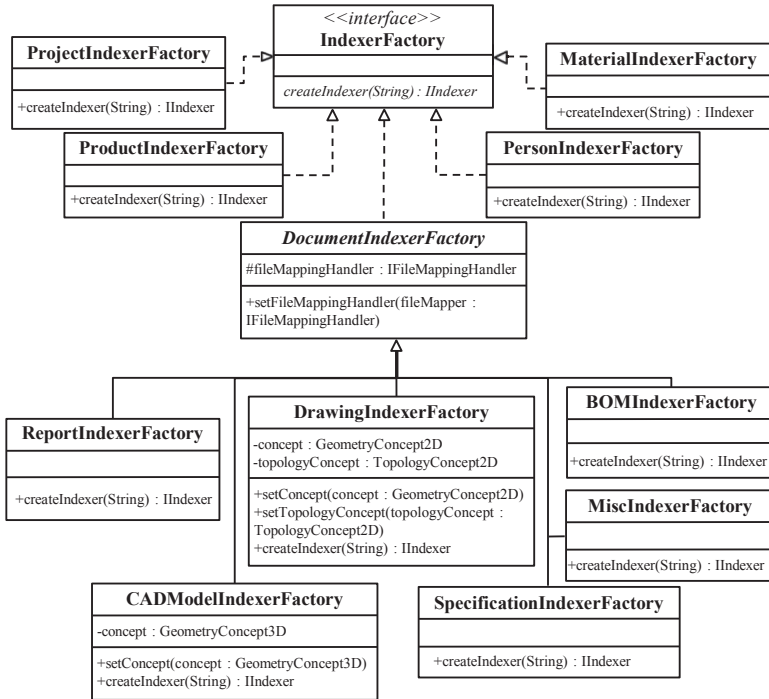


Abbildung 5.2. — Struktur der vom Interface `IndexerFactory` abgeleiteten Subklassen: Differenzierung nach dem Artefakttyp (erste Ebene) und Differenzierung nach dem Dokumenttyp (zweite Ebene)

beziehen sich die Ausprägungen bei erstmaliger Anwendung einer derartigen Facette (hier Facette `Location`) auf die oberste Ebene der Pfadangabe (hier die Kontinente). Sobald auf dieser Ebene eine Selektion erfolgt, sind die Werte der nachfolgenden Ebene (hier die Länder des ausgewählten Kontinents) als Ausprägungen der Facette anzugeben. Gleiches gilt für eine dritte Selektion, die

bei Auswahl des Landes `Deutschland` zur Anzeige der deutschen Orte als Ausprägungen führt. Um diese dynamische Änderung der Facettenausprägungen während des Retrievalprozesses zu ermöglichen, sind im Rahmen der Indexierung zwei unterschiedliche Vorgehensweisen möglich. Die einfachere der beiden Möglichkeiten ist die von Hatcher [Hat07] vorgeschlagene Speicherung des gesamten Pfades als Ausprägung

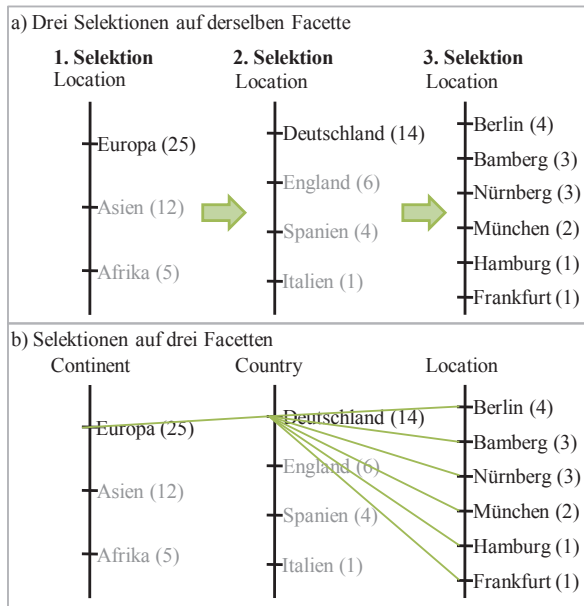


Abbildung 5.3. — Möglichkeiten zur Filterung anhand einer Ortsangabe: a) Anwendung einer einzigen hierarchischen Facette; b) Anwendung von drei eigenständigen Facetten

einer einwertigen Facette. Allerdings würde dieses Vorgehen während des Retrievalprozesses einen erhöhten Aufwand zur Ermittlung der Selektionsmöglichkeiten auf den oberen Ebenen (hier auf der Ebene der Kontinente und der Länder) erfordern. Schließlich sind hierzu alle Ausprägungen zunächst auf die entsprechende Selektionsebene zu transformieren, indem man die Pfadangaben beispielsweise mit Hilfe regulärer Ausdrücke bis auf die jeweilige Ebene reduziert. Da eine derartige Reduktion aber durchaus auch zu identischen Ausprägungen und damit zu Duplikaten führen kann, sind diese in einem anschließenden Schritt zu identifizieren und deren zugehörige Anfragevorschauren zu aggregieren. Dieser Ermittlungsaufwand lässt sich verringern, wenn man die von Hostetter [Hos08] vorgeschlagene Indexierung mehrerer Pfadangaben pro Artefakt anwendet. Anstatt eine einzige Pfadangabe „Europa/Deutschland/Bamberg“ zu nutzen, empfiehlt er, diese in mehrere nummerierte Pfadangaben der Form „0/Europa“, „1/Europa/Deutschland“ und „2/Europa/Deutschland/Bamberg“ zu zerlegen und als Ausprägungen einer mehrwertigen Facette zu speichern. Auf diese Weise lassen sich zumindest die für die jeweiligen Selektionsschritte erforderlichen Ausprä-

gungen einfacher und v. a. schneller ermitteln. Insgesamt betrachtet ist allerdings zu berücksichtigen, dass bei Verwendung von nur einer einzigen Facette, die einzelnen getätigten Selektionsschritte in den parallelen Koordinaten visuell verloren gehen. Da sich mit jeder Selektion die Facettenausprägungen und deren Semantik verändern, ist, wie Abbildung 5.3 unter a) zeigt, eine Nachvollziehbarkeit und somit auch eine nachträgliche Modifikation einer Anfrage nur sehr schwer möglich. Aus diesem Grund wird die in Abbildung 5.3 unter b) dargestellte Möglichkeit der Verwendung mehrerer Facetten bevorzugt — auch deshalb, weil die Bildung der einzelnen Facetten ohne großen Zusatzaufwand automatisch erfolgen kann und während des Retrievalprozesses keine gesonderten Berechnungen notwendig sind. Hierzu ist lediglich das in Anhang D definierte Facettenschema dahingehend zu erweitern, dass jede Facette mit einem Attribut **hierarchical** versehen wird, das entweder defaultmäßig den Wert **false** besitzt oder mit dem Wert **true** die Existenz eines Subelements **<subfacets/>** angibt. Dieses Subelement enthält die einzelnen Unterkategorien als **<subfacet/>**-Elemente, wobei jedes mit der zugehörigen **facetID** und einer numerischen Angabe der Pfadposition mittels des Attributs **level** charakterisiert wird. Eine derartige Definition, wie sie beispielhaft in Listing 5.5 gezeigt ist, ermöglicht schließlich eine automatische Zerlegung einer importierten Pfadangabe (z. B. „Europa/England/London“) in einzelne Werte und die Bildung von entsprechenden Facetten für diese Werte (Facette **Continent** mit dem Wert „Europa“, Facette **Country** mit dem Wert „England“ und Facette **Location** mit dem Wert „London“).

Listing 5.5 — Exemplarische Definition einer Facette für Ortsangaben

```
<facet facetId="location" facetType="NOMINAL" multiIndexable
    ="false" hierarchical="true">
  <synset>
    <compTerm>Ort</compTerm/>
    <compTerm>Firmensitz</compTerm/>
  </synset/>
  <subfacets>
    <subfacet facetId="continent" level="0" />
    <subfacet facetId="country" level="1" />
    <subfacet facetId="location" level="2" />
  </subfacets/>
</facet/>
```

Auf den ersten Blick könnte man annehmen, dass derartige Pfadangaben im Bereich der technischen PE insbesondere auch zur Beschreibung der Produktgruppe eines Produktes verwendet werden können. Schließlich können auch hier Angaben wie „Dichtung/O-Ring“, „Dichtung/Flachdichtung“, „Schraube/Kopfschraube“ oder „Schraube/Stiftschraube“ gebildet werden. Allerdings ist zu berücksichtigen, dass diese Pfadangaben nicht durch einheitliche Unterkategorien beschreibbar sind, wie dies

bei geographischen Angaben der Fall ist. So werden Dichtungen nämlich anhand des Dichtungstyps und Schrauben nach ihrer Schraubenform differenziert. Folglich wären bei der Definition der `<subfacet/>`-Elemente mehrere Facetten — im Beispiel die Facette für den Dichtungstyp und die für die Schraubenform — mit dem gleichen Wert für das Attribut `level` zu versehen, was eine eindeutige Zuordnung der Werte zu Facetten erschwert. Da auch die zuerst beschriebene Möglichkeit der Verwendung einer einzelnen Facette aufgrund der erläuterten Nachteile nicht empfohlen wird, sollte insbesondere in diesem Fall auf die Verwendung von Pfadangaben verzichtet und statt dessen auf die Nutzung von Name-Wert-Paaren für den Import der Informationen zurückgegriffen werden.

5.3. Szenario 2: Import von XML-Datei und Dokumentverzeichnis

Werden mit der Containerdatei beide Datenobjekte, d. h. XML-Datei und Dokumentverzeichnis übergeben, so wird für das weitere Vorgehen davon ausgegangen, dass jedes im Dokumentverzeichnis enthaltene Dokument auch mit Hilfe von Metadaten im Drittsystem verwaltet wird. Folglich muss sich in der XML-Datei für jedes Dokument ein entsprechendes `<Artefakt/>`-Element finden lassen. Um dies zu gewährleisten, wird vorausgesetzt, dass die Pfadangabe zur Beschreibung der `artefaktURI` mit dem Pfad zur Dokumentdatei im Verzeichnisbaum übereinstimmt. Durch ihren Vergleich kann somit für jede Dokumentdatei das zugehörige Artefakt identifiziert werden. Zusätzlich wird die Annahme zugrunde gelegt, dass für jedes Dokument dessen Dokumenttyp bekannt ist. Dies kann entweder durch eine spezielle Strukturierung des Verzeichnisbaumes (vgl. Abbildung 4.12 in Kapitel 4.4) oder mittels eines Metadatums mit der entsprechenden Information, wie dies beispielsweise bei manchen PMS, DMS oder auch dem FORFLOW-Prozessnavigator gegeben ist, erreicht werden. Diese Information stellt eine notwendige Bedingung dar, um eine Verarbeitung der Dokumentinhalte in der richtigen Art und Weise gewährleisten zu können. Schließlich ist zu berücksichtigen, dass sich Dokumente zwar allgemein durch das zu ihrer Speicherung verwendete Dateiformat⁷⁵ unterscheiden; diese Information alleine reicht jedoch nicht aus, um die zur Verarbeitung entwicklungsspezifischer Dokumentinhalte richtige Arbeitsweise des `DocumentIndexer` zu garantieren. Der Grund hierfür ist, dass Dokumente zwar häufig im selben Dateiformat vorliegen, aber völlig unterschiedlichen Zwecken im PEP dienen und damit einen voneinander abweichenden Informationsgehalt aufweisen. Als Beispiele hierfür dienen u. a. Pflichtenhefte und Testberichte, die als DOC/DOCX-Dokumente zwar beide ein Produkt beschreiben, dabei aber

⁷⁵Besonders häufige, in der industriellen Praxis verwendete Dateiformate sind das Portable Document Format (PDF), das Document-Format von MS Word (DOC/DOCX) und das Extensible Spreadsheet-Format von MS Excel (XLS/XLSX).

jeweils andere Produktaspekte fokussieren. Demzufolge sind zur Initialisierung des **DocumentIndexer** Informationen über den Dokumenttyp als auch das Dateiformat nötig (siehe hierzu auch Kapitel 5.4). Letzteres kann beispielsweise als Metadatum in der XML-Datei enthalten sein oder direkt aus dem Verzeichnispfad ermittelt werden, welcher üblicherweise am Ende die Form *Dateiname.Dateiformat* aufweist.

Sind die genannten Annahmen erfüllt, lassen sich die Artefaktinformationen durch eine Verarbeitung der Datenobjekte in der Reihenfolge 1. *XML-Datei* → 2. *Dokumentverzeichnis* indexieren. Hierzu erzeugt das in Abbildung 5.1 dargestellte Modul *Import-Parser* gemäß Kapitel 5.2 die Menge der zu indexierenden Artefakte, wobei die ein Dokument repräsentierenden Artefakte um ihre zugehörige, im Dokumentverzeichnis enthaltene Datei in Form eines **ContextualFile**-Objektes ergänzt werden. Diese Artefaktobjekte werden im Anschluss mittels des *Mapper*-Moduls an ihren jeweils zuständigen Indexer weitergeleitet, welcher die Erstellung der entsprechenden Polyrepräsentationen übernimmt. Auf diese Weise wird jedes **IndexableArtifact**-Objekt mindestens mit Attributfacetten und gegebenenfalls mit Beziehungsfacetten versehen. Zusätzlich führt die Verarbeitung des Dokumentinhaltes bei einem Dokument zu einer Erweiterung seiner Polyrepräsentation um weitere Facetten, weshalb im folgenden Kapitel 5.4 eine detaillierte Betrachtung dieser dokumentspezifischen Inhaltsverarbeitung erfolgt.

5.4. Szenario 3: Import eines Dokumentverzeichnisses

Das dritte Importszenario liegt dann vor, wenn die importierte Containerdatei lediglich einen Verzeichnisbaum mit Dokumenten enthält, so dass keine expliziten Metadaten für die Dokumente existieren. Dieser Fall ist v. a. dann gegeben, wenn einfache Dateiverwaltungssysteme, wie beispielsweise der bei Microsoft Windows-Betriebssystemen verfügbare Explorer, zur Archivierung der Dokumente eingesetzt werden. Folglich muss hier für jedes Dokument mindestens die Definition des zugehörigen Dokumenttyps — beispielsweise über eine entsprechende Ordnerbenennung — gewährleistet sein. Ist dies gegeben, kann der Verzeichnisbaum rekursiv durchlaufen werden, wobei jede identifizierte Dokumentdatei als Artefakt mit dem Artefakttyp *Dokument*, der Datei als **ContextualFile**-Objekt und sämtlichen weiteren verfügbaren Informationen wie dem Verzeichnispfad, dem Dateinamen, dem Dateiformat und dem Dokumenttyp als Metadaten an das *Mapper*-Modul übergeben wird. Dieses Modul erstellt gemäß dem in Kapitel 5.2 erläuterten Vorgehen zunächst ein **IndexableDocument**-Objekt, welches mit einer eindeutigen **artefaktID**, die gemäß Kapitel 4.1.2 als UUID generiert werden sollte, zu versehen ist. Anschließend leitet es das erzeugte **IndexableDocument**-Objekt an den für Dokumente zuständigen **DocumentIndexer** weiter, welcher allerdings für die richtige Verarbeitung des Dokumentes entsprechend zu initialisieren ist. Dazu wird im Schritt *Dokumenttyp-Mapping* abhängig vom jeweiligen Dokumenttyp

das entsprechende, für den Dokumenttyp verantwortliche und von der Superklasse `DocumentIndexerFactory` abgeleitete Subklassenobjekt ermittelt (siehe zweite Ebene in Abbildung 5.2). Dieses wiederum erzeugt mit der vom Interface geerbten Methode `createIndexer(String)` im folgenden *Dateiformat-Mapping* den `DocumentIndexer`. Dabei wird diesem ein `DocumentHandler` zugewiesen (vgl. Abbildung 5.4), dessen `processInput(InputStream, String, ExtractionSource)`-Methode das Dateiformat als `String` übernimmt und somit eine dateiformatspezifische Verarbeitung des jeweiligen Dokumentes ermöglicht. Hierfür können dem `DocumentHandler` zusätzlich eventuell benötigte Ähnlichkeitskonzepte zugewiesen werden (vgl. `CADDocumentHandler` in Abbildung 5.4). Die Erzeugung der Abhängigkeiten zwischen den nicht nur hierfür, sondern für das gesamte LFRP-Framework benötigten Objekten, wird dabei durch die Verwendung des Open Source-Frameworks Spring von SpringSource (bzw. VMware)⁷⁶ erleichtert. Da Spring das Prinzip der sogenannten *Dependency Injection* unterstützt, lassen sich alle erforderlichen Objekte extern in einer Datei namens *applicationContext.xml* als normale Java-Objekte, sogenannte Plain Old Java Objects (POJOs) oder auch Spring-Beans, spezifizieren. Ihnen werden die abhängigen Objekte mittels einer einfachen XML-Konfiguration zugewiesen, was eine ausprogrammierte Erzeugung der Objekte unnötig macht und die Flexibilität der Anwendung erhöht [Wol10]. Somit können je nach Initialisierung des `DocumentIndexer`

- unterschiedliche Parser eingesetzt,
- unterschiedliche Informationen extrahiert und damit schließlich auch
- unterschiedliche Facetten erzeugt werden.

Wie eine Verarbeitung im Detail aussieht, wird aufgrund der großen Anzahl relevanter Dokumenttypen im Folgenden anhand ausgewählter Beispiele erläutert. Um hierbei eine möglichst übergreifende und produktneutrale Abdeckung des PEP zu erreichen, erfolgte die Auswahl entlang der für den PEP charakteristischen Prozessphasen. Da diese jedoch abhängig vom verwendeten VM v. a. in ihrer Anzahl und ihrem Detaillierungsgrad variieren, im Prinzip aber alle das gleiche Vorgehen vom Groben zum Detail abbilden (siehe Tabelle A.1 und A.2 in Anhang A), wird für die weitere Betrachtung das von Pahl et al. in [PBF07] beschriebene und alle wichtigen Tätigkeiten umfassende, vierstufige Phasenmodell herangezogen. Dieses setzt sich aus den nicht immer streng voneinander abzugrenzenden Phasen *Klären*, *Konzipieren*, *Entwerfen* und *Ausarbeiten* zusammen, in denen gemäß Abbildung 5.5 diverse Dokumenttypen als Inputfaktoren für die nachfolgenden Phasen zu erzeugen sind oder für deren Erstellung benötigt werden⁷⁷. Dabei dient die erste Phase der Aufgabenklärung dazu, die

⁷⁶<http://www.springsource.org>

⁷⁷Ergebnisse einer detaillierten Analyse von Dokumenten der technischen PE hinsichtlich ihrer Zuordnung zu den einzelnen Phasen des PEP (genau genommen den Phasen des FORFLOW-Prozessmodells) sowie in Bezug auf ihren Informationsgehalt können Anhang E entnommen werden.

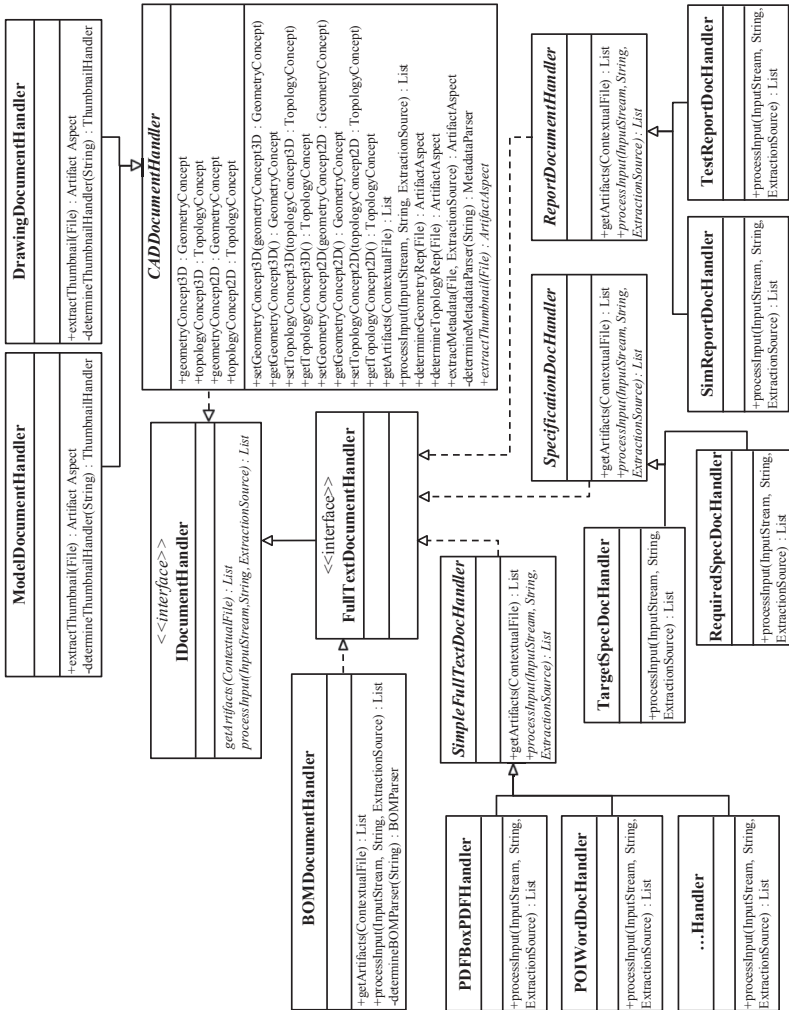


Abbildung 5.4. — Hierarchie der zur Dokumentverarbeitung verfügbaren DocumentHandler

Problemstellung zu präzisieren und die Anforderungen an das zu entwickelnde Produkt zu erfassen [AB⁺07, PBFG07, Con08b]. Hierzu sind grundsätzlich die beiden Szenarios der kundenanonymen und der kundenindividuellen PE zu unterscheiden.

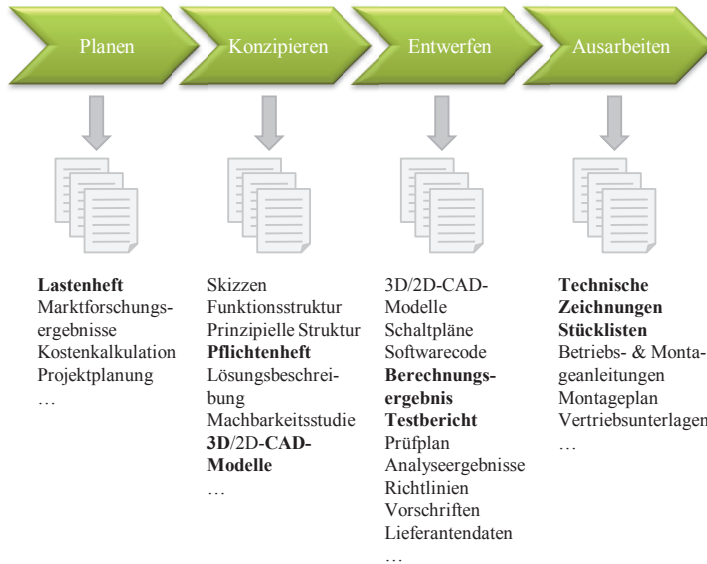


Abbildung 5.5. — Vierstufiges Phasenmodell nach Pahl et al. [PBFG07] mit Beispielen für zugehörige Dokumenttypen

Während nämlich im ersten Szenario Produkte für den Markt entwickelt werden, dessen Bedürfnisse und Wünsche neben Konkurrenzprodukten mittels Methoden der Marktforschung analysiert werden müssen, sind die Rahmenbedingungen bei einer kundenindividuellen Entwicklung enger gefasst. Hier gibt der Kunde seine Anforderungen an das Produkt vor, deren Realisierbarkeit zu überprüfen und mit dem Kunden abzustimmen ist. Als Ergebnis wird in beiden Entwicklungsszenarios ein Lastenheft erstellt, das eindeutig spezifiziert, welchen Zweck und welche Eigenschaften das zu entwickelnde Produkt erfüllen muss bzw. soll. Mit der folgenden Konzipierungsphase beginnt schließlich der kreative Teil der PE, in dem verschiedene Möglichkeiten zur Lösung des gegebenen Problems und damit zur Umsetzung der Anforderungen auszuarbeiten sind [Con08b]. Hierfür ist die vom Produkt zu erfüllende Gesamtfunktion unter Bezug auf den Energie-, Stoff- und Signalumsatz lösungsneutral zu formulieren und soweit wie möglich in Teilfunktionen zu zerlegen [AB⁺07]. Für diese Teilfunktionen wiederum sind geeignete Wirkprinzipien zu finden, die teilweise eigene Produkte

bzw. Produktkomponenten mit entsprechenden Schnittstellen erfordern und zu Funktionsstrukturen — den prinzipiellen Lösungen — kombiniert werden. Um dabei die Machbarkeit neuer und innovativer Ideen bzw. Lösungswege zu verifizieren, werden oftmals Ideenskizzen, erste einfache CAD-Modelle als Designstudien und Auslegungsberechnungen erstellt, die allerdings nur auf eine Grobgestalt des Produktes zielen und zur Veranschaulichung für die Wirkstruktur des Gesamtproduktes dienen [PBFG07]. Von diesen Lösungsvarianten ist schließlich eine auf Basis von Bewertungsergebnissen und gegebenenfalls in Abstimmung mit dem Kunden auszuwählen und gemeinsam mit den Inhalten des Lastenheftes in einem Pflichtenheft zu dokumentieren. Die ausgewählte Lösung ist schließlich in der dritten Phase, der Entwurfsphase, durch eine Feingestaltung der CAD-Modelle für die einzelnen Produktkomponenten zu konkretisieren (= Festlegung der Baustruktur [AB⁺07]) sowie durch verschiedene computerunterstützte Simulationsverfahren und Tests hinsichtlich Festigkeit, Bewegungsverhalten und anderen Anforderungen zu überprüfen und zu optimieren [PBFG07, Con08b]. Erst wenn hier eine durchgängige, alle Tests bestandene und alle Anforderungen erfüllende Lösung erreicht ist, wird in der Ausarbeitungsphase die Produktdokumentation angefertigt. Diesbezüglich sind neben Fertigungsunterlagen, wie beispielsweise technischen Zeichnungen, Stücklisten oder Numerical Control (NC)-Programmen, welche zur Herstellung des Produktes benötigt werden, auch sämtliche Dokumente für den Vertrieb, den Betrieb, die Wartung und die Entsorgung des Produktes zu erstellen [AB⁺07, Con08b].

Zusammenfassend zeigt sich, dass mit jeder Phase, d.h. von der informativen über die prinzipielle und die gestalterische bis hin zur herstellungstechnischen Festlegung eines Produktes [PBFG07], die in den Ergebnisdokumenten enthaltenen Informationen zunehmend konkreter werden. Um welche Informationen es sich hierbei im Detail handelt, wird in den folgenden Kapiteln anhand ausgewählter, für diese vier Phasen wichtigen Dokumenttypen erläutert. Dazu werden diese in Abbildung 5.5 hervorgehobenen Dokumenttypen jeweils zunächst allgemein definiert und anschließend hinsichtlich ihres Informationsgehalts und der damit verbundenen Extraktionsmöglichkeiten im Rahmen des LFRP-Frameworks genauer betrachtet.

5.4.1. Lasten- und Pflichtenheft

Vor der eigentlichen Entwicklung eines Produktes sind insbesondere zwei Dokumenttypen von besonderer Bedeutung, welche beide die wesentlichen Aspekte und Rahmenbedingungen der zu erbringenden Aufgabenstellung enthalten. Der Unterschied zwischen beiden Dokumenttypen besteht darin, dass die Aufgabenstellung aus unterschiedlichen Blickwinkeln betrachtet und zu unterschiedlichen Zeitpunkten im PEP beschrieben wird. Während nämlich im Lastenheft der Kunde bzw. Auftraggeber, welcher sowohl innerhalb als auch außerhalb des Unternehmens angesiedelt sein kann,

zu Beginn der Klärungsphase seine Bedürfnisse, d. h. das gewünschte Produkt mit seinen Eigenschaften und Funktionen, verdeutlicht, erfolgt spätestens am Ende der Konzeptionsphase im Pflichtenheft eine Übersetzung dieser überwiegend nicht-formalen Kundenwünsche in formale technische Anforderungen sowie eine Erläuterung des zu ihrer Realisierung geplanten Lösungsansatzes [FL97, Ste05]. Somit findet hier eine Beschreibung der Aufgabenstellung aus Sicht der verantwortlichen Produktentwickler statt. Diese Transformation der Aufgabenbeschreibung von der Kunden- in die Entwicklersicht ist mit dem Kunden abzustimmen und stellt eine notwendige Voraussetzung dafür dar, ein, den Kundenwünschen entsprechendes Produkt liefern zu können. Daher ist das Pflichtenheft als Kommunikationsgrundlage zwischen Auftraggeber und Auftragnehmer zu sehen, die eindeutig spezifiziert, welchen Zweck ein Produkt unter welchen Bedingungen wie erfüllen soll [Kol98]. Häufig wird in diesem Zusammenhang in der industriellen Praxis zwar auch von der Spezifikation eines Produktes gesprochen; allerdings reicht der Inhalt eines Pflichtenheftes weit über die in einer Produktspezifikation erläuterten technischen und/oder funktionalen Aspekte eines Produktes [Wit10] hinaus.

In welcher Art und Weise die Beschreibung der Aufgabenstellung erfolgt, hängt zum einen vom Ersteller des jeweiligen Dokumentes und zum anderen von den unternehmensspezifischen Gegebenheiten ab. Demzufolge kann nicht von einer, für alle Unternehmen allgemein gültigen und einheitlichen Struktur für Lasten- bzw. Pflichtenhefte ausgegangen werden. Stattdessen erfolgt ihre Erstellung häufig auf Basis von unternehmensspezifischen Dokumentvorlagen, wodurch zwar eine gewisse Strukturierung vorgegeben ist, diese aber nicht zwangsläufig eingehalten werden muss. Oftmals sind abhängig von der konkreten Situation Ergänzungen oder Änderungen vorzunehmen, wodurch ganze Strukturelemente wegfallen oder neu hinzukommen können. Aus diesem Grund wird im Folgenden zunächst der Informationsgehalt beider Dokumenttypen genauer betrachtet. Im Anschluss werden die im Rahmen des LFRP-Frameworks erforderlichen Extraktorkomponenten mit ihren Besonderheiten vorgestellt.

5.4.1.1. Informationsgehalt eines Lastenheftes

Das Lastenheft stellt die erste Informationsquelle für Produktentwickler dar, die gemäß DIN 69901 Teil 5 alle Forderungen und Wünsche enthält, die ein Auftraggeber an die Leistungen und Lieferungen eines Auftragnehmers stellt [Deu09]. Folglich sollten in einem Lastenheft sowohl organisatorische als auch produktspezifische Informationen enthalten sein. Unter organisatorischen Daten werden hier neben dem Erstellungsdatum und sonstigen Metadaten des Dokumentes auch Angaben über den Auftraggeber verstanden. Letztere werden häufig entweder in Form von Kontaktdaten oder — im Fall der Lastenhefterstellung durch den Auftragnehmer — in Form eines Verweises auf den Kunden anhand seines Personen- bzw. Firmennamens oder seiner Kundennummer verdeutlicht. Als produktspezifische Informationen hingegen sind

sämtliche Erläuterungen zu sehen, die Auskunft darüber geben, was und wofür etwas gemacht werden soll. Dabei sind neben eventuell relevanten Rahmenbedingungen v. a. Angaben zum gewünschten Soll-Zustand eines Produktes, d. h. zur Zielsetzung, der gewünschten Produktfunktion sowie den Produkteigenschaften, zu treffen. Allerdings ist zu beachten, dass diese Beschreibung üblicherweise in natürlicher Sprache erfolgt, was eine gewisse Vagheit, Unvollständigkeit, Unstrukturiertheit — in manchen Fällen sogar Widersprüchlichkeit — der Angaben nicht ausschließen lässt. Zudem hat eine genauere Analyse von Beispieldokumenten gezeigt, dass in Bezug auf die inhaltliche Gestaltung von Lastenheften in der industriellen Praxis sehr unterschiedliche Ausprägungen existieren. Diese weisen von sehr gut strukturierten Dokumenten bis zu einer einfachen Auflistung von Anforderungen, wie es ein in Listing 5.6 enthaltener und aus [Ste05, S. 488] entnommener Auszug aus einem Lastenheft zur Entwicklung eines Kugelschreibers zeigt, eine große Bandbreite auf.

Listing 5.6 — Auszug aus einem Lastenheft für einen Kugelschreiber

- Der Kugelschreiber hat einen neuartigen Mechanismus, sodass ein unbeabsichtigtes Verschmutzen nahezu unmöglich ist.
- Der Kugelschreiber muss ein modernes ansprechendes Design aufweisen.
- Optional kann das Schreibgerät mit individueller Werbung ausgestattet werden.
- Die eingesetzten Materialien sollen einen ausgesprochen soliden Eindruck von hoher Qualität vermitteln.
- ...

Doch auch bei strukturierten Dokumenten — selbst wenn diese im gleichen Unternehmen erstellt wurden (vgl. Beispiele in Abbildung 5.6) — kann nicht immer von ein und derselben Struktur oder von einem einheitlichen Layout ausgegangen werden, was eine Extraktion von Informationen für eine Anwendung im LFRP-Framework wesentlich erschwert. Prinzipiell hat sich allerdings gezeigt, dass für die Anfertigung eines Lastenheftes in der industriellen Praxis primär Textverarbeitungssysteme zum Einsatz kommen, wobei sich insbesondere zwei Dateiformate zur Speicherung dieser Textdokumente etabliert haben: das DOC/DOCX-Format von Microsoft⁷⁸ und das PDF-Format von Adobe Systems⁷⁹ [Ado08]. Für beide Dateiformate stehen frei zugängliche Bibliotheken, wie beispielsweise die Java-Bibliotheken PDFBox⁸⁰ und POI⁸¹ der Apache Software Foundation, welche auch im Rahmen der prototypischen Imple-

⁷⁸<http://www.microsoft.com>

⁷⁹<http://www.adobe.com>

⁸⁰<http://pdfbox.apache.org>

⁸¹<http://poi.apache.org>

mentierung zur Extraktion von Informationen aus derartigen Dokumenten genutzt wurden, zur Verfügung.

Lastenheft	
Technische Projektverantwortung:	Tel:
Email:	Fax:
Teilbezeichnung:	
Teilnummer:	
Zeichnungsindex:	
Kunde:	
1. Projektierung <ul style="list-style-type: none"> - Zum Projektstart muß vom Werkzeugmacher ein Statusbericht erstellt werden. Dieser muß undigefordert alle 2 Wochen an den technischen Projektverantwortlichen bei gesondert werden, inhaltlich sollte dieser dem Beispiel im Anhang entsprechen. - Zusätzlich zum Terminplan können Photos vom Werkzeugstand angefordert werden. - Vom Lieferanten ist eine Konstruktions-FMEA für das Werkzeug zu erstellen. Bei Bedarf ist diese zur Einsicht vorzulegen. - Die Werkzeugkonstruktion muß vor der Formherstellung zur Einsicht vorgelegt werden (Papierzeichnungen oder CAD Daten). - Erkennt der Lieferant, daß der vereinbarte Endtermin nicht eingehalten werden kann, so hat er dies unverzüglich unter Angabe der genauen Gründe und der voraussichtlichen Dauer der Verzögerung schriftlich mitzuteilen. 	
2. Teilbeschreibung und Zeichnungen <ul style="list-style-type: none"> - Die 2D/ CAD Daten werden im Normalfall im CATIA-Format von zur Verfügung gestellt. - Grundsätzlich sind die von übermittelten CAD Daten und die Teilzeichnung verbindlich. - Fehlende Entwürfsfragen müssen vom Werkzeugmacher eingebracht werden. - Sollte eine Änderung am Teil aus werkzeugtechnischen Gründen erforderlich sein, so muß diese von schriftlich freigegeben werden. 	
3. Grundsätzliche Formgestaltung <ul style="list-style-type: none"> - Das Werkzeug muß in seiner Ausführung dem Stand der Technik entsprechen. Speziell Aufbaugröße, Aufteilung der Nester, Wahl des Anguß usw. müssen optimal gestaltet werden. - Die Formmaster sind im Regelfall als Einstücke auszuführen. Es ist nicht gestattet die Form direkt in die Formplatten zu anfertigen. - Bereiche die eine Abstimmung nach der Erstmusterung erfordern (Verrastungen, Flimschmelze,...) müssen ohne großen Aufwand und Zusatzkosten veränderbar sein. Hierzu sollten im Normalfall Einstücke vorgesehen werden. - Normteile wie Auswürger, Führungen, usw. müssen Metrisch ausgeführt werden. 	

ww. 20.09.02

Lastenheft für xxx

Projekt:
Anzahl Fahrzeuge:
Anzahl Werkzeuge:
Zeichnungsstand:

Nr.	Kriterien	Erforderlich		Forderungen
		Ja	Nein	
1.0	Werkzeugaufbau:			
1.1	Formmaster nach CAD-Daten (werden von xxx zur Verfügung gestellt)	X		Zur Angebotserstellung werden Bilder zur Verfügung gestellt.
1.2	Formstellungsebene (Nutenvermeidung)	(siehe 1.3)	-	
1.3	Lage der Formmaster (siehe 1.3.1)	-	-	
1.4	Stiefler Aufbau (Oben und Unterteil)	X		
1.5	Stiefler Aufbau (Oberteil 2 geteilt)		X	
1.6	Oberteil schalen (siehe Punkt 2.0)	X		
1.7	Öffnungswinkel zum Unterteil	-	-	110° bis 115°
1.8	Dämpfungskinder am Werkzeugoberteil für Dreiviertelentlastung	X		Angabe über Art in Prospekt
1.9	Sonstige Ausführungen	-	-	Zentriert für Fokussierung der Werkzeughälften
2.0	Schalene:			
2.1	Schalene Ausführung	X		
2.2	Stabile Ausführung (Betätigungszellen beachten)	X		Unterteil verstärkt
2.3	Sonstige Ausführungen	-	-	
3.0	Werkzeugmaterial:			
3.1	Werkzeugmaterial:	-	-	Angebot über Stahl bzw. hochverfestes Aluminium
3.2	Sonstige Ausführungen	-	-	
4.0	Temperierung:			
4.1	Temperierung Oberteil	X		180° Wasserkehlau

Abbildung 5.6. — Auszüge aus unterschiedlichen Lastenheften eines Unternehmens

5.4.1.2. Informationsgehalt eines Pflichtenheftes

Wie bereits zuvor erwähnt, ist das Pflichtenheft vom Lastenheft dadurch abzugrenzen, dass es die im Lastenheft enthaltenen Anforderungen in für die PE relevante Kriterien transformiert und diese um erarbeitete Realisierungsvorgaben zu ihrer Erfüllung ergänzt [Deu09]. Demnach weist auch ein Pflichtenheft sowohl organisatorische als auch produktspezifische Daten auf, wobei letztere allerdings in einem höheren Strukturiertheits- und v. a. Detaillierungsgrad beschrieben werden. Dies äußert sich darin, dass der laut Literatur wichtigste Teil eines Pflichtenheftes aus einer Anforderungsliste besteht [Kol98, Ste05, PBFG07, Hab08], welche im Rahmen der am Ende des PEP durchzuführenden Endabnahme des Kunden zum Abgleich von realisierten Ist- und vereinbarten Soll-Eigenschaften eines Produktes dient. Aus diesem Grund sollten sämtliche Anforderungen, Bedingungen und Restriktionen in dieser Liste möglichst konkret, d. h. in Form quantifizierbarer und damit messbarer Größen, angegeben werden [Ste05, Hab08]. Zusätzlich werden diese teilweise mit Toleranzangaben versehen, die eine Zulässigkeit in einem bestimmten Intervall gewährleisten [Ste05]. Des

Weiteren ist jede Anforderung entweder als sogenannte Fest- bzw. Fixforderung, die unter allen Umständen zu realisieren ist, oder als Wunschforderung, die nur dann zu erfüllen ist, wenn sie keinen oder nur einen geringen Zusatzaufwand erfordert, zu markieren [Ste05, Hab08, Wie08]. Eine Anforderung ist dabei gemäß Koller allgemein als eine erwartete Leistung, Fähigkeit, Eigenschaft oder einzuhaltende Parameterwertvorgabe zu verstehen [Kol98], die einen bestimmten Aspekt eines Produktes adressiert. Demzufolge werden Anforderungen üblicherweise in Anforderungsgruppen eingeordnet, für die in der Literatur unterschiedliche Kategorisierungen existieren. So unterscheidet beispielsweise Koller zwischen den vier Anforderungskategorien *Marktbedingungen*, *Werdegangsbedingungen*, *Umwelt- und gesellschaftsbedingte Forderungen* und *Systemeigene Bedingungen* [Kol98], welche wiederum hierarchisch in Unterkategorien gegliedert werden. Während die Gruppe der Anforderungen, die durch den Markt bzw. den Kunden vorgegeben werden, u. a.

- Leistungsdaten (= Leistung, Geschwindigkeit, Drehmoment, ...),
- Informationen zur Lebensdauer und Zuverlässigkeit,
- Schnittstellenbeschreibungen (= geometrische und elektrische Anschlussdaten, zulässige Abmessungen, Gewichte, ...),
- Informationen zum Aussehen (= Design, Farbe, ...),
- Informationen über eventuell erforderliche Sonderausführungen,
- Preise und
- Termine

enthält, beziehen sich Werdegangsbedingungen auf die bei der Entwicklung und Herstellung eines Produktes zu berücksichtigenden DfX-Aspekte (z. B. Fertigungs-, Lager-, Transport-, Reparatur-, Recyclinggerechtigkeit, ...). Die dritte Kategorie der Umwelt- und gesellschaftsbedingten Forderungen schließt neben einzuhaltenden Gesetzen, Normen und Richtlinien, auch Faktoren ein, die aus der Umwelt auf das Produkt einwirken (z. B. Temperatur, Spritzwasser, Staub usw.) oder die umgekehrt das Produkt auf die Umwelt ausübt (z. B. Schadstoff- / Strahlungsemissionen, Lärm, Verletzungsgefahren usw.). Alle weiteren produktspezifischen Merkmale, wie beispielsweise Angaben im Hinblick auf Eigenwärme, Reibung oder Resonanzerscheinungen, werden schließlich der vierten Kategorie, den systemeigenen Bedingungen, zugeordnet. Selbstverständlich lassen sich die hier aufgezeigten Anforderungen aber auch in andere Gruppierungen einteilen, wie sie beispielsweise von Haberhauer in [Hab08] oder von Ullman in [Ull03] vorgeschlagen werden (siehe Tabelle 5.1).

Insgesamt wird aus all diesen Strukturierungen deutlich, dass nicht nur technische, sondern auch marktwirtschaftliche, ökologische und auch rechtliche Anforderungen dokumentiert werden, um eine möglichst umfassende Abdeckung aller Produktaspek-

Tabelle 5.1. — Kategorisierung von Anforderungen nach Haberhauer [Hab08] (links) und nach Ullman [Ull03] (rechts) mit Beispielen

<i>Haberhauer-Kategorisierung</i>	<i>Ullman-Kategorisierung</i>
Geometrie (Abmessungen, Anschlussmaße, verfügbarer Bauraum)	Funktionale Anforderungen (Energie-, Informations-, Materialfluss)
Technische Daten (Leistung, Kräfte, Drehzahlen)	Ergonomische Faktoren (Erscheinung, Bedienbarkeit)
Stoff/Material (geforderte Werkstoffe, Schmiermittel)	Physikalische Anforderungen (Physikalische Eigenschaften, verfügbarer Raum)
Qualität (Prüfvorschriften, Kontrollgrößen, Toleranzen)	Zuverlässigkeitswerte (Sicherheit, mittlere störungsfreie Zeit)
Sicherheit (Arbeits-, Umwelt-, Überlastsicherheit)	Lebenszyklusaspekte (Vertrieb, Wartbarkeit, Instandhaltung, Recycling)
Fertigung/Montage (gefordertes Herstellungsverfahren)	Ressourcenbezogene Faktoren (Zeit, Kosten, Maschinen)
Instandhaltung (Verschleißteile, Wartungsumfang, -intervalle)	Fertigungsspezifische Anforderungen (Werkstoffe, Mengeneinheiten, Standards)
Ergonomie	
Signal (Steuerung, Messgrößen)	
Gebrauch (Einsatzort, Zweck, Lebensdauer)	
Spezielle Anforderungen (besondere Kundenwünsche)	
Stückzahlen (geforderte Produktionsmenge)	
Kosten (Herstell-, Werkzeugkosten)	
Termine (Prototyp, Serienbeginn, Lieferung)	

te zu erreichen. Welche Strukturierung jedoch im konkreten Anwendungsfall sinnvoll und v. a. wie umfangreich diese ist, hängt primär von dem zu entwickelnden Produkt und seinen Rahmenbedingungen ab. Außerdem sollte dabei immer auch berücksichtigt werden, dass nicht nur die zu erfüllenden Anforderungen festgelegt werden. Vielmehr sind auch nicht-realiserbare Anforderungen zu dokumentieren, um eventuelle Missverständnisse und daraus resultierende Projektverzögerungen möglichst zu vermeiden.

Zusätzlich zu einer Anforderungsliste sollte ein Pflichtenheft aber auch immer über weitere, nicht direkt produktbezogene Abschnitte verfügen, welche sich in Anlehnung an Koller prinzipiell folgendermaßen gliedern lassen [Kol98]:

- Organisatorische Daten: Auftraggeber, Auftragnehmer (Verantwortlicher), Datum, Version, Projektdaten, zeitlicher Rahmen (Termine), . . .
- Stand der Technik: Informationen zur Ausgangssituation, d. h. Ergebnisse aus Wettbewerbs- bzw. Marktanalysen, Patente, Normen, Fertigungsverfahren u. a.
- Markt: Kunden, Länder, Zielgruppen
- Problemstellung: Lösungsneutrale Beschreibung der Zielsetzung mit Erläuterung der zu erfüllenden Produktfunktion(en)
- Lösungsbeschreibung: Erläuterung des Lösungsprinzips zur Realisierung der Anforderungen

Dabei finden sich v. a. bei den beiden letztgenannten Elementen häufig auch Verweise auf andere Dokumente wie beispielsweise auf technische Zeichnungen oder Richtlinien, die bestimmte Aspekte näher erläutern und daher sozusagen als Bestandteil des Pflichtenheftes zu sehen sind. Insgesamt hat sich aber auch bei der Analyse von Pflichtenheften gezeigt, dass es sich eher um semi-strukturierte Textdokumente handelt, die wie Lastenhefte üblicherweise im DOC/DOCX- bzw. PDF-Format vorliegen und nicht alle aufgezeigten Inhaltsbestandteile enthalten müssen. Demzufolge hängt die Extraktion von Informationen auch hier stark vom tatsächlichen Aufbau und Inhalt des jeweiligen Dokumentes ab.

5.4.1.3. Extraktorkomponenten für Lasten- bzw. Pflichtenhefte

Aufgrund der inhaltlichen und auch strukturellen Varietät von Lasten- und Pflichtenheften werden im Folgenden zunächst anhand von Auszügen aus Beispieldokumenten verschiedene extrahierbare Informationen, die für eine Bildung von Facetten in Frage kommen können, aufgezeigt. Welche von diesen allerdings in einem konkreten Anwendungsfall von Bedeutung sind, hängt von der tatsächlichen Gestaltung dieser Dokumententypen ab und ist damit unternehmensspezifisch festzulegen. Folglich sind entsprechende, für diese Dokumententypen zuständige Extraktorkomponenten zu erstellen, die als Subklassen des `SpecificationDocHandler` (vgl. Abbildung 5.4) eine entsprechende Verarbeitung der Dokumente bzw. ihrer Inhalte gewährleisten. Auf die hierfür einsetzbaren Methoden wird im Anschluss näher eingegangen.

TECHNISCHES PFLICHTENHEFT

Türle Stirnwand

Entwicklungsziele und Anforderungen

Änderungsindex / Änderungsbeschreibung

Datum	Seite	Kapitel	Stabhalter	Änderung
25.12.02		geändertes Dokument	Weber	Erstellung
05.11.02		Kap. 2.1	Weber	Fein korrig.

Auftraggeber: Musterkunde AG

Ansprechpartner: Fr. Weber 0951/863-2883

Änderungsindex	Änderungsbeschreibung	CAD	2D	Auftraggeber	Reise	Prof. Datum	Stem. Datum
YAP16420.02	Stirn Wand	IG	001	001	---	2002-09-25	2002-09-25

Pflichtenheft

Türle Stirnwand

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

Reise Datum: 2002-09-25

4 Anforderungen

4.1 Allgemeine Anforderungen

Durch das Handling beim Kunden darf die Türle nicht beschädigt, aus dem Durchbruch gedrückt oder undicht werden.

- Definierte Aufnahmen des Lugs, Stirn- und durchbruch (Biegeradius $r > 50 \text{ mm}$)
- Schutz des Lugs, Stirn- und durchbruch vor Beschädigung
- Abbildung des Durchbruches zum Fahrzeuginnenraum
- Geometrische Dimensionierung
- Erfüllung der Kundenanforderungen gem. Lastenheft

4.2 Montage-/Demontageaspekte

4.2.1 Montage

Für die Montage der Türle sind Standardwerkzeuge zu verwenden. Der Montageaufwand soll minimiert werden.

Die Montage der Türle verläuft vom Innenraum in den Motorraum. Der Einbau der Türle in den Stirn- und durchbruch soll mit möglichst geringem manuellem Kraftaufwand erfolgen.

4.2.2 Demontage

Die Demontage der Türle muß ohne Zerstörung derselben möglich sein. Die Wiedereinbau muß unter Verwendung der demontierten Türlebestandteile (neue Schaumstoffdichtung) erfolgen.

4.2.3 Nachrüstung von Zusatzumfängen / Reparaturlösung

Für die Nachrüstung von Zusatzumfängen sind Nachrüstlösungen im Kunststoffteil gefertigt vorgesehen, welche bei Bedarf durchgebohrt werden können.

4.3 Thermische Beanspruchungen

Es dürfen keine sichtbaren Verformungen auftreten. Kleine maßliche Abweichungen sind tolerierbar, wenn der Anschluß an benachbarte Teile nicht beeinträchtigt wird. Sichtbare Oberflächenveränderungen oder Verfärbungen sind nicht zulässig.

4.4 Dichtheitsanforderungen

Die Dichtheitsanforderungen gelten für die im Fahrzeug montierte und mit Leitungskabeln besetzte Türle.

Wassersäule über der montierten Türle 300 mm

Abbildung 5.7. — Zwei Beispielseiten aus einem Pflichtenheft mit extrahierbaren dokumentenspezifischen (grün umrandet), personenspezifischen (blau umrandet) und produktspezifischen (orange umrandet) Daten

Analyse von Beispieldokumenten

Unter der Annahme, dass alle Lasten- als auch Pflichtenhefte über organisatorische Daten verfügen, lassen sich aus diesen bereits erste wichtige Facetteninformationen auslesen. Hierzu gehören Informationen über das Dokument selbst, die entweder aus den Dokumenteigenschaften — sofern diese sinnvoll gepflegt werden — oder teilweise auch aus dem Inhalt des Dokumentes selbst extrahiert werden können. Zu letzteren können gemäß dem in Abbildung 5.7 links gezeigten Beispiel der Titel, das Erstellungsdatum, das Änderungsdatum, die Auftragsnummer sowie der Ersteller bzw. Autor des Dokumentes gehören. Doch bereits hier ist zu differenzieren, welche Artefakte die im Dokument enthaltenen Daten tatsächlich adressieren. So stellen beispielsweise die ersten vier der genannten Daten (vgl. grüne Umrandungen in Abbildung 5.7 links) spezifischere Informationen des Dokumentes dar, weshalb sie als Ausprägungen dokumentenspezifischer Attributfacetten dem das Dokument repräsentierenden `IndexableDocument`-Objekt zuzufügen sind. Der Dokumentautor bzw. dessen Daten hingegen sind als personenspezifische Informationen (siehe blau umrandete Daten in Abbildung 5.7 links) zu sehen, was die Erstellung und Indexierung eines Artefakts vom Artefakttyp `Person` (= ein `IndexablePerson`-Objekt) mit den entsprechenden Attributfacetten als Repräsentationen erfordert. Dies wiederum resultiert in

der Notwendigkeit, beide Artefakte jeweils um eine Beziehungsfacette in Form eines **ArtifactRelationship**-Objektes zu ergänzen, welches die Beziehung zwischen den beiden Artefakttypen beschreibt (siehe hierzu das Beziehungsnetz in Abbildung 4.10). Gleiches gilt für die Kontaktdaten des Auftraggebers, falls dieser nicht mit dem Dokumentautor übereinstimmt. Liegt eine Übereinstimmung vor, so sollte das bereits erzeugte **IndexablePerson**-Objekt um die zusätzlichen Daten (hier Telefonnummer und Firmenname) ergänzt werden. Sind in den organisatorischen Daten noch andere Artefakttypen als die bisher genannten adressiert (z. B. das zugehörige Projekt oder zusätzliche Dokumente), so sollten selbstverständlich auch für diese eigenständige **IndexableArtifact**-Objekte mit Attribut- und Beziehungsfacetten als Repräsentationen erstellt werden.

Bei der Generierung zusätzlicher **IndexableArtifact**-Objekte ist allerdings zu beachten, dass eine eindeutige Identifizierbarkeit dieser Artefakte gewährleistet sein muss, um eine redundante Ablage des gleichen Artefakts im Index zu vermeiden. Wird nämlich ein Artefakt, wie beispielsweise der Dokumentautor, nicht durch eine eindeutige ID, sondern wie im linken Auszug von Abbildung 5.7 durch einen Nachnamen referenziert, so kann eine im Rahmen der späteren Konsistenzprüfung notwendige Erweiterung von eventuell im Index bereits vorhandenen Daten zu diesem Autor nicht durchgeführt werden. Der Grund hierfür ist, dass der Nachname durchaus auf mehrere Personen im Umfeld des Unternehmens zutreffen kann, was eine eindeutige Zuordnung unmöglich macht. Demzufolge sollte prinzipiell bereits bei der Erstellung von Dokumenten auf diese Notwendigkeit geachtet werden. Ist dies nicht gegeben, sind grundsätzlich zwei Varianten für das weitere Vorgehen denkbar. Zum einen könnte der *Indexierer* selbst eine ID für das Artefakt erzeugen, womit allerdings nicht nur eine redundante Datenhaltung sondern in gewisser Weise auch ein Informationsverlust in Kauf genommen werden muss. Zum anderen könnte man sich die Befragung eines Nutzers oder eines Systemverantwortlichen vorstellen, was allerdings nur in dem Fall möglich ist, in dem die Indexierung während des laufenden Geschäftsbetriebes abläuft. Findet die Indexierung dagegen zu festen Zeitpunkten (z. B. nachts) statt, müssten die nicht-indexierbaren Artefakte mit ihren Informationen beispielsweise in einer XML-Datei protokolliert werden. Orientiert sich diese an dem in Kapitel 4.4 erläuterten Aufbau der Importdatei für Artefaktinformationen, könnte diese Datei im Nachhinein (z. B. am nächsten Tag) auf einfache Weise manuell bearbeitet und indiziert werden.

Zusätzlich zu den organisatorischen Daten sind in beiden Dokumenttypen auch immer produktspezifische Informationen enthalten, so dass nicht nur das Dokument, sondern immer auch das darin beschriebene Produkt als eigenständiges **IndexableArtifact**-Objekt generiert werden sollte. Dieses wird, wie in Abbildung 5.7 links anhand der orange umrandeten Daten veranschaulicht, üblicherweise durch eine Sachnummer und gegebenenfalls durch eine Bezeichnung — den Produktnamen — im Dokument referenziert. Vor allem die Sachnummer stellt hier ein wichtiges Datum dar, um ent-

sprechend dem obigen Vorgehen auch für das Produkt ein **IndexableProduct**-Objekt erstellen zu können. Dies wiederum impliziert die Bildung von *Produkt* \leftrightarrow *Dokument*-Beziehungen, so dass dieses nicht nur um Attributfacetten, sondern auch um eine Beziehungsfacette erweitert werden kann. Beispiele für Attributfacetten sind diesbezüglich der bereits erwähnte Produktname oder Informationen über die Funktion(en), die das Produkt erfüllen soll. Werden letztere, wie im Beispiel in Abbildung 5.8, unter einem separaten Gliederungspunkt aufgelistet, so können die einzelnen Elemente dieser Liste als Ausprägungen einer mehrwertigen Facette zur Beschreibung der Produktfunktion genutzt werden. Allerdings muss dazu der Extraktorkomponente bekannt sein, dass die Funktion(en) unter diesem Gliederungspunkt erläutert werden. Dies kann einerseits durch eine feste Codierung in der Extraktorkomponente erreicht werden, was allerdings bereits bei einer leicht veränderten Reihenfolge der einzelnen Gliederungspunkte sehr unflexibel ist. Aus diesem Grund wird ein Pattern Matching, d. h. ein Zeichenkettenvergleich der Überschriften mit den im Facettenschema in den `<synset/>`-Elementen definierten Begrifflichkeiten empfohlen. Ein ähnliches Vorgehen könnte man beispielsweise auch im Hinblick auf die für ein Produkt durchzuführenden Prüfungen (siehe Abbildung 5.8 unter Gliederungspunkt 5) anwenden, indem man sozusagen die Überschriften der zweiten Gliederungsebene (hier *Dichtheitstest, Nachweis der Nachrüstbarkeit von Leitungsumfängen, Wärmewechselprüfung*) extrahiert und als Ausprägungen einer zugehörigen Attributfacette betrachtet. Damit würde sich beispielsweise eine Filterung nach Produkten, die bestimmte Prüfungen bestehen bzw. für die bestimmte Prüfungen existieren müssen, realisieren lassen.

Des Weiteren besteht die Produktbeschreibung im Wesentlichen aus der Angabe der zu erfüllenden Anforderungen. Erfolgt diese in Form einer strukturierten Auflistung, wie sie in Abbildung 5.7 rechts veranschaulicht ist, könnten auch hier die Überschriften der zweiten Gliederungsebene (*Allgemeine Anforderungen, Montage / Demontageaspekte, Thermische Beanspruchungen, Dichtheitsanforderungen, ...*) die Ausprägungen einer Attributfacette **Anforderung** darstellen. Diese könnte beispielsweise genutzt werden, um eine Filterung nach Produkten durchzuführen, die Dichtheitsanforderungen erfüllen. Allerdings könnte man sich anstelle einer produktspezifischen Facette hier durchaus auch eine dokumentspezifische Facette vorstellen, die eine Suche nach für eine Problemstellung relevanten Lasten- bzw. Pflichtenheften unterstützt. Demzufolge ist zu entscheiden, welchem Artefakttyp man eine derartige Facette zuordnen würde. Zudem sind hinsichtlich der Beschreibung der einzelnen Anforderungen selbst zwei Szenarios zu differenzieren. Zum einen werden Anforderungen rein textuell beschrieben, wie dies beispielsweise für die Gliederungspunkte 4.1 und 4.2 im rechten Auszug von Abbildung 5.7 der Fall ist. Um aus ihnen brauchbare Informationen zu extrahieren, ist im Vergleich zum Nutzen, den man durch Anwendung der zugehörigen Facette haben würde, ein sehr hoher Aufwand erforderlich. Folglich sollten diese Textbeschreibungen als Ausprägungen einer dokumentspezifischen Ähnlichkeitsfacette **QbT** indexiert werden, welche den textuellen Inhalt im Retrievalprozess über eine

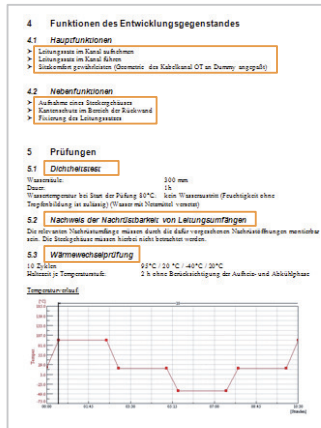


Abbildung 5.8. — Exemplarischer Auszug aus einem detaillierten Lastenheft, welcher die gewünschten Produktfunktionen sowie die vom Auftraggeber geforderten Prüfungen beschreibt

Schlagwortsuche zugreifbar macht. Zum anderen können Anforderungen aber auch durch numerische Werte erläutert werden, wie es am Beispiel der beiden Anforderungen *Betriebstemperatur* und *Kurzzeittemperatur* im rechten Auszug von Abbildung 5.7 gezeigt ist. Hier werden detaillierte Informationen zu den Produkteigenschaften definiert, so dass derartige Anforderungen prinzipiell als eigene, produktspezifische Attributfacetten mit den jeweils angegebenen Werten als Ausprägungen indexiert werden könnten.

Unabhängig von den organisatorischen und produktspezifischen Daten sollte für jedes Lasten- bzw. Pflichtenheft prinzipiell die bereits erwähnte dokumentspezifische Ähnlichkeitsfacette *QbT* gebildet werden. Sie wird benötigt, um ein Auffinden sämtlicher Informationen, für die keine eigenen Facetten gebildet werden können, über eine Schlagwortsuche zu gewährleisten. Ist zudem eine gewisse Strukturierung der Dokumente z. B. gemäß der in Kapitel 5.4.1.2 vorgestellten Bestandteile (State of the Art, Problemstellung, ...) gegeben, so könnten diese Strukturelemente darüber hinaus als dokumentspezifische Filterkriterien eingesetzt werden. Ihre Verwendung wäre beispielsweise insbesondere für noch unerfahrene Produktentwickler ein hilfreiches Kriterium, um nach Dokumenten zu suchen, die nützliche Informationen zu einem bestimmten Bestandteil (z. B. dem Stand der Technik) enthalten.

Methoden zur Informationsextraktion

Für die Extraktion von Informationen aus natürlichsprachigen semi-strukturierten Dokumenten können verschiedene Methoden aus dem Forschungsbereich des Natural Language Processing (NLP) herangezogen werden. Hierfür ist prinzipiell ein mehrstufiger Prozess oder Schritte umfassender Verarbeitungsprozess — eine sogenannte Verarbeitungspipeline — zu definieren, in der in sequentieller Abfolge verschiedene Aufgaben abgearbeitet werden. Diese Abfolge besteht gemäß Appelt und Israel meist aus vier elementaren Bausteinen [AI99]. So hat im ersten Schritt eine Textsegmentierung mit Hilfe eines sogenannten Tokenizers oder Tokenscanners zu erfolgen, welcher die Textstruktur z. B. auf Basis regulärer Ausdrücke identifiziert und den Text in spezielle Zeichenketten, die sogenannten Tokens, zerlegt [Neu01]. Diese Tokens sind im zweiten Schritt lexikalisch und morphologisch weiter zu verarbeiten, indem beispielsweise ihre Wortarten durch ein sogenanntes Part-of-speech Tagging bestimmt und ihre Flexions- oder Kompositaformen näher analysiert werden. Diese werden im Rahmen einer syntaktischen Analyse zur Erkennung von Sätzen und schließlich in der domänenspezifischen Analyse zur Erkennung von Eigennamen, Co-Referenzen oder auch Beziehungen verwendet [AI99, Neu01]. Welche Bausteine und welche Komponenten jedoch in einem konkreten Anwendungsfall benötigt werden, hängt dabei von vier Faktoren ab [AI99]:

- *Dokumentsprache*: Da die deutsche Sprache über mehr Flexionsformen als beispielsweise die englische Sprache verfügt, sind für sie andere Verarbeitungsregeln erforderlich.
- *Textart*: Bei der Verarbeitung von technischen Texten sind andere Informationen (Produktnamen, -gruppen usw.) von Bedeutung als es z. B. bei Nachrichtenanzeigen der Fall ist.
- *Textlänge und -layout*: In tabellarisch gestalteten Dokumenten wird die Identifizierung und Extraktion von Daten mittels Pattern Matching wesentlich erleichtert; in langen unstrukturierten Texten müssen die Abschnitte zunächst identifiziert bzw. gefunden werden.
- *Zu extrahierende Informationen*: Während eine Erkennung von Eigennamen relativ einfach mit Hilfe von Namenslisten realisiert werden kann, gestaltet sich die Erkennung von Eigenschaften zu einem Objekt aufgrund der hierzu erforderlichen Analyse ganzer Textfragmente um einiges aufwändiger.

Dabei sind für die Realisierung der einzelnen Komponenten grundsätzlich zwei verschiedene Ansätze denkbar [AI99]. Der erste Ansatz basiert auf der Erstellung einer Menge von Regeln und Grammatiken, die eine mit dem System vertraute Person — im Idealfall in Abstimmung mit einem Domänenexperten — erstellt, formalisiert, in das System einpflegt und schließlich anhand von Testläufen optimiert. Da für eine

gute Ergebnisqualität hierbei unter Umständen sehr viele Regeln manuell zu erstellen und iterativ zu optimieren sind, ist diese Vorgehensweise sehr arbeitsaufwändig und zeintensiv. Außerdem hängt die Ergebnisqualität von der Erfahrung und den Fähigkeiten des üblicherweise nicht mit der entsprechenden Domäne vertrauten Regelersellers sowie von der Existenz linguistischer Ressourcen wie Glossaren oder Lexika ab. Aus diesem Grund wird in vielen Fällen auf ein maschinelles Lernverfahren auf Basis statistischer Modelle (z. B. dem Hidden Markov Modell [Fin03]) zurückgegriffen, bei dem Gesetzmäßigkeiten, Regeln und Muster anhand von Beispieldaten gelernt bzw. antrainiert werden, so dass der zugrunde liegende Algorithmus auch bei Nicht-Trainingsdaten gute Ergebnisse liefert. Dazu wird allerdings eine ausreichend große Menge an Testdaten benötigt, die von einem Experten der entsprechenden Domäne beispielsweise manuell zu annotieren sind.

Da die Konzeption einer solchen Verarbeitungspipeline für Lasten- und Pflichtenhefte und die dafür erforderliche Entwicklung beziehungsweise Anpassung der einzelnen Komponenten den Rahmen der vorliegenden Arbeit sprengen würde bzw. ein Forschungsthema für sich darstellt, soll im Folgenden mit GATE (General Architecture for Text Engineering)⁸² ein Werkzeug vorgestellt werden, das für die Entwicklung derartiger Pipelines eine hilfreiche Unterstützung bietet. GATE ist sozusagen als ein Java-basierter Werkzeugkasten für die Entwicklung und Durchführung verschiedenster NLP-Aufgaben zu betrachten. Dieser wird seit dem Jahr 1995 an der Universität von Sheffield ständig weiterentwickelt und kann als Open Source-Werkzeug kostenlos genutzt werden. Darüber hinaus unterstützt GATE bei der Verarbeitung nicht nur verschiedene Dateiformate wie PDF, DOC, XML oder Rich Text Format (RTF), sondern auch diverse Sprachen (Englisch, Deutsch, Spanisch, Chinesisch, Arabisch, ...). Das bei Anfertigung dieser Arbeit verfügbare Release 5.2.1 setzt sich dabei aus mehreren Bestandteilen zusammen, die vorgefertigte Softwarekomponenten für die verschiedenen Aufgaben auf unterschiedliche Weise bereitstellen. Allerdings können diese Softwarekomponenten nicht nur genutzt, sondern auch an die eigenen spezifischen Bedürfnisse angepasst werden. Dazu steht mit dem Modul GATE Developer (siehe Screenshot in Abbildung 5.9) eine integrierte Entwicklungsumgebung zur Verfügung, die nicht nur zum Erzeugen und Testen von Pipelines aus verschiedenen regel- und lernbasierten NLP-Komponenten dient. Vielmehr wird dem Nutzer auch die Möglichkeit zur manuellen Annotation von Dokumenten und damit eine relativ einfache Art und Weise der Generierung von Trainingsdaten geboten. Zusätzlich existieren neben dieser GUI zur direkten Anwendung aber auch andere Module wie beispielsweise GATE Teamware als Web-Anwendung für Annotationen oder GATE Cloud als Cloud Computing-Lösung. Für den im Rahmen der vorliegenden Arbeit betrachteten Anwendungsfall der Entwicklung von Extraktorkomponenten ist jedoch GATE Embedded von besonderem Interesse. Dieses stellt eine Programmierschnittstelle zur Verfügung, die eine Nutzung von sämtlichen in GATE verfügbaren Services und damit

⁸²<http://gate.ac.uk>

eine Integration von NLP-Pipelines in andere Anwendungen, wie beispielsweise in das LFRP-Framework, ermöglicht. [CMBT02]

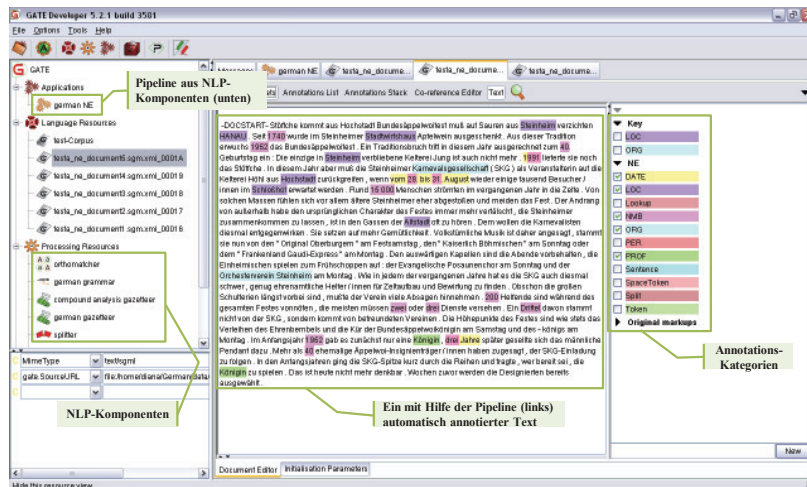


Abbildung 5.9. — Screenshot des GATE Developer

5.4.2. CAD-Modell

CAD-Modelle sind, wie die beiden vorhergehenden und auch die noch folgenden Dokumenttypen, ebenfalls zur Kategorie der Produktmodelle zu zählen. Obwohl sie zwar teilweise bereits in der Konzipierungsphase (hier jedoch nur im Sinne einer Grobgestaltung) erstellt werden, sind sie genau genommen als das Resultat der auf die Konzeptionsphase folgenden Entwurfsphase zu sehen. Folglich ist ein CAD-Modell als ein Produktentwurf zu betrachten, der in möglichst detaillierter Form die Gestalt eines Produktes beschreibt bzw. visualisiert und heutzutage mittels modernster Computertechnologie erzeugt wird. Dies spiegelt sich auch in der für diesen Dokumenttyp verwendeten Bezeichnung wider, welche sich aus den beiden Wörtern *CAD* und *Modell* zusammensetzt. So verdeutlicht zum einen der Begriff des Modells, dass Dokumente dieses Typs ein Abbild des zu fertigenden Produktes enthalten. Zum anderen wird dieses Abbild rechnerunterstützt, d. h. mit Hilfe sogenannter CAD-Systeme — AwS, welche die Tätigkeiten technisches Zeichnen und geometrisches Modellieren auf digitale Weise ermöglichen — konstruiert, was durch das Akronym CAD ausgedrückt wird [PBFG07, Wie08]. CAD steht dabei für die englische Bezeichnung Computer Aided

Design, welche sich gegen die auch manchmal verwendete Bezeichnung des Computer Aided Drafting durchgesetzt hat [Sch87] und gemäß dem Ausschuss für wirtschaftliche Fertigung wie folgt definiert ist:

„**CAD** ist ein Sammelbegriff für alle Aktivitäten, bei denen die EDV direkt oder indirekt im Rahmen von Entwicklungs- und Konstruktionstätigkeiten eingesetzt wird.

Dies bezieht sich im engeren Sinn auf die grafisch-interaktive Erzeugung und Manipulation einer digitalen Objektdarstellung, z. B. die zweidimensionale Zeichnungserstellung oder die dreidimensionale Modellbildung.“ [AWF85] nach [Con08d, S. 560]

Für eine detailliertere Analyse dieses Dokumenttyps werden im Folgenden zunächst die zur Erstellung derartiger Dokumente in der Praxis eingesetzten CAD-Systeme (Kapitel 5.4.2.1) sowie die für einen Datenaustausch verwendeten Dateiformate (Kapitel 5.4.2.2) näher betrachtet. Daraufhin werden mit Hilfe von Beispieldaten aus der industriellen Praxis der Informationsgehalt (Kapitel 5.4.2.3) und die Möglichkeiten zur Repräsentation dieser Informationen für eine Verwendung im LFRP-Framework (Kapitel 5.4.2.4) erläutert. Kapitel 5.4.2.5 schließt diese Betrachtung letztlich mit einer Darstellung der Evaluationsergebnisse für die zur Bildung von Ähnlichkeitsfacetten prototypisch implementierten Ähnlichkeitskonzepte ab.

5.4.2.1. 3D-CAD-Systeme: Historie und Charakteristik

Während die Computerunterstützung zu Beginn der CAD-Einführung Anfang der 1980er Jahre noch auf den 2D-Raum beschränkt war und folglich 2D-Zeichnungssysteme, die auf Arbeiten zur rechnerunterstützten Programmierung von Werkzeugmaschinen des MIT¹³ in Cambridge basierten, Einzug in die Unternehmen hielten, sind mit der Zeit immer leistungsfähigere Konstruktionsprogramme entstanden [Wie08]. Nach einer Weiterentwicklung dieser *CAD-Systeme der 1. Generation* zu $2\frac{1}{2}$ D-Systemen, die mit Hilfe von Verschiebe- oder Rotationsvektoren bereits die dritte Dimension simulierten, entstanden schließlich Ende der 1980er Jahre die ersten *CAD-Systeme der 2. Generation*. Aufgrund der fortschreitenden Technologie ermöglichten diese nun die geometrische Beschreibung einer Produktgestalt in drei Dimensionen, weshalb seitdem auch häufig von 3D- anstatt von CAD-Modellen gesprochen wird. Für die Beschreibung selbst bzw. die Speicherung der Geometriedaten wird dabei ein rechnerinternes Modell genutzt [Sch87], das im Laufe der Zeit ebenfalls immer weiter optimiert wurde. Während am Anfang nur einfache Kanten- bzw. Drahtmodelle zum Einsatz kamen, welche ein Produkt ausschließlich über Punkte und Kanten definierten und folglich kreisförmige Objekte nur über Polygonzüge annähern konnten [Sch87, Con08d], ist man bereits nach kurzer Zeit auf die Verwendung von Flächenmodellen zur internen Darstellung umgestiegen. Diese erweiterten das Kantenmodell um Daten über

die Objektoberfläche, indem diese durch Ebenen, Torus- und Freiformflächen beschrieben wurde [VWSS94, Con08d]. Doch auch diese Informationen reichten nicht aus, um beispielsweise Volumenberechnungen oder Kollisionsbetrachtungen vornehmen zu können. Aus diesem Grund wurden schließlich Volumenmodelle entwickelt, die auch als Festkörpermodelle bezeichnet werden und grundsätzlich mit Hilfe von zwei Methoden beschrieben werden können [Sch87, Con08d]. Die erste Methode besteht in der Beschreibung des Modellvolumens als Boundary Representation (BRep), wobei die Begrenzungsflächen des Modells eine Hülle bilden, deren Inhalt als Körper verwaltet wird. Bei Anwendung der zweiten Methode Constructive Solid Geometry (CSG) hingegen erfolgt die Beschreibung durch Verwendung einer Menge von Grundkörpern (Quader, Zylinder, ...), die mit Hilfe logischer Operatoren (UND, ODER, NICHT, ...) miteinander verknüpft werden [Sch87, VWSS94]. Beide Methoden sind in den volumenorientierten *CAD-Systemen der 3. Generation* vorzufinden, die heute in vielen Unternehmen eingesetzt werden. Diese weisen zusätzlich weitere Fähigkeiten auf, welche die folgenden Techniken unterstützen: [GEG05, ABB⁺06, Wie08]

- Das *parametrische Konstruieren* (Parametric Design), bei dem die Geometrie mit den zu definierenden Abmessungen assoziativ verknüpft ist, so dass Änderungen der Abmessungen automatisch zu Änderungen der Geometrie führen.
- Das *Konstruieren mit technischen Elementen* (Feature-based Design), bei dem ein CSG-Grundkörper als Basis verwendet und mit Form-Features wie Nuten, Bohrungen, Fasen usw. als Gestaltungselemente solange versehen wird, bis die gewünschte Produktgestalt erreicht ist.
- Das *Konstruieren mit Zwangsbedingungen* (Constraint Modeling), bei dem die Verwendung technischer Elemente mit der Möglichkeit, umfangreiche Regelwerke im System zu hinterlegen, kombiniert wird. Unter Zwangsbedingungen sind hier beispielsweise die Parallelität oder die Rechtwinkligkeit von Kanten zueinander zu verstehen.

Obwohl zunehmend in die Entwicklung von 4D- und sogar 5D-Systemen investiert wird [KF98, HM04, CKS08, Aut09] und auch 2D-Systeme noch immer zum Einsatz kommen, stellen 3D-CAD-Systeme zum derzeitigen Zeitpunkt den Stand der Technik in diesem Bereich dar. Folglich hat sich insbesondere in den letzten Jahren in vielen Unternehmen ein Wandel von 2D zu 3D vollzogen, was primär damit verbunden ist, dass die Generierung einer 3D-Beschreibung einige gravierende Vorteile mit sich bringt. So ist neben einer verbesserten Darstellung, welche die Betrachtung eines Produktes aus verschiedenen Blickwinkeln ermöglicht, v. a. die Verlagerung von Simulations- und Testmethoden in den digitalen Bereich als wesentlicher Treiber für diesen Wandel zu sehen [EM91, VWSS94]. Durch die Verwendung von 3D-Modellen als digitale Prototypen — sogenannte Digital Mock-Ups (DMUs) [PBFG07] — können Fehler frühzeitig im PEP erkannt und beseitigt werden, ohne Geld in die Fertigung physischer Prototypen (Physical Mock-Ups) investieren zu müssen. Des Weiteren

ren können die digitalen Produktdaten auch von anderen Unternehmensabteilungen (z. B. Marketing und Vertrieb) für diverse Zwecke (z. B. Produktpräsentationen) genutzt werden [SW08]. Außerdem haben sich 3D-CAD-Systeme immer mehr zu AwS entwickelt, die alle Aufgabenbereiche der Konstruktion bis hin zur Erstellung der Fertigungsunterlagen für ein Produkt unterstützen [Sch87, Con08d]. Folglich sind bereits heutige CAD-Systeme im Allgemeinen modular aufgebaut, so dass für die Erfüllung verschiedener Aufgaben (Modellieren, Analysieren, Zeichnen, . . .) jeweils eigene Module zur Verfügung stehen [ABB⁺06]. Diese allerdings unterscheiden sich von System zu System, wobei die Anzahl der am Markt und damit auch in der industriellen Praxis verfügbaren CAD-Systeme sehr umfangreich ist. Für einen Überblick werden daher im Folgenden nur die Systeme kurz vorgestellt, die zum gegenwärtigen Zeitpunkt der Erstellung dieser Arbeit eine große Akzeptanz und damit Verbreitung in der Industrie erlangt haben.

AutoCAD bzw. Autodesk Inventor Einer der weltweit führenden Anbieter von CAD-Systemen in den Bereichen Mechanik & Maschinenbau, Architektur, Geographische Informationssysteme, Tiefbau sowie Media & Entertainment ist das Unternehmen Autodesk, Inc. aus Kalifornien, dessen bereits im Jahr 1982 auf den Markt gebrachtes CAD-System AutoCAD noch heute zu den am weitesten verbreiteten Konstruktionsprogrammen gehört. Dabei handelt es sich bei AutoCAD grundsätzlich um ein vektororientiertes Zeichenprogramm, das mittlerweile in der 25. Version als AutoCAD 2011 erhältlich ist und je nach Anwendungsgebiet mit unterschiedlichen Konstruktionsmodulen ausgestattet werden kann. Somit gibt es speziell für den mechanischen Bereich die Erweiterung AutoCAD Mechanical, die allerdings rein speziell für die mechanische 2D-Konstruktion entwickelt wurde. Um den Wandel hin zur 3D-Konstruktion nicht zu versäumen, hat Autodesk ausgehend von dieser Basis mit Autodesk Inventor ein 3D-CAD-System entwickelt, welches derzeit als Version Inventor 2011 zusammen mit AutoCAD Mechanical in einem Paket — der Autodesk Inventor Suite — vermarktet wird. Dabei kombiniert Inventor die beiden Techniken des parametrischen und Feature-basierten Konstruierens, indem jedes Feature (d. h. jede Bohrung, jede Fase usw.) als unabhängiges Objekt behandelt wird. Dessen Geometrie kann anschließend durch einfaches Ändern der Abmessungen variiert werden, ohne das Feature neu erstellen zu müssen. Insgesamt stehen für den Produktentwurf hierbei acht Module zur Verfügung, die neben der Konstruktion von Einzelteilen (Part Modeling) u. a. die Erstellung von Baugruppen aus diesen Einzelteilen (Assembly Modeling), das Konstruieren von Blechteilen (Sheet Metal Modeling) oder auch eine Speicherung der beim Entwerfen auftretenden Gedanken als graphische Notizen (Engineer's Notebook) ermöglichen. Demzufolge können durch Anwendung dieser Software diverse Produkte entworfen werden, was in den meisten Fällen mit der Konstruktion von Einzelteilen (parts) beginnt. Diese werden als Dateien mit der Endung .ipt für Inventor Part ge-

speichert und eventuell zu Baugruppen (assembly) kombiniert. Letztere werden als Inventor Assembly-Dateien mit der Endung .iam archiviert, welche jedoch nicht die Einzelteile selbst, sondern Referenzen auf deren .ipt-Dateien enthalten. [BJ02, Aut08, Aut10b]

Pro/ENGINEER Ein anderes, sehr weit verbreitetes CAD-System wurde vom Unternehmen Parametric Technology Corporation, kurz PTC, im Jahr 1988 als Konstruktionssoftware mit dem Namen Pro/ENGINEER entwickelt. Dieses parametrische 3D-CAD-System, welches neben CAD- auch Computer Aided Engineering (CAE)- und Computer Aided Manufacturing (CAM)-Funktionalitäten integriert, wird seit dem Jahr 2002 unter dem Namen Pro/ENGINEER Wildfire vermarktet und war damit laut PTC das erste CAD-System, das auch Web-Dienste unterstützt. Mittlerweile ist Pro/E, wie es auch häufig bezeichnet wird, in der Version Wildfire 5.0 erhältlich und ist vor allem in der Automobilbranche vorzufinden. Für die Speicherung der konstruierten Produkte werden auch bei Pro/E proprietäre Datenformate mit den Endungen .prt für Einzelteile und .asm für Baugruppen verwendet. [HK02, Par10]

CATIA und SolidWorks Ebenfalls in der industriellen Praxis häufig anzutreffen sind die CAD-Systeme CATIA und SolidWorks des französischen Unternehmens Dassault Systèmes. Das ursprünglich für den Flugzeugbau entwickelte Konstruktionsprogramm CATIA, dessen Name als Akronym für **C**omputer **A**ided **T**hree-Dimensional **I**nteractive **A**pplication steht, hat sich seit seiner Markteinführung im Jahr 1981 zu einem professionellen Werkzeug für die technische PE qualifiziert und damit Dassault Systèmes als einen ernst zu nehmenden Konkurrenten auf dem CAD-Markt etabliert. Mittlerweile ist CATIA in der Version V6R2011 auf dem Markt und erweitert damit die Vorgängerversion V5 insbesondere um Techniken der dynamischen Verhaltensmodellierung und -simulation sowie der 3D-Datenoptimierung. Für die Archivierung der konstruierten Produkte werden hierbei weiterhin die mit Version V5 eingeführten Dateiformate mit den Endungen .CATPart für Einzelteile und .CATProduct für Baugruppen verwendet; jedoch findet man in der industriellen Praxis auch immer noch sogenannte .model-Dateien der früheren Version CATIA V4, welche Beschreibungen von sowohl Einzelteilen als auch Baugruppen enthalten können. Neben CATIA zählt auch SolidWorks zu der von Dassault Systèmes angebotenen Produktfamilie, dessen Rechte sich das Unternehmen im Jahr 1997 sicherte. Ziel der Entwicklung von SolidWorks war es, eine einfach zu verwendende und v. a. benutzerfreundliche CAD-Software anzubieten, die ohne teure Hard- und Software auskam und demzufolge keine hohen Investitionen erforderte. Somit ist das primär für den Maschinenbau konzipierte System innerhalb kürzester Zeit zu einem der bekanntesten CAD-Systeme geworden, das in der Version Premium 2010 laut [Das10c, Abschnitt „Unsere Errungenschaften“] mit „(...) mehr als 700 000 an mehr als 97 000 Standorten in über 100 Ländern (...)“ die höchste

Anwenderzahl weltweit aufweist. Was das Vorgehen beim Konstruieren betrifft, gleicht SolidWorks den bereits vorgestellten Systemen. Folglich sind auch hier zunächst Einzelteile zu erstellen, deren Speicherung in Form von Solid Part-Dateien (Endung `.sldprt`) erfolgt. Diese können anschließend zu Baugruppen zusammengesetzt und in Dateien mit der Endung `.sldasm` für Solid Assembly abgelegt werden. [Vog06, Das10a, Das10b]

NX und Solid Edge Zwei weitere bekannte CAD-Systeme sind die Werkzeuge NX und Solid Edge, welche ursprünglich vom Softwarehersteller Unigraphics Solutions Inc. (UGS) entwickelt wurden und seit 2007 aufgrund einer Firmenübernahme von der Siemens AG — genauer gesagt dessen Geschäftsbereich Siemens PLM Software — vertrieben werden. Beide Systeme basieren dabei auf dem ebenfalls von UGS entwickelten Parasolid-Modellierkern und nutzen die von Siemens PLM Software entwickelte Technik *Synchronous Technology*. Dabei handelt es sich gemäß [Sie10d] um eine Kombination aus direkter Geometriemodellierung und der Verwendung eines Constraint-Solvers, wodurch eine direktere, interaktivere, schnellere und v. a. flexiblere Modellierung — jedoch unter parametrischer Kontrolle — ermöglicht werden soll. Trotz dieser Gemeinsamkeiten weisen beide Systeme aber auch Unterschiede auf. So sind beide Systeme einerseits auf verschiedene Zielmärkte ausgerichtet, weshalb sie diese neue Technologie auf unterschiedliche Weise nutzen. Zum anderen lassen sie sich hinsichtlich der zur Archivierung der konstruierten Produkte verwendeten Datenformate voneinander abgrenzen. Während das aktuell in der Version 7 erhältliche NX-System sowohl Einzelteile als auch Baugruppen mittels nur eines proprietären Datenformates, das die Endung `.prt` besitzt, archiviert, wird bei Solid Edge hingegen zwischen drei Formaten unterschieden. So werden Einzelteile als Part-Dateien (Endung `.par`), Baugruppen im Dateiformat mit der Endung `.asm` für Assembly und schließlich Blechteile in Form von Part Sheet Metal-Dateien (Endung `.psm`) gespeichert. [Sch05a, Sie10b, Sie10c]

Wie die Erläuterung zeigt, verfügen alle CAD-Systeme über proprietäre Datenformate zur Speicherung der CAD-Modelle. Diese allerdings erfordern für einen Zugriff auf die in ihnen enthaltenen Daten die Verwendung von kostspieligen Konvertern (= Post- und Preprozessoren), was bei einer Nutzung von unterschiedlichen CAD-Systemen, wie es beispielsweise bei Unternehmen der Automobilindustrie vonnöten ist, hohe Kosten verursachen würde. Diese Kosten resultieren jedoch nicht nur aus den hohen Lizenzgebühren für die diversen Konverter, sondern entstehen v. a. aufgrund des erforderlichen hohen manuellen Aufwands zur Nachbearbeitung der Modelle, da jeder Konvertierungsvorgang mit einem Datenverlust verbunden ist [Wie08]. Aus diesem Grund kam schon bald die Forderung nach einem Format, das „(...) sicherstellen muss, daß die Produktdaten auch noch nach vielen Jahren z. B. für die Ersatzteilfertigung von den dann benutzten CAD-Systemen lesbar sind.“ [Wie08, S. 130], auf. Dies führte schließlich zur Entwicklung verschiedener Standarddateiformate, die einer-

seits diese Form der Archivierung gewährleisten und andererseits den Datenaustausch mit Lieferanten und Kunden vereinfachen. Zudem bieten alle gängigen CAD-Systeme eine automatische Konvertierung ihrer nativen Daten in diese neutralen Formate an [AM09]. Daher sind diese Austauschformate für den vorliegenden Anwendungsfall von besonderem Interesse, weshalb sie im Folgenden u. a. im Hinblick auf die von ihnen bereitgestellten Möglichkeiten zur Speicherung von Informationen untersucht werden.

5.4.2.2. Austauschformate für 3D-Modelle

Um eine Übertragung von CAD-Daten in andere CAD- oder sonstige IT-Systeme zu vereinfachen, wurden im Verlauf der Jahrzehnte verschiedene Austauschformate von Normungsinstituten, Verbänden oder auch Unternehmen entwickelt. Demzufolge findet man je nach Unternehmen andere Formate, von denen neben bereits teilweise nicht mehr weiterentwickelten Standards wie VDA-FS (Flächenschnittstelle des Verbandes der Automobilindustrie) oder STL⁸³ (StereoLithography-Format) sowie dem noch eher selten anzutreffenden 3D-PDF-Format von Adobe insbesondere vier eine hohe Relevanz in der industriellen Praxis erlangt haben. Hierbei handelt es sich um die Austauschformate IGES, STEP, JT und VRML, welche im Folgenden etwas detaillierter betrachtet werden.

IGES – Initial Graphics Exchange Specification Bereits im Jahr 1979 bildete sich auf Drängen diverser CAD-Anwender ein Konsortium aus diesen und verschiedenen CAD-Anbietern, das unterstützt vom US National Bureau of Standards⁸⁴ sowie dem amerikanischen Verteidigungsministerium das Ziel verfolgte, ein Datenformat zur Übertragung einfacher 2D-Geometriedaten zwischen Systemen verschiedener Hersteller zu entwickeln. Damit war das nach zwei Jahren vom American National Standards Institute (ANSI) als Norm Y14.26M standardisierte Format IGES geboren, welches zwar zunächst auf die Übertragung von Kantenmodellen beschränkt war, jedoch ab Version 4.0 auf den Transfer einfacher CSG-basierter und schließlich mit Version 5.0 BRep-basierter Volumenmodelle erweitert wurde. Dieses Datenformat mit der Endung .igs legt dabei eine für die Datenkonvertierung benötigte Menge von geometrischen und nicht-geometrischen Elementen mit ihren jeweiligen Darstellungsformen fest, wobei prinzipiell zwischen den vier, in Abbildung 5.10 veranschaulichten Elementgruppen *Objektbeschreibung*, *Strukturelemente*, *Bemaßung* und *vordefinierte Assoziationen* differenziert wird. Folglich werden alle produktdefinierenden Daten eines Produktes durch eine Teilmenge dieser Elemente und ihrer zugehörigen

⁸³<http://www.ennex.com/~fabbers/StL.asp>

⁸⁴Hierbei handelt es sich heute um das National Institute of Standards and Technology (<http://www.nist.gov>).

Ausprägungen repräsentiert [US 96]. Zur Übertragung dieser Elemente einigte man sich auf den Aufbau einer IGES-Datei aus Datensätzen im American Standard Code for Information Interchange (ASCII)-Format, welche eine feste Länge von 80 Spalten besitzen. Diese Datensätze bilden dabei eine Sequenz von insgesamt sechs oder fünf Abschnitten, die mit einer optionalen *Flag Section* zur Kennzeichnung komprimierter ASCII-Dateien oder einer *Start Section*, welche einen für den Anwender lesbaren Kommentar enthält, beginnt. Darauf folgen eine *Global Section* mit notwendigen organisatorischen Informationen u. a. für die Post- und Preprozessoren, eine *Directory Entry Section* mit einer Auflistung der zu übertragenden Elemente sowie eine *Parameter Data Section*, welche die zu den vorher definierten Elementen zugehörigen Parameterwerte enthält. Eine *Terminate Section*, in der Prüfdaten für die vorhergehenden Abschnitte und die Abbruchanweisung für den Prozessorlauf festgelegt sind, zeigt schließlich das Ende der Datei an. Durch diesen beschriebenen Aufbau lassen sich insgesamt relativ kleine Datenmengen und folglich geringe Speicheranforderungen realisieren. Allerdings wurde die Entwicklung von IGES mit Aufkommen des neuen Standards STEP nicht mehr weiter vorangetrieben, so dass die Version 5.3 aus dem Jahr 1996 den letzten veröffentlichten Standard hierzu darstellt. [And93, US 96, Kem01, Nat08]

STEP – Standard for the Exchange of Product Model Data (ISO 10303) Um die Übertragung von Produktdaten nicht nur allein auf den geometrischen Aspekt zu beschränken, begann man Mitte der 1980er Jahre damit, die Eigenschaften verschiedener bereits existierender Austauschformate in einem übergreifenden Standard zu vereinen. Aus der Zusammenarbeit verschiedener Länder und des von der ISO eingesetzten Komitees ISO TC184/SC4⁸⁵ resultierte schließlich im Jahr 1994 mit der Normenreihe ISO 10303 ein internationaler Standard, welcher ständig erweitert wurde und bis heute fortentwickelt wird. Im Gegensatz zu anderen Austauschformaten definiert STEP dabei ein integriertes Produktmodell, welches die Berücksichtigung aller in einem Unternehmen eingesetzten Systeme (CAD, CAE, CAM, PDM, ...) und damit einen schnellen und sicheren Austausch aller während des Produktlebenszyklus entstehenden interdisziplinären Informationen zu einem Produkt ermöglichen soll. Hierzu werden die verschiedenen Sichten auf ein Produkt durch Partialmodelle erfasst, die durch Anwendungsprotokolle (AP) auf der Grundlage anwendungsunabhängiger Basismodelle — sogenannter Integrated Resources zur Definition grundlegender Datenklassen — repräsentiert werden. Mittels dieser Basismodelle beschreibt jedes AP dabei die Produktdaten aus einem anderen anwendungsspezifischen Aspekt. Während beispielsweise das AP 203 (Configuration Controlled 3D Designs of Mechanical Parts and Assemblies) die für einen Austausch

⁸⁵Technical Committee 'Industrial Automation Systems & Integration' / Subcommittee 'Industrial Data'

Objektbeschreibung		Strukturelemente	vordefinierte Assoziationen
Geometrie Point Connect Point Conic Arc Coplous Data 2D Path 3D Path Closed Planar Curve Surface Line Offset Curve Parametric Spline Curve Parametric Spline Surface Plane Ruled Surface Surface of Revolution Tabulated Cylinder Rational B-Spline Curve Rational B-Spline Surface Right Circular Cylindrical Surface Right Circular Conical Surface Spherical Surface Toroidal Surface Flash Transformation Matrix	FEM-Modell Finite Element Nodal Displacement/Rotation Nodal Results Element Results CSG-Modell Block Right Ang. Wedge Right Circ. Cylinder Right Circ. Cone Frustum Sphere Torus Solid of Revolution Solid of Linear Extrusion Ellipsoid Boolean Tree Solid Instance Solid Assembly Selected Component	Associativity Definition Associativity Instance Drawing Line Font Definition Macro Definition Macro Instance Property Subfigure Definition Network Subfigure Definition Singular Subfigure Instance Rectangular Array Subfigure Instance Circular Array Subfigure Instance Network Subfigure Instance Text Font Definition View External Reference Nodal Load/Constraint Text Display Template Text Display Instance Colour Definition Attribute Table Definition Attribute Table Instance Null Node Units Data	Group w/o Backpointers Extern. Ref. File Index Views Visible Views Visible, Colour Line Weight Entry Label Display Single Parent Ass. Dimensional Geometry Ass. Ordered Group w/o Backp. Planar Ass. Flow
	B-Rep Solid Modell Manifold Solid B-Rep Object Vertex, Edge, Loop Face, Shell		Bemaßung Angular Dimension Diameter Dimension Flag Note, General Label General Note/new Gen. Note Leader (Arrow) Linear Dimension Ordinate Dimension Point Dimension Radius Dimension General Symbol Sectionated Area Copius Date(Center Line, witness line section)

Abbildung 5.10. — Elemente des IGES-Datenformates in der Version 5.1 (nach [And93, S. 185])

von Geometrie-, Topologie- und Konfigurationsdaten (z. B. Version, Freigabestatus) eines mechanischen Produktes verwendbaren Elemente definiert, beschreibt AP 218 (Ship structures) die Elemente, die für den Transfer von Schiffsstrukturen relevant sind. Darüber hinaus ist jedes AP aus sogenannten Konformitätsklassen aufgebaut, welche jeweils Teilmengen des APs definieren und für eine Konformität mit dem Standard von einem Übersetzungsprogramm mindestens zu implementieren sind. So enthält beispielsweise das AP 203 gemäß Pratt sechs Konformitätsklassen, von denen die Konformitätsklasse 1 (Configuration-controlled design information without shape) in jedem Fall zu unterstützen ist, da die anderen fünf diese lediglich um Gestaltinformationen unterschiedlicher rechnerinterner Modelle ergänzen [Pra01]. Zusätzlich zu diesen Partialmodellen bzw. APs stellt STEP u. a. auch Beschreibungs- und Implementierungsmethoden bereit, die angeben, wie die Datenmodelle mit Hilfe der eigens hierfür entwickelten objektorientierten, formalen Sprache EXPRESS zu spezifizieren bzw. wie diese in einem Anwendungssystem zu implementieren sind. Zur tatsächlichen Übertragung der Produktdaten beschreibt zudem ISO 10303-21 (Clear Text Encoding of the Exchange Structure), wie eine ASCII-basierte Datei mit der Endung .stp oder .step aufgebaut sein muss. Danach sind zwei Abschnitte, ein

HEADER- und ein darauf folgender DATA-Abschnitt zu definieren, die jeweils durch das Schlüsselwort ENDSEC begrenzt werden. Der HEADER hat hierbei einer fixen Struktur zu folgen, welche aus einer Sequenz der Unterelemente FILE_DESCRIPTION, FILE_NAME, FILE_SCHEMA, FILE_POPULATION, SECTION_LANGUAGE und SECTION_CONTEXT zur Definition organisatorischer Daten besteht. Der Datenabschnitt hingegen beinhaltet schließlich die eigentlichen zu übertragenden Produktdaten, wobei jedes Element zur Beschreibung dieser Daten in einer separaten Zeile mit einer eindeutigen Instanznummer, dem Elementnamen sowie in Klammern gesetzten Attributwerten zu definieren ist. Letztere können dabei neben konkreten Werten u.a. auch Referenzen auf zuvor definierte Elemente unter Angabe deren Instanznummer enthalten. [And93, AT00, Pra01, SCR06, Int10b]

JT – Jupiter Tessellation Ein noch relativ junges Austauschformat ist das ursprüngliche von UGS auf den Markt gebrachte und mittlerweile insbesondere von Siemens PLM Software unterstützte JT-Format. Dessen Entwicklung begann im Jahr 1999 mit der Übernahme des vom Softwarehersteller Engineering Animation, Inc. entwickelten Direct Model Formates durch UGS, um einen Industriestandard für die systemneutrale Visualisierung, Analyse und Optimierung von 3D-Geometrien und Produktstrukturen zu schaffen. Da ein Standard allerdings nicht von einem Unternehmen allein etabliert werden kann, gründete UGS nach vier Jahren die Initiative JT Open, welche von zahlreichen Unternehmen wie Siemens, PTC oder SAP unterstützt wird und damit die Weiterentwicklung und Verbreitung des Formates in der Industrie vorantreibt. Nachdem Anfang 2007 die erste offizielle Spezifikation des Datenformates veröffentlicht wurde, stellt JT laut [Wik10d] mit über vier Millionen Lizenzen mittlerweile das am häufigsten eingesetzte Datenformat zur Visualisierung von 3D-Daten dar. Folglich wurde seine Spezifikation im September 2009 auch von der ISO zur Veröffentlichung als ISO/PAS⁸⁶ 14306 akzeptiert. Der Vorteil dieses Datenformates ist insbesondere in seiner hohen Flexibilität zu sehen, da Dateien mit der Endung .jt einerseits nur wenige, einfache Daten, aber andererseits auch umfangreiche und präzise Geometriedaten, Produktstrukturen und Fertigungsinformationen enthalten können. Um dies zu ermöglichen, setzt sich das JT-Datenmodell insgesamt aus den neun Komponenten *Product Structure*, *Facet*, *Lighting*, *Textures*, *Precise Geometry* und *Topology*, *Geometry Primitives*, *Product Manufacturing Information*, *Attributes/Properties* und *Meta data* zusammen. Durch den Einsatz spezieller Kompressionsverfahren verfügen JT-Dateien außerdem über eine deutlich geringere Größe als native CAD-Daten (laut [Wik10d] nur etwa 1–10 % der Originalgröße), was den Datenaustausch und damit die Zusammenarbeit wesentlich erleichtert. Aus diesem Grund wird JT auch häufig als sogenanntes

⁸⁶PAS steht als Akronym für Publicly Available Specification und resultiert aus der zweiten Phase des ISO-Normungsprozesses [Int10c].

Lightweight-Format bezeichnet. Zur Abbildung von Produktstrukturen differenziert JT zwischen den drei Knotentypen *JtkAssembly*, *JtkPart* und *JtkInstance*, die mittels drei unterschiedlicher Methoden gespeichert werden können. So wird beispielsweise bei der als *Per Part* bezeichneten Speicherung jeder *JtkAssembly*-Knoten in einer separaten JT-Datei und die zugehörigen *JtkPart*-Knoten in entsprechenden Unterdateien abgelegt. Während dagegen die *Fully shattered*-Methode für jeden Knoten der Produktstruktur eine eigene JT-Datei erzeugt, ist bei der *Monolithic*-Speicherung die gesamte Produktstruktur in einer einzigen JT-Datei enthalten. [Sie09, Pro09, Sie10e, Wik10d]

VRML – Virtual Reality Modeling Language Obwohl VRML ursprünglich 1994 als 3D-Standard für das Internet entwickelt wurde, ermöglichen die meisten der gängigen CAD-Systeme heutzutage den Import und Export von Dateien mit der Endung *.wrl* als Akronym für *world*. Dies ist u. a. darauf zurückzuführen, dass es mit Hilfe dieser einheitlichen Spezifikation nicht nur möglich ist, virtuelle Welten, d. h. interaktive 3D-Objekte und -Szenen zu erzeugen, sondern diese auch mit jedem beliebigen Webbrowser nutzen zu können. Zusätzlich wurde die erweiterte Spezifikation namens VRML97 im Jahr 1997 als ISO/IEC 14772-Standard weltweit anerkannt und von der Industrie akzeptiert, so dass VRML auch im Bereich der technischen PE Anwendung fand und hier häufig zum Austausch von 3D-Daten — insbesondere zur Weitergabe an Animations- und Renderingsoftware — eingesetzt wird. Dabei handelt es sich bei einer VRML-Datei um ein Dokument im ASCII-Format, das auf dem sogenannten Szenegraphen-Konzept basiert und ähnlich zu HTML von einem Webbrowser interpretiert werden kann. Dies wird erreicht, indem vordefinierte Knotentypen wie beispielsweise für geometrische Primitive (Kugeln, Quader, Zylinder usw.), Lichtquellen oder Anker zur Verfügung gestellt werden, die hierarchisch zu Objekten gruppiert und schließlich in einem Szenegraphen zur Beschreibung der Beziehungen zwischen den einzelnen Objekten angeordnet werden. Da im Zuge des technologischen Fortschritts weitere Anforderungen eine Optimierung des Standards erforderten, wurde VRML im Jahr 2004 durch das ebenfalls standardisierte Nachfolgeformat X3D, was für Extensible 3D steht, ersetzt. Dieses erweitert seinen Vorgänger zwar u. a. um neue Knotentypen, multiple Möglichkeiten der Datenkodierung sowie eine nun XML-basierte Beschreibung der virtuellen Welten, konnte sich jedoch bis zum jetzigen Zeitpunkt nicht wirklich übergreifend etablieren. Folglich ist VRML weiterhin als ein wichtiges Format im Bereich der technischen PE zu betrachten. [Sch98, Web10]

5.4.2.3. Informationsgehalt von CAD-Modellen

Wie die vorherigen Kapitel gezeigt haben, werden CAD-Modelle in diversen Dateiformaten gespeichert. Da für einen Zugriff auf proprietäre Datenformate teure Konverter

benötigt werden, die im Rahmen des dieser Arbeit zugrunde liegenden FORFLOW-Teilprojektes nicht zur Verfügung standen, basieren die folgenden Erläuterungen auf der Analyse von industriellen, in den zuvor erläuterten Austauschformaten vorliegenden Testdaten und den zugehörigen Formatspezifikationen. Dabei ist jedoch zu berücksichtigen, dass die getroffenen Aussagen grundsätzlich auch für proprietäre Datenformate Gültigkeit besitzen und folglich auf diese übertragbar sind. Des Weiteren wird darauf hingewiesen, dass Austauschformate aufgrund erforderlicher Mechanismen zum Schutz des geistigen Eigentums beim Datenaustausch prinzipiell einen niedrigeren Informationsgehalt besitzen als proprietäre Datenformate. Obwohl letztere durchaus über einen größeren, teilweise auch qualitativ besseren Informationsgehalt verfügen können, ist es nicht Ziel der vorliegenden Arbeit, alle prinzipiell unterstützten Informationsarten vollständig zu berücksichtigen. Vielmehr sollen Möglichkeiten aufgezeigt werden, die sich bereits bei Verwendung einfacherer Austauschformate realisieren lassen.⁸⁷

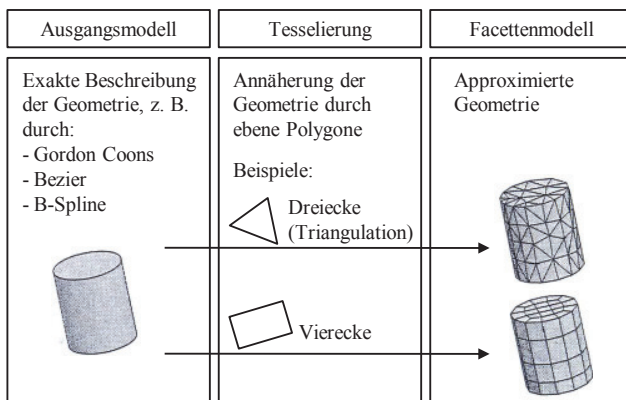


Abbildung 5.11. — Approximation einer Zylindergeometrie durch Tessellierung (nach [AB⁺07, S. Y 15])

Der Informationsgehalt von CAD- bzw. 3D-Modellen lässt sich prinzipiell in mehrere Bestandteile untergliedern, welche jeweils unterschiedliche Aspekte eines Produktentwurfes fokussieren. Der dabei wichtigste Aspekt ist die Beschreibung der geometrischen Gestalt bzw. Form eines Produktes in Form einer maßstäblichen, dreidimensionalen Darstellung, aus der nicht nur die Funktion und Struktur, sondern auch die Herstellung des Produktes ersichtlich ist [Hab08]. Somit bestehen 3D-Modelle zum

⁸⁷Gleiches gilt im Übrigen auch für die Erläuterungen im Kapitel über technische Zeichnungen (siehe Kapitel 5.4.4).

größten Teil aus Geometrieinformationen, die bei einer Transformation in ein Austauschformat üblicherweise vereinfacht werden. Der Grund für diese Vereinfachung ist in der Anwendung von Basisfunktionalitäten zum Schutz des geistigen Eigentums, d. h. des unternehmensspezifischen Produktwissens, als auch in der Notwendigkeit einer Komplexitätsverringerung zur Visualisierung und Optimierung durch CAE-Methoden zu sehen [BKK⁺04, AM09]. Allerdings führt diese Vereinfachung in den meisten Fällen dazu, dass alle unternehmensspezifischen Daten zu einem Produkt (z. B. dessen Feature-Historie oder die während des Entwurfs getroffenen Entscheidungen) entfernt werden und die Beschreibung der Produktgestalt in einer approximierten Flächendarstellung — einem sogenannten Facettenmodell — erfolgt [AB⁺07]. Wie Abbildung 5.11 verdeutlicht, wird dabei die Gestalt eines Produktes einer Tessellierung unterzogen, was eine Annäherung der Oberfläche durch ein Netz von geometrisch einfachen, zueinander angeordneten Ebenen bewirkt. Als Ebenen können dabei verschiedene Polygone zum Einsatz kommen, wobei in den meisten Fällen für einen höheren Genauigkeitsgrad Dreiecke verwendet werden. Demzufolge wird dieser Vorgang auch als Triangulation und die daraus resultierenden Geometriedaten als triangulierte Daten bezeichnet. Zwar ermöglichen die Standards JT, STEP und IGES auch die Beschreibung von Geometriedaten mittels der BRep-Methode, was jedoch aufgrund der oben genannten Gründe eher selten erfolgt. Sind diese Daten allerdings doch in einer CAD-Datei enthalten, können daraus unter Umständen weitere produktbezogene Merkmale wie beispielsweise Abmessungen ermittelt werden. Dieser Aspekt wird durch ein Beispiel im folgenden Kapitel 5.4.2.4 noch genauer verdeutlicht.

Zusätzlich zur Geometrie eines Produktes sollten CAD-Modelle auch Aufschluss über den Aufbau eines Produktes, d. h. dessen Bau- bzw. Produktstruktur, geben. Dies gilt insbesondere für Baugruppen, die sich aus mehreren Komponenten, d. h. aus Einzelteilen und Unterbaugruppen, zusammensetzen. Diese Information ist in proprietären Datenformaten enthalten, da ihre Erzeugung auf der Definition der einzelnen Komponenten basiert, die in einer hierarchischen Baumstruktur im CAD-System angelegt und gespeichert werden. Aber auch die betrachteten Austauschformate — mit Ausnahme von VRML — bieten die Möglichkeit einer topologischen Produktbeschreibung, die angibt, aus welchen Komponenten und in welcher Art und Weise ein Produkt aufgebaut ist.

Des Weiteren verfügen alle Austauschformate über die Möglichkeit Metadaten zu transportieren. Dabei gruppiert allerdings jedes Format andere Daten unter diesem Begriff, weshalb im Folgenden jedes Austauschformat diesbezüglich separat betrachtet wird.

VRML Ausgehend von der Formatspezifikation ist für VRML-Dateien kein eigener Abschnitt für die Definition von Metadaten vorgesehen. Vielmehr ist VRML rein auf die Beschreibung der Produktgestalt und der mit diesem Produkt verbundenen Interaktionsmöglichkeiten fokussiert. Allerdings besteht prinzipiell die

Möglichkeit, bei einem Export der nativen Daten in das VRML-Format Metadaten über das Produkt oder sonstige organisatorische Daten mit in die Datei zu schreiben. Dazu müsste jedes Metadatum lediglich als Name-Wert-Paar deklariert, mit einem '#'-Zeichen als Kommentar gekennzeichnet und vor die eigentliche Szenenbeschreibung eingefügt werden. Wie in Listing 5.7 verdeutlicht, lassen sich so diverse zusätzliche Informationen nicht nur auf einfache Weise mit übertragen, sondern können auch relativ einfach mittels Pattern Matching aus der Datei wieder ausgelesen werden.

Listing 5.7 — Möglichkeit zur Speicherung von Metadaten in VRML-Dateien

```
#VRML V2.0 utf8
#
#Ersteller = Weber
#Erstellungsdatum = 01.07.2006
#Exportsystem = Catia V4
#Produktname = Sechskantmutter
#Sachnummer = 4951-005
#Funktion = befestigen
#Innenradius = 1.75 cm
#
Group {
  children [
    DEF _X1_X2 Shape {
      appearance Appearance {
        material Material {
          diffuseColor 0.594824 0.594824 0.594824
          ambientIntensity 0.200000
          specularColor 0.662588 0.662588 0.662588
          emissiveColor 0.158118 0.158118 0.158118
          shininess 0.210000
          transparency 0.000000
        }
      }
      geometry IndexedFaceSet {
        coord Coordinate {
          point [
            0.002282 0.001244 -0.002349,
            0.002487 0.001207 -0.002230,
            ...
          ]
          ...
        }
      }
    }
  ]
}
```

IGES Für die Speicherung und Übertragung von Metadaten wird bei IGES der Dateiabschnitt *Global Section* verwendet, der gemäß Abbildung F.1 in Anhang F diverse Parameter für unterschiedlichste organisatorische Daten zur Verfügung stellt. Hierzu gehören zum einen produktbeschreibende Parameter wie die ProduktIDs des Sender- und des Empfänger-Systems (vgl. Index 3 und 12), der Modellmaßstab (Index 13) sowie die verwendete Maßeinheit (Index 14 und 15). Darüber hinaus können dokumentspezifische Informationen hinterlegt werden, die Auskunft geben über den Dateinamen (Index 4), das Quell-CAD-System (Index 5), die Version des verwendeten Preprozessors (Index 6), das Datum und den Zeitpunkt der Dokumenterstellung (Index 18), die der Dateierstellung zugrunde liegende Version der Formatspezifikation (Index 23) sowie das Datum der letzten Dateiänderung (Index 25). Außerdem sind spezielle Parameter für den Namen des Dokumenterstellers (Index 21) sowie für Informationen zur Organisation, welcher der Autor angehört (Index 22), in dieser *Global Section* verfügbar.

STEP Neben den im HEADER einer STEP-Datei enthaltenen allgemeinen organisatorischen Metadaten (vgl. Kapitel 5.4.2.2), können weitere Informationen auch im DATA-Abschnitt einer Datei gespeichert werden. Hierbei ist jedoch zu berücksichtigen, dass die Art und der Umfang dieser Metadaten dabei vom jeweils unterstützten AP abhängen. Da im Bereich der technischen PE primär die beiden APs 203 und 214 (Core Data for Automotive Mechanical Design Processes) eingesetzt werden, sollen deren Möglichkeiten hier stellvertretend aufgezeigt werden. Gemäß [SCR06] unterstützt das AP 203 dabei die Speicherung folgender Informationen:

- Daten zur Definition bzw. Identifikation eines Produktes sowie dessen entwurfsspezifischer Konfigurationssteuerung (d. h. Version, Freigabestatus, ...)
- Änderungsdokumentationen
- Informationen über zu beachtende rechtliche, industrielle oder unternehmensspezifische Richtlinien
- Daten zur Identifikation der in einem Produkt enthaltenen Norm- bzw. Zukaufteile
- Informationen zu Lieferanten, gegebenenfalls mit notwendigen Qualifikationskriterien
- Auftragsbezogene Daten (sofern ein Kundenauftrag existiert)
- Für Produkttests oder -analysen erforderliche Daten bzw. deren Ergebnisse, solange diese Designentscheidungen verdeutlichen
- Textuelle Anmerkungen zu einer Produktgestalt

Diese werden im AP 214 schließlich um Informationen erweitert, die speziell für den Automobilbereich von Bedeutung sind. Dazu gehören Elemente zur Speicherung von Werkstoffdaten, Fertigungswerkzeugen und ihren Eigenschaften, Fertigungsplänen, NC-Daten, Produkteigenschaften, Daten für kinematische Simulationen, Toleranzinformationen oder auch von bei der Produktkonstruktion verwendeten Form-Features. [Pra01, SCR06]

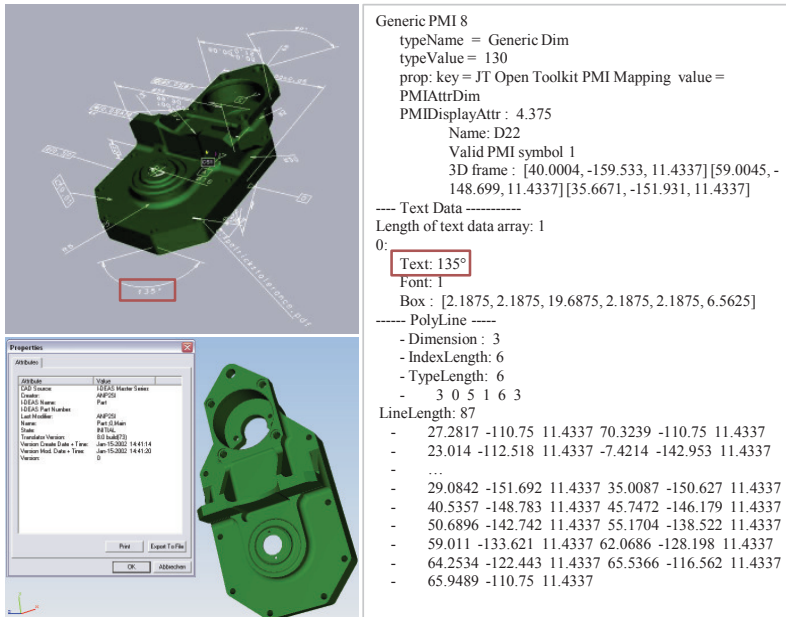


Abbildung 5.12. — Visualisierung eines Produktes im JT-Format mit PMI-Daten (links oben) und als *Properties*-Elemente gespeicherten Metadaten (links unten) sowie die zu einem PMI-Datum (rote Umrandung) extrahierbaren Informationen (rechts)

JT Auch das JT-Format stellt für die Speicherung von Metadaten zwei separate Abschnitte bereit [Sie09]. Zum einen können mit Hilfe des sogenannten *PMI Data Segment*⁸⁸ alle erforderlichen Informationen zur Herstellung des Produktes, wie Lage- und Formtoleranzen, Bemaßungen, Oberflächengüte, Schweißvermerke, und andere aufgenommen werden. Zur Verdeutlichung zeigt Abbildung 5.12 im linken oberen Teilbild die Visualisierung eines Produktes mit seinen zugehörigen

⁸⁸PMI steht für Product and Manufacturing Information

gen PMI-Daten. Das Auslesen dieser Daten aus der zugehörigen Datei resultiert allerdings darin, dass zwar der textuelle Inhalt dieser PMI-Daten extrahiert werden kann, die Semantik dahinter jedoch verborgen bleibt. Als Beispiel hierfür zeigt Abbildung 5.12 im rechten Teilbild die zu dem in Abbildung 5.12 links oben, rot markierten PMI-Datum extrahierbaren Informationen. Zum anderen können Metadaten, die nicht die Fertigung eines Produktes adressieren, in einem sogenannten *Meta data segment* gespeichert werden. Hierzu werden *Properties*-Elemente bereitgestellt, die in Form von Name-Wert-Paaren beispielsweise zur Definition organisatorischer Daten dienen. Wie Abbildung 5.12 im linken unteren Teilbild illustriert, können damit u. a. Informationen zum Quellsystem (*CAD Source*), zur Person, die das Dokument erstellt (*Creator*) bzw. geändert (*Last Modifier*) hat, zum Erstellungs- (*Version Create Date + Time*) und Änderungsdatum (*Version Mod. Date + Time*) oder zur Version (*Version*) enthalten sein.

Neben diesen drei wesentlichen Informationsarten Geometrie, Produktstruktur und Metadaten bieten alle Austauschformate die Option, die im CAD-System festgelegten Materialinformationen bzw. die Angabe, aus welchem Werkstoff ein Produkt besteht, zu speichern. Dazu sind diese entweder als Metadaten zu deklarieren (STEP) oder können durch eigens hierfür bereitgestellte Elemente definiert werden (VRML, JT, IGES). Analysiert man allerdings letztere, so zeigt sich, dass die Angabe des Werkstoffes jeweils durch Farbinformationen approximiert wird. Als Beispiel hierfür wird der in Listing 5.7 veranschaulichte Dateiinhalt eines VRML-Modells herangezogen, welcher unter Verwendung des *Material*-Elementes verschiedene Farbparameter für die Visualisierung des jeweiligen Werkstoffes definiert. Gleiches gilt im Übrigen auch für das JT-Format mit den hierfür verfügbaren *JtkMaterial*-Knoten, was darauf zurückzuführen ist, dass es sich hierbei ebenfalls um ein Datenformat vorrangig zur Visualisierung von 3D-Objekten handelt. Allerdings stellen diese Farbinformationen keine sinnvollen Suchkriterien dar, weshalb sie nicht weiter betrachtet werden.

5.4.2.4. Extraktion und Repräsentation von CAD-Daten

Um die in CAD-Modellen enthaltenen Informationen für eine Suche im LFRP-Framework nutzen zu können, müssen die Informationen in einem ersten Schritt extrahiert, d. h. aus den Dateien gelesen, und anschließend geeignet aufbereitet, d. h. repräsentiert, werden. Aus diesem Grund wird im Folgenden zunächst ein Überblick über die zur Extraktion der Informationen verwendbaren Softwarebibliotheken für eine Implementierung in der Programmiersprache Java gegeben. Danach wird erläutert, wie die einzelnen Bestandteile des Informationsgehalts eines 3D-Modells repräsentiert werden sollten.

Java-Bibliotheken zur Informationsextraktion

Um eine Verarbeitung von dreidimensionalen Daten zu gewährleisten, kann grundsätzlich Java3D — eine Klassenbibliothek zur Erzeugung, Manipulation und Darstellung dreidimensionaler Grafiken — eingesetzt werden. Aktuell in der Version 1.5.2 für mehrere Plattformen verfügbar, ist Java3D als Open Source-Bibliothek über die Internetseite <https://java3d.dev.java.net> erhältlich. Darüber hinaus erfordert die Berücksichtigung von 3D-Dateien in den unterschiedlichen Austauschformaten den Einsatz weiterer Softwarebibliotheken, welche nachfolgend kurz erläutert werden:

VRML Zum Lesen und Schreiben von VRML-Dateien existieren prinzipiell zwei Möglichkeiten. Zum einen kann Java3D um das Subprojekt J3D-VRML97⁸⁹ erweitert werden, welches eine `Java3DLoader`-Klasse für 3D-Modelle im VRML97-Format enthält. Zum anderen kann als Alternative auch die von Satoshi Konno entwickelte CyberVRML97 for Java-Bibliothek genutzt werden, welche ebenfalls kostenfrei erhältlich ist und bei SourceForge.net über die Internetadresse <http://sourceforge.net/projects/cv97java/> heruntergeladen werden kann.

JT Um JT-basierte Anwendungen erstellen zu können, wird von der Vereinigung JT Open grundsätzlich das JT Open Toolkit⁹⁰ zur Verfügung gestellt. Dessen Nutzung ist jedoch mit zwei negativen Aspekten verbunden. Der erste negative Aspekt ist in der mit Kosten verbundenen Nutzung der Bibliothek zu sehen, da nur Mitglieder der JT Open-Gemeinschaft Zugriff auf diese Bibliothek erhalten. Der zweite negative Aspekt betrifft die Fokussierung der Bibliothek auf nur eine Programmiersprache — nämlich C++. Allerdings kann durch einen Einsatz der standardisierten Anwendungsprogrammierschnittstelle Java Native Interface (JNI), welche einen Aufruf von plattformspezifischen und damit auch C++-Funktionalitäten aus einer Java-Anwendung ermöglicht, diesem Mangel begegnet werden. Ein positiver Aspekt der Nutzung des JT Open Toolkits ist allerdings in der Unterstützung verschiedener Zugriffsoptionen auf JT-Dateien zu sehen, welche zur Extraktion der einzelnen Informationen genutzt werden können. So erlaubt beispielsweise die `TESS_ONLY`-Option den Zugriff auf nur tesselierte Geometriedaten, während die `BREP_ONLY`-Option zur Anzeige BRep-basierter Geometrien oder die `NO_GEOM`-Option einen Lesezugriff allein auf die Produktstruktur ermöglichen.

STEP Auch für den Zugriff auf STEP-Dateien stehen prinzipiell zwei Möglichkeiten zur Verfügung. Die erste Möglichkeit besteht in der Verwendung der Java-Bibliothek jCAE, welche über SourceForge.net (<http://jcae.sourceforge.net>) in der aktuellen Version 0.16 erhältlich ist. Hierbei handelt es sich, wie der

⁸⁹<https://j3d-vrml97.dev.java.net>

⁹⁰http://www.plm.automation.siemens.com/en_us/products/open/jtopen/technology/jt_open_toolkit.shtml

Name zeigt, um eine speziell für den CAE-Bereich entwickelte Bibliothek, die über einen JNI-Wrapper Funktionalitäten der frei verfügbaren C++-Software-entwicklungsplattform Open CASCADE Technology⁹¹ bereitstellt. Letztere wurde vom Unternehmen Open CASCADE S.A.S.⁹² entwickelt, um speziell die Entwicklung von CAD-, CAM- und CAE-Anwendungen zu unterstützen. Daher bietet die Plattform u. a. Funktionalitäten für das Lesen und Schreiben von STEP-Dateien, die auf den APs 203, 214 und 209 basieren. Eine andere Alternative steht mit der vom Unternehmen LKSoftWare GmbH⁹³ entwickelten Bibliothek JSDAI zur Verfügung, welche seit Ende des Jahres 1996 kontinuierlich weiterentwickelt wurde und auf der Internetseite <http://www.jsdai.net> erhältlich ist. Diese Bibliothek ist mit den beiden im STEP-Standard genormten Implementierungsmethoden ISO 10303-22 (Standard Data Access Interface, SDAI) und ISO 10303-27 (Java programming language binding to the SDAI) konform, was auch durch das Akronym JSDAI, welches für **S**tandard **D**ata **A**ccess **I**nterface for the **J**ava programming language steht, ausgedrückt wird. Obwohl zunächst nur kommerziell erhältlich, steht JSDAI seit Mitte 2008 auch als Open Source-Bibliothek zur Verfügung und kann unter der AGPL v3-Lizenz⁹⁴ zum Lesen, Schreiben und Manipulieren von objektorientierten, durch ein EXPRESS-basiertes Datenmodell definierte Daten eingesetzt werden. Dabei wird für die derzeit verfügbare Version v4 (aktuell Release v4.2.1) auch ein Plugin für die integrierte Entwicklungsumgebung Eclipse⁹⁵ angeboten.

IGES Da die Entwicklungsplattform Open CASCADE Technology auch Funktionalitäten zum Lesen und Schreiben von IGES-Modellen enthält, kann für die Verarbeitung dieser Daten ebenfalls die für STEP einsetzbare jCAE-Bibliothek genutzt werden (siehe Abschnitt STEP).

Repräsentation von Metadaten

Die Analyse der in CAD-Modellen enthaltenen Metadaten hat gezeigt, dass in dieser Gruppe eine Vielzahl unterschiedlichster Informationen enthalten sein kann. Um einen besseren Überblick zu geben, verdeutlicht Tabelle 5.2 eine Einordnung dieser Informationen in verschiedene Kategorien. Daraus ist ersichtlich, dass es sich bei Metadaten nicht immer nur um dokumentbeschreibende Informationen, sondern auch um Informationen zu dem im Dokument beschriebenen Produkt sowie zu Personen, Werkstoffen oder zusätzlichen Dokumenten (z. B. Richtlinien) handelt. Demzufolge sollten zusätzlich zum eigentlichen Dokument zunächst diejenigen Artefakte, welche in den Metadaten adressiert werden und im Retrievalprozess auffindbar sein sollen, als

⁹¹<http://www.opencascade.org>

⁹²<http://www.opencascade.com>

⁹³<http://www.lksoft.com>

⁹⁴GNU Affero General Public License

⁹⁵<http://www.eclipse.org>

Tabelle 5.2. — Einordnung der potentiell verfügbaren Metadaten in Kategorien

Kategorie	Metadaten
<i>Dokumentinformationen</i>	Dateiname bzw. Pfad
	Erstellungs- & Änderungsdatum
	Dateigröße
	Dokumentversion
	Quell-CAD-System & Release
	Version bzw. AP der Formatspezifikation
<i>Produktbezogene Informationen</i>	Produktname
	Sachnummer bzw. ProduktID
	Produktgranularität
	Produktfunktion (optional)
	Modellmaßstab & Maßeinheit
	Produktversion & Freigabestatus
	Angaben zur Fertigung (Oberflächen- güte, Schweißvermerke, Toleranzen, . . .)
	Sonstige textuelle Anmerkungen
<i>Personenbezogene Informationen</i>	Dokumentautor
	Lieferantendaten
	Unternehmensdaten
<i>Referenzierende Daten</i>	Dokumentverweise (Richtlinien, Zeichnungen, Auftrag, . . .)
<i>Werkstoffdaten</i>	Farbdefinition
	Werkstoffnummer (optional) & -name

eigenständige **IndexableArtifact**-Objekte erstellt und mit den jeweils zugehörigen Beziehungsfacetten versehen werden. Anschließend können diese Artefakte durch die in den Metadaten verfügbaren Informationen bzw. deren Repräsentationen ergänzt werden. Dabei sind prinzipiell alle Metadaten in Form beschreibender Eigenschaften, d. h. als Name-Wert-Paare, verfügbar. Damit können sie zur Generierung von Attributfacetten genutzt werden, indem der Wert eines Name-Wert-Paares als Ausprägung der, zum Namen korrespondierenden Attributfacette gesetzt wird. Ein Problem hier-

bei ist allerdings, dass in den Unternehmen die Möglichkeiten zur Speicherung von Metadaten häufig nicht wirklich genutzt bzw. nur die notwendigen Standardwerte gesetzt werden (vgl. Abbildung 5.13). Folglich liegen hier Potentiale frei, die in Anspruch genommen werden sollten, um Informationen für hilfreiche Suchkriterien auf relativ einfache Art und Weise automatisch erfassen zu können.

```
ISO-10303-21;
HEADER;
FILE_DESCRIPTION(('CATIA V5 STEP Exchange'),'2;1');
FILE_NAME('\\\\10.11.131.4\\transit\\07-05-04\\hein\\Step-daten-fuer H.
G\\X2\\00FC\\X0\\ck\\Part1.stp','2007-05-04T13:13:55+00:00','(none)','(none)','CATIA Version
5 Release 16 (IN-10)';CATIA V5 STEP AP214','none');
FILE_SCHEMA(('AUTOMOTIVE DESIGN { 1 0 10303 214 1 1 1 }'));
ENDSEC;
DATA;
#5=PRODUCT('Part1','',( #2 ));
#2=PRODUCT_CONTEXT('','#1,'mechanical');
#1=APPLICATION_CONTEXT('automotive design');
#10=PRODUCT_DEFINITION('','', #6, #3);
#3=PRODUCT_DEFINITION_CONTEXT('part definition', #1, '');
#11=PRODUCT_DEFINITION_SHAPE('','', #10);
#42328=MECHANICAL_DESIGN_GEOMETRIC_PRESENTATION_REPRESENTATION
('',( #30 ), #17);
#20=SHAPE_REPRESENTATION('',( #19 ), #17);
#19=AXIS2_PLACEMENT_3D('','#18,$,$);
#36=AXIS2_PLACEMENT_3D('Plane Axis2P3D', #33, #34, #35);
#252=AXIS2_PLACEMENT_3D('Ellipse Axis2P3D', #249, #250, #251);
#267=AXIS2_PLACEMENT_3D('Ellipse Axis2P3D', #264, #265, #266);
...
ENDSEC;
END-ISO-10303-21;
```

Abbildung 5.13. — Beispiel einer STEP-Datei eines Deckels für einen Motorkanal, deren Metadaten nur folgende Standardinformationen enthalten: Pfad inkl. Dateiname (blau), Erstellungszeitpunkt (rot), Quell-CAD-System (grün), Verwendetes AP (orange), Produktname (lila), Produktgranularität (türkis)

Repräsentation von Form-Feature-Informationen

Nachdem v. a. proprietäre Datenformate, aber auch manche APs des STEP-Standards über die Möglichkeit zur Speicherung von Informationen über Form-Features verfügen, sollen im Folgenden denkbare Optionen zur Repräsentation dieser Informationen aufgezeigt werden. Dabei wird die Annahme zugrunde gelegt, dass diese in Form sogenannter Feature-Bäume oder auch Feature-Historien vorliegen, welche die Reihenfolge der zur Erstellung der Produktgeometrie verwendeten Form-Features (Boh-

rungen, Fasen, ...) dokumentieren. Sind diese Feature-Historien verfügbar, so können sie grundsätzlich für einen Ähnlichkeitsvergleich von Produkten verwendet werden. Dazu ist das jeweilige Produkt bzw. `IndexableProduct`-Objekt lediglich um eine Ähnlichkeitsfacette zu ergänzen, die als Ausprägung eine geeignete Beschreibung dieser Informationen enthält. Wie eine derartige Beschreibung aussehen kann, wurde bereits in Kapitel 2.5.1 anhand von Forschungsansätzen in diesem Bereich erläutert. Allerdings ist zu berücksichtigen, dass ein Ähnlichkeitsvergleich von Feature-Historien nicht notwendigerweise alle zu einem Produkt ähnlichen Produkte findet. Der Grund hierfür liegt in der Varietät der Reihenfolge, in der Form-Features zur Erzeugung eines Produktes angewendet werden können.

Neben einer Verwendung zur Bildung einer Ähnlichkeitsfacette, können Feature-Informationen aber auch hilfreich sein, um Produkte mit bestimmten Form-Features zu finden. So wurde in Gesprächen im Rahmen dieser Arbeit beispielsweise das Informationsbedürfnis geäußert, Kabelkanäle anhand der Anzahl ihrer Befestigungsbohrungen finden zu wollen. Indem man diese Information aus den vorhandenen Feature-Historien extrahiert, kann folglich eine Attributfacette gebildet werden, die eine Selektion von Produkten erlaubt, welche entweder über eine exakte Anzahl oder eine Mindest- bzw. Maximalmenge von Bohrungen verfügen. Dieses Beispiel zeigt aber auch, dass mit der Bildung einer derartigen Facette wiederum eine neue Fragestellung verbunden ist. Schließlich muss nicht jede Bohrung eines Produktes notwendigerweise eine Befestigungsbohrung sein, so dass hier entweder nur eine Selektion anhand der allgemein verfügbaren Bohrungen möglich ist oder gegebenenfalls weitere Kriterien zur Identifikation von Befestigungsbohrungen herangezogen werden müssen.

Repräsentation der Produktstruktur

Die Struktur eines Produktes lässt sich üblicherweise in Form einer hierarchischen Anordnung einzelner Elemente extrahieren, wie dies im Beispiel in Abbildung 5.14 links anhand eines JT-Modells verdeutlicht wird. Dabei beschreibt jeder Knoten eine Komponente des Produktes, die selbst wieder ein Produkt ist. Folglich sollten für diese Komponenten ebenfalls eigenständige Artefakte erzeugt werden, die mit den aus der Produktstruktur extrahierbaren Beziehungsfacetten versehen werden (vgl. hierzu die in Kapitel 4.2 erläuterten 'besteht aus'- und 'ist enthalten in'-Beziehungen). Zusätzlich kann anhand dieser Produktstruktur für jede einzelne Baugruppe bestimmt werden, aus wie vielen Einzelteilen bzw. Baugruppen diese selbst zusammengesetzt ist. Somit lassen sich für jede Baugruppe zusätzlich die zwei Attributfacetten *Anzahl-BG* und *Anzahl-ET* bilden.

Des Weiteren bietet die Produktstruktur die Möglichkeit zur Generierung einer Ähnlichkeitsfacette, welche eine Suche nach ähnlich strukturierten Produkten ermöglicht. Hierfür muss die Produktstruktur in eine geeignete Beschreibung überführt werden, die einem Ähnlichkeitsvergleich unterzogen werden kann. Allerdings muss dabei zu-

<pre> Level: 0 JtkASSEMBLY: B120516173_D-STROMGENERATOR (32 children) JtkTRANSFORM 1, 0, 0, 0 0, 1, 0, 0 0, 0, 1, 0 0, 0, 0, 1 Level: 1 JtkASSEMBLY: C120940958_002-SPREIZBUCHSE (1 children) JtkTRANSFORM 0.866025, -0.5, 0, 0 0.5, 0.866025, -0, 0 0, 0, 1, 0 -81.5, -45, -147.6, 1 Level: 2 JtkPART: C120940958_002-SPREIZBUCHSE_SOLIDS JtkTRANSFORM 1, -2.98023e-008, -1.06226e-009, 0 2.98023e-008, 1, -2.37518e-008, 0 1.06226e-009, 2.37518e-008, 1, 0 -3.95208e-007, -3.46726e-007, 0, 1 JtkMATERIAL ambient = (0.0539986, 0.107997, 0.0279973, 1) diffuse = (0.269993, 0.539986, 0.139986, 1) specular = (0.8, 0.8, 0.8, 1) emission = (0, 0, 0, 1) shininess = 15 Level: 1 JtkASSEMBLY: I123421002_002-SENKSSCHRAUBE (1 children) JtkTRANSFORM 1, 0, 0, 0 0, 1, 0, 0 0, 0, 1, 0 27.0459, -7.24693, -1.03502, 1 ... </pre>	<pre> JtkXTBody { 145 Generate texture coordinates = 0 Write nominal geometry = 0 Number of regions = 2 JtkXTRegion { 146 Solidity = 0 Number of shells = 1 JtkXTShell { 148 Number of facades = 4 Number of edges = 0 JtkXTFacade { 188 Sense = 1 JtkXTFace { 149 Ambient = [0.2000000029802322, 0.2000000029802322, 0.2000000029802322, 1] Diffuse = [0.6000000238418579, 0.6000000238418579, 0.6000000238418579, 1] Specular = [0.75, 0.75, 0.75, 1] Emission = [0, 0, 0, 1] Number of loops = 2 JtkXTCylinder { 150 Location = (0, 0, 1) Reference = (-1, 0, 0) Axis = (0, 0, -1) Radius = 4.4 } JtkXTLoop { 158 Number of fins = 1 JtkXTFin { 160 Edge direction = 1 Interval = [0, 0] JtkXTEdge { 161 Precision = 1e-005 } } } } } } } } ... </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Abbildung 5.14. — Exemplarische Daten eines CAD-Modells zu dessen Produktstruktur (links) bzw. BRep-basierter Produktgestalt (rechts) im JT-Format

nächst geklärt werden, wann zwei Produktstrukturen als ähnlich zueinander betrachtet werden. Gemäß Tsai, Tien und Pan sind dabei grundsätzlich zwei Ähnlichkeiten von Bedeutung [TTP04]. Zum einen ist dies die als *configuration similarity* bezeichnete strukturelle Ähnlichkeit, die sich rein auf den strukturellen Aufbau eines Produktes aus Unterkomponenten bezieht. Zu ihrer Ermittlung sind zunächst die Komponenten jeder einzelnen Strukturebene auf eine Übereinstimmung hinsichtlich ihrer Anzahl miteinander zu vergleichen, wobei zu beachten ist, dass das Fehlen einer Komponente in einer höheren Strukturebene bedeutender ist als in einer niedrigeren Ebene. Folglich ist, wie im oberen Bereich von Abbildung 5.15 dargestellt, für ein Anfrageprodukt Q_2 die Produktstruktur des Produktes D_3 mit einer höheren Ähnlichkeit zu bewerten als die des Produktes D_4 . Zum anderen können aber auch bei struktureller Gleichheit auf einer Ebene der Produktstruktur Unterschiede in den Produkten bzw. Komponenten

selbst existieren, indem beispielsweise verschiedene Varianten (vgl. unterer Bereich in Abbildung 5.15) oder auch verschiedene Produkte eingesetzt werden. Aus diesem Grund sollte zur Berechnung der Gesamtähnlichkeit neben einer rein strukturellen Ähnlichkeit auch immer die von Tsai, Tien und Pan als *mode similarity* bezeichnete Typähnlichkeit zwischen den einzelnen Komponenten mit einbezogen werden [TTP04].

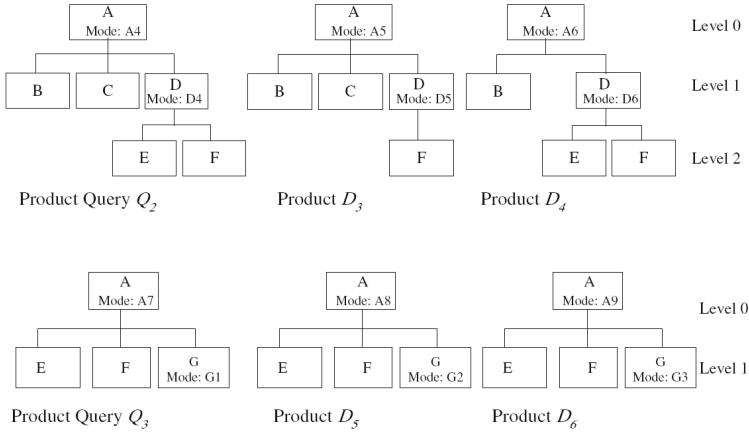


Abbildung 5.15. — Vergleich von Produktstrukturen D_i mit der Struktur eines Anfrageproduktes Q_i (entnommen aus [TTP04, S. 1509])

Um dies zu realisieren schlagen Tsai, Tien und Pan eine tiefenbasierte Überführung der Produktstruktur in eine zweidimensionale Matrix vor, aus der unter Verwendung beider Ähnlichkeiten die Gesamtähnlichkeit einer Produktstruktur ermittelt werden kann. Dabei wird für jede Ebene der Produktstruktur eine Spalte in der Matrix erzeugt, so dass jede Zeile schließlich eine Zeichenkette enthält, welche die für einen Pfad in der Produktstruktur sequentielle Anordnung der Komponenten von der Wurzel bis zum Blatt repräsentiert. Auf diese Weise wird beispielsweise das in Abbildung 5.15 dargestellte Produkt Q_2 durch eine dreispaltige Matrix mit den vier Reihen AB, AC, ADE und ADF beschrieben. Zusätzlich ist für jede Komponente ihr Komponententyp zu speichern. Der Vergleich zweier Matrizen QM und DM erfolgt schließlich in einem dreistufigen Vorgehen, indem zunächst ein Ähnlichkeitsvergleich der Matrizenreihen erfolgt. Dabei wird für jede Reihe bzw. jede Zeichenkette r_j aus QM und jede Zeichenkette r_k aus DM ein Ähnlichkeitswert $SV_{j,k}$ ermittelt. Dieser nimmt den Wert 1 an, wenn die Zeichenketten und zugehörigen Komponententypen vollkommen über-

einstimmen. Bestehen die Zeichenketten aus völlig unterschiedlichen Zeichen, so ist $SV_j, k = 0$. In allen anderen Fällen ermittelt sich der Wert als

$$SV_j, k = \frac{\sum_{i=1}^{count} (((1 - Position_i - 1) / AdjEditDist_{j,k}) * ModeSim_{i,j})}{Length(QM[j])}$$

mit

- $AdjEditDist_{j,k}$ ist die bereinigte Edit-Distanz zwischen den Zeichenketten r_j und r_k ,
- $Position_i$ ist die relative Position der übereinstimmenden Komponente i ,
- $count$ ist die Gesamtzahl der übereinstimmenden Komponenten der Zeichenketten r_j und r_k ,
- $Length(QM[j])$ ist die Länge der Zeichenkette r_j aus der Matrix QM und
- $ModeSim_{i,j}$ ist die Typähnlichkeit zwischen der Komponente i der Zeichenkette r_j und der übereinstimmenden Komponente in Zeichenkette r_k , die mit Hilfe eines fuzzy-basierten Moduls (vgl. [TTP04]) ermittelt wird.

Auf dieser Basis lässt sich nun für jede Zeichenkette r_j aus QM ein temporärer Ähnlichkeitswert TSV_j als Maximum der Ähnlichkeitswerte SV_j, k und schließlich ein struktureller Ähnlichkeitswert SSV als normalisierte Summe aller temporären Ähnlichkeitswerte TSV_j ermitteln [TTP04].

Speichert man folglich für jede Produktstruktur deren 2D-Matrix und die zugehörigen Komponententypen als Ausprägungen einer Ähnlichkeitsfacette, können diese Informationen im Rahmen des Retrievalprozesses zur Berechnung der Ähnlichkeitswerte genutzt werden. Dazu muss lediglich das beschriebene Vorgehen zur Ermittlung der Ähnlichkeiten als Ähnlichkeitsmaß für das entsprechende Suchmodul der Suchkomponente zur Verfügung gestellt werden.⁹⁶

Ferner sollte bei einer Indexierung bzw. Informationsrepräsentation nicht nur an den Suchprozess, sondern auch an den Aspekt der Ergebnispräsentation gedacht werden. Will man beispielsweise dem Nutzer auf der GUI eine Erklärungskomponente anbieten, die aufzeigt, warum ein Produkt als relevant für eine Ähnlichkeitsanfrage eingestuft wurde, könnte man ihm hierzu ein Bild der zu diesem Produkt gehörigen Produktstruktur anzeigen. Dafür muss allerdings dieses Bild während der Indexierung erzeugt und schließlich mit gespeichert werden. Aus diesem Grund kann jedes **IndexableArtifact**-Objekt, also jede Polyrepräsentation, um weitere Facetten — die sogenannten Thumbnailfacetten — ergänzt werden, welche generell Bilddaten als Ausprägungen besitzen.

⁹⁶Für die Implementierung solcher Ähnlichkeitsmaße stellt das LFRP-Framework das Interface **SimilarityMeasure** zur Verfügung.

Repräsentation der Produktgeometrie

Abhängig von der Form, in der die Geometriedaten aus einem CAD-Modell extrahiert werden können, lässt sich das im Dokument beschriebene Produkt um unterschiedliche Repräsentationen erweitern. Sind beispielsweise BRep-Daten verfügbar, können unter Umständen produktspezifische Eigenschaften hieraus ermittelt werden. Dies soll am Beispiel einer JT-Datei zur Beschreibung eines O-Rings verdeutlicht werden. Wie Abbildung 5.14 rechts verdeutlicht, sind in den Parametern bestimmter Elemente — hier einem *JtkXTCylinder*-Element und seinem *Radius*-Attribut — spezifische Abmessungen für O-Ringe enthalten. Extrahiert und analysiert man diese Elemente, können Werte für den Außenradius und den Innenradius eines O-Rings bestimmt werden. Zusätzlich besteht bei manchen Produkten, wie beispielsweise bei O-Ringen oder auch Scheiben, die Möglichkeit, durch Nutzung der das Produkt umgebenden Bounding Box deren Höhe bzw. Dicke zu bestimmen. Auf diese Weise können somit entsprechende Attributfacetten gebildet und dem jeweiligen Produkt als Repräsentationen hinzugefügt werden. Allerdings ist natürlich immer zu berücksichtigen, dass diese Parameterermittlung nur produktabhängig erfolgen kann und folglich eine spezifische Ausgestaltung der Extraktorkomponente erfordert.

Anstelle von BRep-basierten Daten werden geometrische Informationen jedoch viel häufiger in Form tessellierter Daten gespeichert. Diese wiederum können einerseits dazu genutzt werden, Vorschaubilder für die Produkte zu erzeugen, die als Ausprägungen von Thumbnailfacetten dem Produkt zugefügt und schließlich bei der Präsentation der Ergebnisse auf der GUI angezeigt werden können. Des Weiteren ermöglichen tessellierte Daten einen Vergleich von Produkten anhand der geometrischen Gestalt, weshalb sie in jedem Fall zur Bildung einer Ähnlichkeitsfacette genutzt werden sollten. Da es sich bei der Geometrie um eine typische Produkteigenschaft handelt, ist diese prinzipiell als produktspezifische Facette zu indexieren. Zusätzlich kann diese Ähnlichkeitsfacette aber auch als eine dokumentspezifische Facette betrachtet werden, weil ein derartiger Geometrievergleich nur bei Dokumenten vom Typ CAD-Modell anwendbar ist. Folglich sollten beide Artefakte — Produkt und Dokument — mit dieser Ähnlichkeitsfacette versehen werden. Hierzu sind die Daten jedoch zunächst aufzubereiten: Eine Analyse von Testdaten hat gezeigt, dass triangulierte Daten prinzipiell in Form von Punktmengen gegeben sind, aus denen die einzelnen Dreiecke nach bestimmten Regeln gebildet werden können. Dabei nutzt allerdings jedes Austauschformat andere Regeln, was entsprechende formatspezifische Methoden zum Erhalt dieser Dreiecksdaten erfordert. So erfolgt beispielsweise in einer VRML-Datei die Definition von Dreiecken und ihren zugehörigen Punkten gemäß Abbildung 5.16 links. Dabei werden zunächst mit Hilfe eines `point[...]`-Elementes alle 3D-Punkte des Polygonnetzes definiert, wobei jede Zeile einen Punkt durch die drei Koordinaten *x*, *y* und *z* sowie eine Positionsnummer (beginnend bei 0) repräsentiert. Diese Punkte werden anschließend in einem `coordIndex[...]`-Element zu Dreiecken kombiniert, wobei die Zahl `-1` zur Trennung der einzelnen Dreiecke dient. Die Dreiecke selbst setzen sich

aus den zu den drei, vor dem Trennzeichen angegebenen Positionsnummern gehörigen Punkten zusammen. Demzufolge definiert die Zeile 0,1,2,-1 in Abbildung 5.16 links ein Dreieck aus den Punkten (16.9037, 0, -4.5293), (17.5, 0, 0) und (16.9037, 0, 4.5293). Eine JT-Datei hingegen beschreibt die Produktgestalt in verschiedenen Detaillierungsstufen, die als Level of Detail (LoD) bezeichnet werden⁹⁷ und jeweils zugehörige geometrische Teilmengen, die sogenannten *geom set*-Elemente, enthalten (vgl. Abbildung 5.16 rechts oben). Dabei wird jede einzelne Teilmenge wiederum durch die im *vertices*-Element angegebenen Punkte charakterisiert. Diese sind aufgrund der Tatsache, dass ein Polygon aus n Punkten immer in $n - 2$ Dreiecke zerlegt werden kann, nach folgender Regel zu Dreiecken zu kombinieren:

Verbinde die ersten drei Punkte eines *geom set*-Elementes zu einem Dreieck. Enthält dieses Element mehr als drei Punkte, so bilde jedes weitere Dreieck aus den zwei letzten Punkten des vorhergehenden Dreiecks und dem nächsten, noch nicht in einem Dreieck enthaltenen Punkt.

Damit lassen sich beispielsweise aus dem in Abbildung 5.16 rechts unten dargestellten Polygon mit fünf Punkten P_1, P_2, P_3, P_4, P_5 drei Dreiecke mit den Punkten (P_1, P_2, P_3) , (P_2, P_3, P_4) und (P_3, P_4, P_5) erzeugen.

Zur Bildung einer für den Ähnlichkeitsvergleich zu verwendenden Repräsentation können nun in einem nächsten Schritt verschiedene Konzepte auf diese Dreiecksdaten angewendet werden. Nachdem 3D-Modelle nicht nur im Bereich der technischen PE, sondern auch für die Entwicklung von Computerspielen, die Produktion von Kinofilmen oder die Planung von Immobilien immer wichtiger werden, hat sich ein eigener Forschungsbereich entwickelt, der sich allein mit der Entwicklung und Optimierung derartiger Konzepte beschäftigt. Aus diesem Grund soll im Folgenden nur ein kurzer Überblick über diesen Bereich der 3D-Ähnlichkeitssuche — häufig auch als *3D Shape Retrieval* bezeichnet — gegeben werden. Für einen umfassenden Überblick wird auf die zahlreiche Literatur zu diesem Bereich verwiesen [CGK03, TV04, IJL⁺05, BKS⁺06]. Konkrete Konzepte, die für eine Anwendung zum Vergleich technischer Produkte in Frage kommen und im Rahmen der prototypischen Implementierung des Frameworks umgesetzt und evaluiert wurden, werden im Anschluss in Kapitel 5.4.2.5 zusammen mit den Ergebnissen dieser Auswertung vorgestellt.

Die Erstellung einer Repräsentation für ein 3D-Modell lässt sich allgemein als mehrstufiges Vorgehen beschreiben, in dem mindestens zwei der in Abbildung 5.17 dargestellten vier Schritte auszuführen sind [BKS⁺06]. Dabei ist ein 3D-Modell häufig zunächst einem Vorverarbeitungsschritt zu unterziehen, in dem eine Normalisierung des Modells hinsichtlich seiner Positionierung im Koordinatensystem vorgenommen wird. Dies ist notwendig, um trotz unterschiedlicher Lagen oder Skalierungen der Objekte Ähnlichkeiten zwischen ihnen feststellen zu können. Hierfür wird eine Haupt-

⁹⁷Der LoD mit dem Wert '0' beschreibt dabei das Produkt mit der höchsten Genauigkeit.

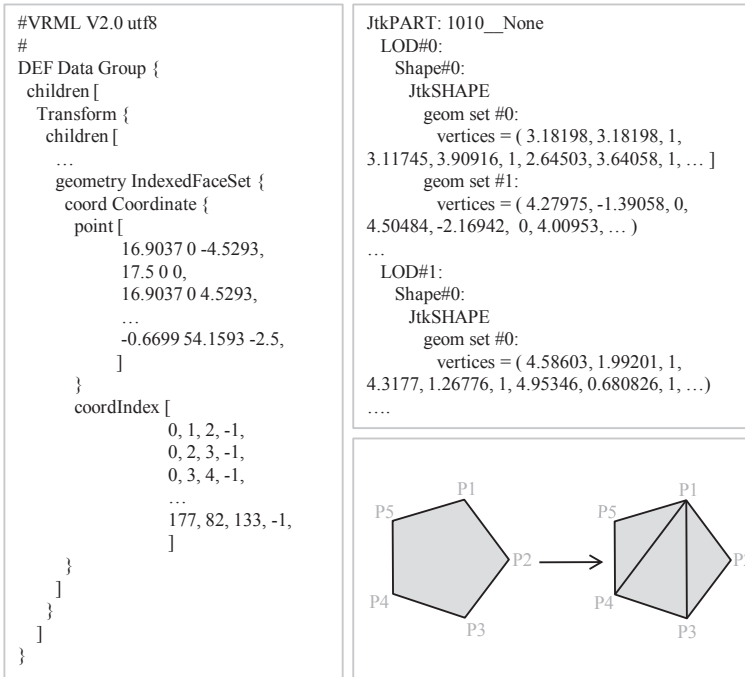


Abbildung 5.16. — Definition der triangulierten Geometriedaten im VRML-Format (links) und im JT-Format (rechts oben) sowie Darstellung der Zerlegung eines Polygons aus fünf Punkten in drei Dreiecke (rechts unten)

komponentenanalyse — in der Literatur auch als Principal Component Analysis oder Karhunen-Loeve-Transformation bezeichnet — auf das 3D-Modell angewendet, um eine Invarianz der Daten gegenüber verschiedenen Transformationen (z. B. Translation, Skalierung, Rotation) zu gewährleisten [LPR04, BKS⁺06]. Im Anschluss sind die tesselierten Geometriedaten zu abstrahieren, so dass diese entweder in Form einer dreidimensionalen Objektoberfläche, eines dreidimensionalen Volumens oder als eines bzw. mehrere zweidimensionale Bilder weiterverarbeitet werden können. Für diese Objektabstraktion werden schließlich im dritten Schritt unter Nutzung verschiedener Methoden, wie beispielsweise einem Sampling oder einer Diskreten Fouriertransformation (DFT), charakteristische Merkmale (im IR als Features bezeichnet) erzeugt, aus denen schließlich eine vektorbasierte, eine histogrammbasierte oder eine graphbasierte Modellrepräsentation erzeugt wird. Wie Bustos et al. in einem umfassenden

Überblick zeigen [BKS⁺05], sind dabei nicht immer alle beschriebenen Prozessschritte erforderlich bzw. können diverse Methoden zu deren Durchführung verwendet werden. Folglich weisen die in der Forschung entwickelten Konzepte zur Repräsentationserstellung eine große Varietät auf, so dass diese in der vorliegenden Arbeit anhand der für eine Repräsentation verwendeten Ausprägung in drei Kategorien eingeteilt werden.

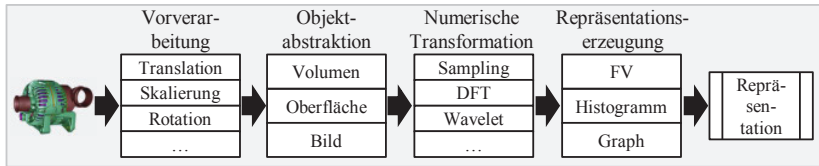


Abbildung 5.17. — Mehrstufiger Prozess der Repräsentationserzeugung aus einem 3D-Modell (nach [BKS⁺06, S. 40])

Die erste Kategorie umfasst dabei die vektorbasierten Ähnlichkeitskonzepte, welche als Repräsentation einen n -dimensionalen Vektor — einen sogenannten Featurevektor (FV) — erzeugen. Dieser enthält die Features der Objektgestalt als numerische Werte, auf die sich relativ einfach mathematische Funktionen wie Distanz- oder Ähnlichkeitsfunktionen anwenden lassen. Allerdings können hierbei verschiedene Merkmale betrachtet werden, weshalb zwischen geometriebasierten und bildbasierten Features zu differenzieren ist. Wie die Bezeichnung vermuten lässt, beziehen sich geometriebasierte Features auf die Geometrie des 3D-Modells, wobei diese entweder global oder lokal fokussiert sind [TV04]. Globale Geometriefeatures betrachten dabei die gesamte Gestalt des 3D-Modells und beschreiben diese mit Hilfe von statistischen Momenten des Umfangs oder Volumens [SV01b, NK04], Verhältniszahlen (z. B. Oberfläche zu Volumen-Verhältnis) [Kei99, CRC⁺02], Fourier-Koeffizienten des Volumens [SV01a] oder Kugelflächenfunktionen [SV01b, KFR03]. Lokale Geometrieigenschaften hingegen konzentrieren sich auf spezielle Aspekte oder Bereiche der Objektform wie beispielsweise das Krümmungsverhalten an einzelnen Punkten der Objektoberfläche [SHI95] oder einzelne Flächensegmente [KSS97]. Werden bildbasierte Features zur Erzeugung eines FV verwendet, erfolgt eine Abstraktion des 3D-Modells durch eines oder mehrere Bilder. Damit lässt sich die Problematik der Objektbeschreibung in den 2D-Raum zurückführen und folglich eine Komplexitätsreduktion erreichen. Abhängig von der Art der Erzeugung dieser Bilder bzw. deren Informationsgehalt existieren diverse Ansätze, zu denen u. a. die von Vranic entwickelten Konzepte auf Basis von Kontur- und Tiefeninformationen [Vra04], der aus unterschiedlichen Blickwinkeln erzeugte Lightfield Descriptor [CTSO03] sowie das von Li und Johan entwickelte Konzept zur Bildung eines sogenannten View Context [LJ10] gehören. Neben dieser Vielzahl an einzel-fokussierten Ansätzen wird zudem an der Verbindung meh-

rerer FV zu einer Repräsentation geforscht, um die Retrievaleffektivität zu erhöhen [BKS⁺04, AKWK06].

Die Grundidee der zweiten Konzeptkategorie besteht in der Ermittlung von Histogrammen, welche die Vorkommenshäufigkeit von gestaltspezifischen Parametern als Verteilungsfunktion abbilden und eine Ähnlichkeitsermittlung von Objekten durch Vergleich der einzelnen Histogrammklassen ermöglichen. Die Parameter, die hierzu verwendet werden, beziehen sich dabei hauptsächlich auf lokale Merkmale der Objektgestalt. So betrachten beispielsweise Kriegel et al. die Lage von Punkten bzw. deren Zugehörigkeit zu einem Bereich des das Objekt umgebenden Raumes, wobei unterschiedliche Methoden der Raumunterteilung (Volumen-, Sektoraufteilung, ...) zum Einsatz kommen [AKKS99, KKM⁺03]. Andere wiederum berechnen die Abstände zwischen einzelnen Punkten der Objektoberfläche [OFCD01, KPNK03, OMT03], ermitteln die Verteilung von Normalenvektoren an zufällig verteilten Punkten auf der Oberfläche [KGS09] oder bestimmen Korrelationswerte zwischen der an einzelnen Punkten vorliegenden Krümmung [ABDP05].

Die dritte Kategorie von Ähnlichkeitskonzepten umfasst schließlich alle nicht-numerischen Ansätze, wobei es sich hierbei primär um graphbasierte Ansätze handelt. Diese legen ihren Fokus zwar auf eine topologische Beschreibung des 3D-Modells in Form eines Graphen, um die räumliche Struktur⁹⁸ des Objektes zu veranschaulichen; erfordern jedoch für eine Extraktion und einen Vergleich relativ komplexe Algorithmen, die v. a. bei großen Datenmengen zu Effizienzproblemen führen [BKS⁺05]. Im Gegenzug ist bei diesen Konzepten kein Vorverarbeitungsschritt zur Normalisierung der Objekte notwendig. Weiterhin bieten manche graphbasierten Ansätze die Möglichkeit, Objekte auf mehreren Detaillierungsstufen zu beschreiben, was v. a. für ein partielles Matching genutzt werden könnte [IJL⁺05]. Insgesamt betrachtet unterscheiden sich die in der Forschung hierzu existierenden Konzepte primär anhand der Art des Graphen, der zur Beschreibung erzeugt wird. Neben Modellgraphen wie dem Model Signature Graph [MPSR01] oder dem Form Feature Adjacency Graph [CH06], kommen hierbei insbesondere skelettartige Graphen wie Skeletons [SSGD03, LPR04, ATC⁺08] oder Reeb-Graphen [HSKK01, BRS03] zum Einsatz.

Wie in Kapitel 2.5.1 erläutert, können 3D-Modelle nicht nur mit ihresgleichen, sondern auch mit 2D-Modellen (z. B. in Form von Skizzen) verglichen werden. Folglich sollten aus den Geometriedaten eines CAD-Modells zwei Ähnlichkeitsfacetten — eine für den 3D / 3D- und eine für den 2D / 3D-Vergleich — gebildet werden. Da jedoch das Anfrageobjekt und das zu vergleichende Objekt immer in der gleichen Repräsentationsform vorliegen müssen, kommen hier schließlich zwei Vorgehensweisen für einen 2D / 3D-Vergleich in Frage. Zum einen kann eine Rückführung des 3D-Modells in den 2D-Raum vorgenommen werden, wie dies bei den bildbasierten Ansätzen erfolgt. Zum

⁹⁸Im Gegensatz zur Produktstruktur beschreibt die räumliche Struktur die Anordnung von geometrischen Elementen (Flächen, Punkte, ...) der Objektgestalt zueinander.

anderen besteht aber auch die Möglichkeit, aus der Anfrage — welche aus einem oder mehreren Skizzen bzw. 2D-Bildern besteht — ein 3D-Modell zu rekonstruieren und für dieses eine entsprechende Repräsentation zu erzeugen. Ansätze hierfür werden in [LS96, CLT06, PR06] vorgestellt.

Für die Implementierung derartiger Ähnlichkeitskonzepte sowie ihrer zugehörigen Ähnlichkeits- bzw. Distanzfunktionen stellt die Indexierungskomponente die beiden Interfaceklassen `SimilarityConcept` und `SimilarityMeasure` zur Verfügung, die jederzeit um Subklassen erweitert werden können. Dabei sollte allerdings bei der Implementierung eines Ähnlichkeitskonzeptes berücksichtigt werden, dass bestimmte, während des Prozesses erzeugte Daten (z. B. Bilder, Histogramme) durchaus hilfreiche Informationen zur Erklärung der Relevanz eines Ergebnisses darstellen. Nutzt man diese zusätzlich als Ausprägungen entsprechend gekennzeichnete Thumbnailfacetten, können diese bei der Ergebnispräsentation für eine Relevanzklärung hilfreich sein.

5.4.2.5. Evaluation implementierter Ansätze für einen 3D / 3D-Vergleich

Aufgrund der Vielzahl bereits existierender 3D-Modelle in Unternehmen wird für die folgende Evaluation von der Annahme ausgegangen, dass ein 3D / 3D-Vergleich für viele Unternehmen den interessanteren und v. a. auch einfacher umzusetzenden Anwendungsfall darstellt. Demzufolge werden im Rahmen der vorliegenden Arbeit Ansätze zur Repräsentationserstellung für dieses Szenario vorgeschlagen. Allerdings ist hierbei zu berücksichtigen, dass in der Forschung bereits eine große Zahl an Konzepten verfügbar ist, so dass es nicht Ziel der vorliegenden Arbeit ist, ein neues, weiteres Konzept für den Vergleich zweier 3D-Modelle zu entwickeln. Vielmehr wird eine Empfehlung gegeben, welches Konzept sich für eine Verwendung im vorliegenden Anwendungsszenario eignet. Aus diesem Grund wurden auf Basis von in der Literatur verfügbaren Evaluationsbetrachtungen, die u. a. im Rahmen des vom europäischen Network of Excellence AIM@SHAPE⁹⁹ erstmals im Jahr 2006 organisierten und seitdem jährlich stattfindenden Evaluationsworkshop SHREC¹⁰⁰ entstanden sind, insgesamt sechs Konzepte ausgewählt, prototypisch implementiert und schließlich anhand von 250 industriellen Testdaten evaluiert. Hierbei handelt es sich um insgesamt vier FV-basierte und zwei histogrammbasierte Konzepte, die in den Tabellen 5.3 und 5.4 überblicksartig zusammengefasst sind. Graphbasierte Konzepte wurden aufgrund der mit ihnen verbundenen Komplexität und Effizienzprobleme bei großen Datenmengen, die auch für den Bereich der technischen PE typisch sind, nicht betrachtet.

⁹⁹<http://www.aimatshape.net>

¹⁰⁰SHREC steht als Akronym für 3D Shape Retrieval Contest und wurde in Anlehnung an die in der IR-Forschung bekannten Workshops TREC, MIREX, INEX usw. ins Leben gerufen, um eine einheitliche Basis für die Evaluation von 3D Shape Retrieval-Algorithmen zu schaffen. (<http://www.aimatshape.net/event/SHREC>)

Tabelle 5.3. — Übersicht über die prototypisch implementierten Konzepte

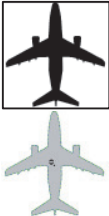
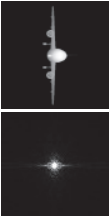


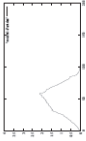


Konzept (Quelle)	Rep. (Dim.)	Basisprinzip	Verschaulichung
Silhouette-	FV	3 Silhouettebilder:	
based FV	3k	<ul style="list-style-type: none">- Parallelprojektion auf den minimal umgebenden Würfel- Bildung der Silhouette (= Kontur)- Anwendung der <i>Fast Fourier Transform</i> (FFT) auf die Konturpunkte- Verwendung der <i>k</i> niederfrequentesten Koeffizienten als Vektorkomponenten	
[Vra04]			
Depth	FV	6 Grauwertbilder:	
Buffer	6k	<ul style="list-style-type: none">- Parallelprojektion auf minimal umgebenden Würfel- Pixel kodiert Abstand zwischen Objekt und Projektionsfläche- Anwendung der <i>Discrete Fourier Transform</i> (DFT)- Verwendung der <i>k</i> niederfrequentesten Koeffizienten als Vektorkomponenten	
[Vra04]			
Ray-based	FV	Objektausdehnung:	
FV with	k(k + 1)/2	<ul style="list-style-type: none">- Erzeugung von Strahlen mit Ursprung im Objektschwerpunkt	
Spherical		<ul style="list-style-type: none">- Ermittlung der entferntesten Schnittpunkte zwischen Polygonnetz und Strahlen und deren Abstände vom Objektschwerpunkt- Definition einer reellwertigen Funktion auf der Einheitskugel, die die Abstände der Schnittpunkte zum Objektschwerpunkt als Objektausdehnung abbildet- Anwendung der <i>Spherical-FFT</i> auf die Funktionswerte- Verwendung der ersten <i>k</i> Koeffizientenreihen als Vektorkomponenten	
Harmonics			
[Vra04]			

Tabelle 5.4. — Übersicht über die prototypisch implementierten Konzepte — Fortsetzung

Konzept (Quelle)	Rep. (Dim.)	Basisprinzip	Veranschaulichung
Hybrid FV [Vra04]	FV	Kombination der drei in Tabelle 5.3 genannten vektorbasierten Konzepte	
Shape Distributions [OFCD01]	Histo- gramm b	Vorkommenshäufigkeit von Funktionswerten: - Ermittlung zufällig verteilter Oberflächenpunkte - Anwendung einer gestaltbeschreibenden <i>shape function</i> - Einordnung der Funktionswerte in Histogramm mit b Klassen	 
Shape Histogram [AKS99] [AKKP05]	Histo- gramm s	Vorkommenshäufigkeit von Punktzahlen - Bildung der umgebenden Kugel und Unterteilung in s Shells - Ermittlung zufällig verteilter Oberflächenpunkte und Zuordnung zu den Shells - Ermittlung der Punktzahl pro Shell & Einordnung in Histogrammklassen	 

Die Evaluation selbst erfolgte unter Rückgriff auf zwei, für den Bereich des IR gängige Methoden. Dabei handelt es sich zum einen um die Ermittlung des Mean Average Precision (MAP)-Wertes für ein Konzept bei Betrachtung der Top-k-Ergebnisse. Dieser gibt den Durchschnitt der über mehrere Suchanfragen ermittelten mittleren Precisionwerte an und lässt sich gemäß folgender Formel berechnen [MRS08]:

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk})$$

Dabei ist $|Q|$ die Anzahl aller durchgeführten Suchanfragen q_j , m_j die Zahl der relevanten Dokumente d_1, \dots, d_{m_j} für die Suchanfrage q_j und R_{jk} die Anzahl der Retrievalergebnisse bis zum relevanten Dokument d_k . In der vorliegenden Evaluation werden im Hinblick auf die Menge der betrachteten Top-k-Ergebnisse zwei Szenarios betrachtet, deren zugehörige MAP-Werte mittels der Bezeichnung MAP@k gekennzeichnet sind. Während im ersten Szenario nur eine Teilmenge des erzeugten Rankings — nämlich nur die ersten 10 Dokumente — betrachtet werden (MAP@10), bezieht das zweite Szenario die gesamte Testdatenmenge ein (MAP@250).

Zum anderen kann jedes Konzept anhand der mit ihm erreichbaren Recall- und Precisionwerte beurteilt werden. Dabei gibt ein Recallwert gemäß Baeza-Yates und Ribeiro-Neto das Verhältnis zwischen den im Ergebnis und den in der Kollektion enthaltenen relevanten Ergebnissen an [BR⁺99]. Somit charakterisiert dieser die Fähigkeit eines Retrievalverfahrens, überhaupt relevante Dokumente im Ergebnis zu liefern. Ein Precisionwert hingegen ist als das Verhältnis zwischen den relevanten Dokumenten im Ergebnis und den insgesamt gefundenen Dokumenten definiert [BR⁺99]. Damit beschreibt dieser, wie gut ein Retrievalverfahren in der Lage ist, nur relevante Dokumente im Ergebnis zu liefern. Demzufolge würde ein ideales IRS in beiden Fällen den Wert 1 aufweisen.

Evaluation der FV-basierten Konzepte

Für die Implementierung aller betrachteten FV-basierten Konzepte wurde auf ausführbare Programmdateien der Entwickler zurückgegriffen¹⁰¹, die den Prozess der Repräsentationserstellung vollständig durchlaufen und aufgrund von Evaluationsläufen der Entwickler über die jeweils optimalen Parameter verfügen [Vra04]. Diese Programme erfordern lediglich die Eingabe einer Datei im Object File Format (Dateiendung .off), was eine — jedoch problemlose — Konvertierung einer in einem Austauschformat vorliegenden Datei bedingt. Aus ihrer Anwendung resultiert schließlich eine .dfv-Datei, die den jeweiligen FV enthält. Dieser weist abhängig vom zugrunde liegenden Konzept sowie dem Parameter k der verwendeten Fourierkoeffizienten die im

¹⁰¹Die Programmdateien konnten ursprünglich über die Internetseite <http://www.informatik.uni-leipzig.de/~vranic/CCCC/index.old.html> bezogen werden und sind nun beim Entwickler anzufragen.

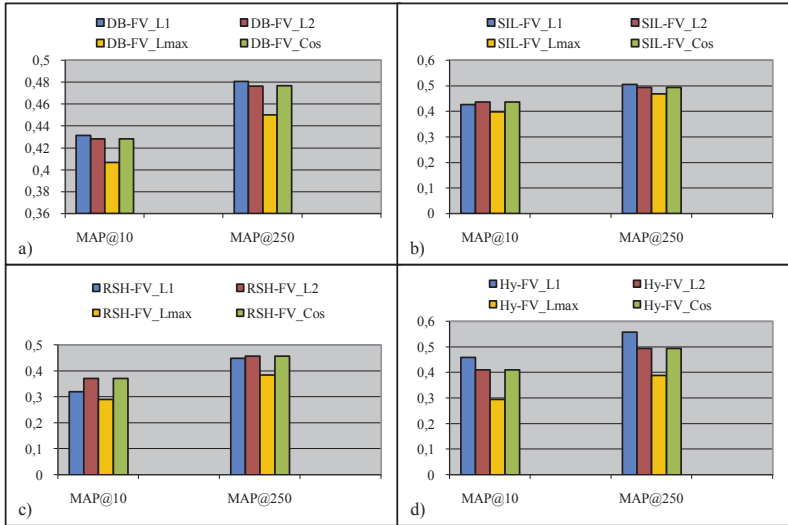


Abbildung 5.18. — Gegenüberstellung der MAP@10- und MAP@250-Werte für die FV-basierten Konzepte: a) DB-FV, b) SIL-FV, c) RSH-FV und d) Hy-FV

Folgenden angegebenen Dimensionen auf. Der Hybrid Feature Vector (Hy-FV) stellt dabei eine Kombination aus den drei anderen Konzepten Depth Buffer Feature Vector (DB-FV), Silhouette-based Feature Vector (SIL-FV) und Ray-based Feature Vector with Spherical Harmonics (RSH-FV) dar.

- DB-FV: Dimension = 438 $\Rightarrow k = 73$
- SIL-FV: Dimension = 300 $\Rightarrow k = 100$
- RSH-FV: Dimension = 136 $\Rightarrow k = 16$
- Hy-FV: Dimension = 472 $\Rightarrow k_{DB-FV} = 31, k_{SIL-FV} = 50, k_{RSH-FV} = 16$

Da durch die Verwendung der Programmdateien bereits die wichtigsten Stellschrauben der Konzepte festgelegt sind, erfolgt im Rahmen der vorliegenden Evaluation eine Betrachtung der für die Konzepte geeigneten Ähnlichkeitsfunktionen. Diesbezüglich werden die Minkowski-Distanzfunktionen Manhattan-Distanz (L1), Euklidische Distanz (L2) und Max- bzw. Tschebyscheff-Distanzfunktion (Lmax) sowie das Kosinusmaß (Cos) untersucht [Sch06]. Wie aus dem Vergleich der MAP-Werte in Abbildung 5.18 zu erkennen ist, sollten für die Konzepte DB-FV und Hy-FV die Manhattan-Distanz

und für das Konzept RSH-FV die euklidische Distanz oder das Kosinusmaß verwendet werden. Für das Konzept SIL-FV liefern die Manhattan-Distanz, die euklidische Distanz und das Kosinusmaß ähnliche Werte, so dass hier eine dieser drei Distanzen verwendet werden sollte. Diese Empfehlung bestätigt sich im Übrigen auch durch den ebenfalls für jedes Konzept durchgeführten Vergleich der ermittelten Recall-Precision-Diagramme, welche in den Abbildungen G.1 bis G.4 in Anhang G veranschaulicht sind.

Evaluation des Konzeptes Shape Histogram (SH)

Bei der Nachimplementierung des SH-Konzeptes sind verschiedene Stellschrauben zu berücksichtigen. Eine dieser Stellschrauben ist die Anzahl N der zufällig zu erzeugenden Oberflächenpunkte, welche in der durchgeführten Evaluation gemäß der Empfehlung der Autoren auf den Wert $N = 50000$ gesetzt wurde. Folglich wurden alle weiteren Stellschrauben auf Basis dieser Festlegung untersucht. Diese sind:

- die Anzahl s der die Kugel unterteilenden Shells
- die Methode, nach der die umgebende Kugel mit Radius r in Shells unterteilt wird:
 - Äquidistante Unterteilung (E), d. h. die Shells sind gleich weit voneinander entfernt, so dass der Abstand zwischen Innen- und Außenradius einer Shell immer dem Wert r/s entspricht.
 - Proportionale Unterteilung (P), d. h. die Kugel wird in Shells gleicher Bedeutung eingeteilt, so dass jede Shell durch die gleiche Anzahl an Punkten $a = \lfloor N/s \rfloor$ charakterisiert ist.
- die Methode, nach der die zufällig erzeugten Punkte den Shells zugeordnet werden:
 - Disjunkte Zuordnung (D), d. h. ein Punkt wird genau einer Shell zugeordnet.
 - Redundante Zuordnung (R), d. h. ein Punkt kann mehreren Shells zugeordnet werden und zwar jeder Shell, für die gilt: euklidischer Abstand zwischen Punkt und Mittelpunkt ist kleiner oder gleich dem Außenradius der Shell.
- die Methode zur Ermittlung der Histogrammwerte:
 - Volumenmodell (Vm), d. h. Ermittlung der Zahl an Punkten, die in einer Shell liegen. Allerdings ist dieses nur sinnvoll, wenn eine äquidistante Unterteilung der Kugel erfolgt, da sonst alle Histogrammwerte gleich sind.
 - Eigenwertmodell (Em), d. h. Ermittlung der Kovarianzmatrix für jede Shell und Verwendung der Eigenwerte.

- die verwendete Ähnlichkeitsfunktion, wobei die euklidische Distanz (L2) mit der von den Autoren vorgeschlagenen Quadratic Form Distance (QFD) verglichen wird.

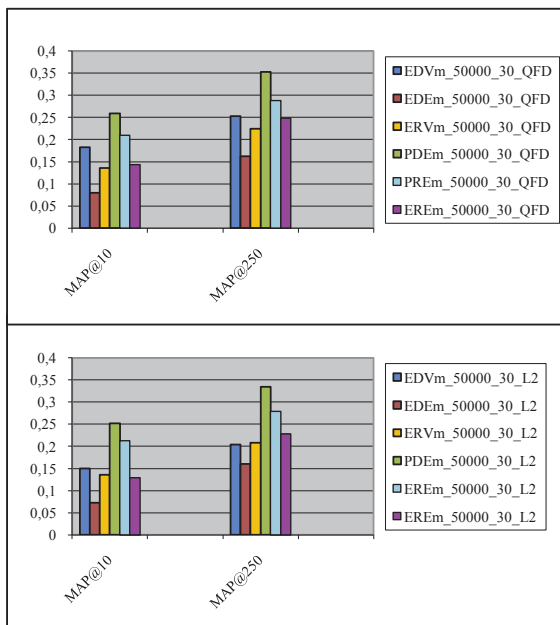


Abbildung 5.19. — Vergleich der MAP-Werte hinsichtlich der Parameterkombination bei Punktzahl $N = 50000$, Shellanzahl $s = 30$ und konstanter Ähnlichkeitsfunktion: QFD (oben) und L2 (unten)

Um allerdings eine Stellschraube untersuchen zu können, ist es notwendig, dass alle Stellschrauben bis auf den zu untersuchenden Parameter konstant bleiben. Folglich wurden für alle folgenden Analysen zunächst die Parameterkombination EDVm und die Ähnlichkeitsfunktion QFD zugrunde gelegt. Dies ist damit zu begründen, dass diese Kombination das ursprüngliche, von den Autoren entwickelte Konzept beschreibt, während alle weiteren Möglichkeiten als Optimierungen erst später entwickelt wurden. Davon ausgehend, wurden somit in einem ersten Schritt verschiedene Werte für die Shellanzahl s ($s = 5, 10, 30, 60, 120, 240$) untersucht, von denen der Wert $s = 30$ die besten Ergebnisse lieferte. Mit diesem Ergebnis wiederum wurden in einer zweiten Analyse die verschiedenen Parameterkombinationen untersucht. Wie der Vergleich

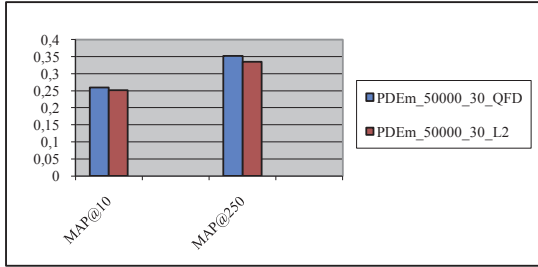


Abbildung 5.20. — Vergleich der MAP-Werte hinsichtlich der Ähnlichkeitsfunktion bei Punktzahl $N = 50000$, Shellerzahl $s = 30$ und konstanter Parameterkombination PDEm

der MAP-Werte für die Ähnlichkeitsfunktion QFD in der oberen Hälfte von Abbildung 5.19 zeigt, sollte eine proportionale Shellerzeugung mit einer disjunkten Punktzuordnung erfolgen, aus der schließlich unter Nutzung des Eigenwertmodells die Histogrammwerte zu ermitteln sind (Kombination PDEm). Diese Empfehlung kann zwar auch bei Verwendung der euklidischen Distanz als Ähnlichkeitsfunktion (vgl. untere Hälfte von Abbildung 5.19) abgeleitet werden; jedoch verdeutlicht Abbildung 5.20, dass sich die besseren Ergebnisse mit der QFD erzielen lassen. Ausgehend von den erreichten Ergebnissen könnte außerdem die Punktzahl N optimiert werden. Aufgrund der Zufälligkeit der erzeugten Punkte gilt jedoch, dass das Modell umso genauer approximiert wird, je mehr Punkte man verwendet. Demzufolge wird diese Analyse nicht weiterverfolgt.

Evaluation des Konzeptes Shape Distributions (SD)

Auch zur Repräsentation von 3D-Modellen durch das von Osada et al. entwickelte Konzept der SD sind Werte für Stellschrauben festzulegen, die im Rahmen der vorliegenden Evaluation untersucht wurden. Zu diesen gehören neben der Anzahl N der zufällig zu erzeugenden Objektoberflächenpunkte, die zur Charakterisierung der Objektgestalt zu verwendende *shape function*, die Anzahl b der Histogrammklassen und die für den Vergleich genutzte Ähnlichkeitsfunktion. Da allerdings die euklidische Distanz in den von den Autoren durchgeführten Evaluationen die besten Ergebnisse hervorgebracht hat, wird für die folgende Analyse dieses Distanzmaß konstant gehalten. Demzufolge wurden die anderen drei Stellschrauben jeweils für sich untersucht, wobei in Bezug auf die Punktzahl die Werte $N = 10, 20, 30, 40, 50, 100, 200, 400$ und hinsichtlich der Menge der Histogrammklassen die Werte $b = 15, 30, 60, 120, 250$ betrachtet wurden. Was die zur Beschreibung der Objektgestalt notwendige Funktion betrifft, erfolgte ein Vergleich der gemäß Osada et al. möglichen fünf *shape functions* [OFCD01]:

- A3: Berechnet den Winkel zwischen drei beliebigen Punkten.
- D1: Ermittelt die Distanz zwischen dem Objektmittelpunkt und einem beliebigen Punkt.
- D2: Misst den euklidischen Abstand zwischen zwei beliebigen Punkten.
- D3: Errechnet die Wurzel der Dreiecksfläche, die zwischen drei beliebigen Punkten definiert ist.
- D4: Bestimmt die Kubikwurzel des Tetraedervolumens, das von vier beliebigen Punkten gebildet wird.

Wie im Recall-Precision-Diagramm in Abbildung 5.21 zu erkennen ist, lassen sich bei diesem Konzept die besten Ergebnisse mit folgenden Einstellungen erzielen: Erzeugung von 400 zufälligen Punkten, wobei für jedes Punktpaar der euklidische Abstand mittels *shape function* D2 berechnet und schließlich in eine von 250 Histogrammklassen eingeordnet wird.

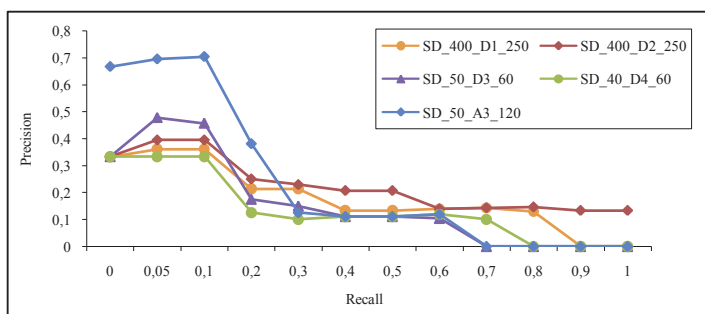


Abbildung 5.21. — Recall-Precision-Diagramm zum Vergleich der verschiedenen Parameter für das SD-Konzept, wobei für jede *shape function* das beste Evaluationsergebnis für den Vergleich verwendet wurde

Vergleichende Betrachtung der jeweils optimalen Konzepte

Nachdem die Konzepte einzeln betrachtet und die für sie jeweils besten Parameterkombinationen evaluiert wurden, sind diese nun in einer abschließenden Betrachtung einander gegenüberzustellen. Dies erfolgt mit Hilfe des in Abbildung 5.22 dargestellten Recall-Precision-Diagramms. Wie man hieraus erkennen kann, liefern prinzipiell sowohl das Konzept Hy-FV als auch das Konzept SIL-FV sehr gute Ergebnisse beim Vergleich von 3D-Modellen. Berücksichtigt man allerdings, dass für die Erstellung des hybriden FV insgesamt drei Ähnlichkeitskonzepte zu implementieren sind, sollte man

den für die Implementierung dieser Konzepte erforderlichen Aufwand zum erreichbaren Nutzen bzw. Mehrwert ins Verhältnis setzen. Da im Vergleich bereits mit weniger Aufwand, d. h. mit nur einem Ähnlichkeitskonzept, fast ebenso gute Ergebnisse erzielt werden können, empfiehlt sich eine Repräsentation von CAD-Modellen mittels des silhouettenbasierten Ansatzes.

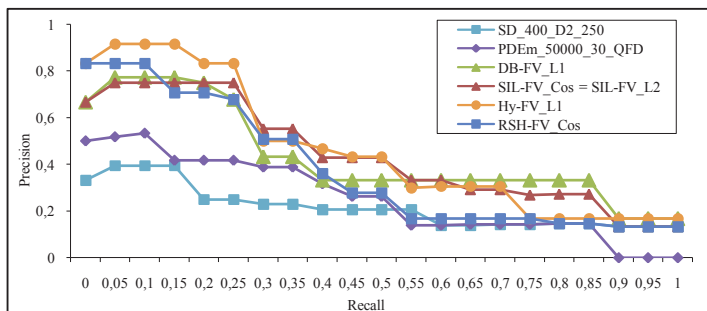


Abbildung 5.22. — Recall-Precision-Diagramm zum Vergleich der Konzepte mit den jeweils besten Parameterkombinationen

5.4.3. Berechnungsergebnis und Testbericht

Da die Tätigkeit des Entwerfens als Optimierungsprozess zu verstehen ist, in dem nach einer Feingestaltung des Produktes dieses im Hinblick auf die zu erfüllenden Anforderungen zu überprüfen und gegebenenfalls anzupassen ist [Hab08], sind neben CAD-Modellen auch Berechnungsergebnisse und Testberichte als wichtige Informationsquellen dieser Phase zu betrachten. Hierbei handelt es sich bei ersteren um die Ergebnisse durchgeführter Berechnungen, welche nach dem Begriffsverständnis in der vorliegenden Arbeit die eigentlichen Aufgaben des CAE darstellen. Da der Begriff CAE in der Literatur unterschiedlich definiert wird, indem er einerseits als Sammelbegriff für den Rechneinsatz im gesamten technischen Bereich, d. h. für CAD, CAM, Computer Aided Planning (CAP) und Computer Aided Quality (CAQ), und andererseits auch häufig als Synonym für den Begriff CAD Verwendung findet, wird in der vorliegenden Arbeit das von Wiendahl, Vajna und auch von Eigner und Stelzer postulierte Verständnis zugrunde gelegt. Danach ist CAE als ein Sammelbegriff für sämtliche AwS zu verstehen, die entweder als eigenständige Softwarelösungen (z. B. MSC Nastran¹⁰², Abaqus¹⁰³ oder COMSOL Multiphysics¹⁰⁴) oder integriert in einem

¹⁰²<http://www.mscsoftware.com/Products/CAE-Tools/MSC-Nastran.aspx>

¹⁰³http://www.simulia.com/products/unified_fea.html

¹⁰⁴<http://www.comsol.de>

CAD-System (z. B. Modul Pro/MECHANICA in Pro/ENGINEER, Abaqus for CATIA V5) zur Lösung von Berechnungs-, Optimierungs- und Simulationsaufgaben in der Entwurfsphase eingesetzt werden können [ES01, Vaj05, Wie08]. Während sich somit der CAD-Bereich auf den gestalterischen Aspekt der PE fokussiert, steht im CAE-Bereich der Prüfungsaspekt im Vordergrund. Ziel dabei ist es, die durch die Gestaltung festgelegten Ist-Produkteigenschaften mit den Soll-Eigenschaften zu vergleichen, wofür das 3D-Modell in einen Prototyp überführt werden muss [Vaj05]. Dieser wird typischerweise in digitaler Form durch Anwendung der sogenannten Finite Elemente Methode (FEM) erzeugt, welche die Geometriedaten des CAD-Modells in eine Menge bzw. ein Netz von endlich kleinen Elementen einfacher Geometrie transformiert [Rus08]. Die Komplexität des auf diese Weise erzeugten DMU ist dabei geringer als die des originalen CAD-Modells, was schnellere Berechnungsmethoden ermöglicht [Vaj05]. Allerdings reicht in manchen Fällen eine digitale Prüfung allein nicht aus, so dass zusätzlich ein physikalischer Prototyp zu fertigen ist. Dieser kann zur Durchführung verschiedener Tests genutzt werden, in deren Rahmen beispielsweise die Oberflächenbeschaffenheit, die Dichtigkeit oder gegebenenfalls auch die Temperaturbeständigkeit überprüft und deren Ergebnisse üblicherweise in einem Testbericht dokumentiert werden. Im Weiteren sollen daher die Ergebnisse beider Prüfverfahren getrennt voneinander betrachtet werden.

5.4.3.1. Berechnungsergebnis: Informationsgehalt und Repräsentationsmöglichkeiten

Die im CAE-Bereich durchgeführten Berechnungen sind gemäß Rust streng genommen in Berechnungen und Simulationen zu differenzieren [Rus08]. Während nämlich eine Berechnung als „(...) die rechnerische Erfassung eines Zustandes (...)“ [Rus08, S. 580] definiert ist, stellt eine Simulation die „(...) Berechnung einer Abfolge aufeinander aufbauender Zustände (...)“ [Rus08, S. 580] dar. Diese Zustände spiegeln dabei das Verhalten eines Produktes unter bestimmten vorgegebenen Rahmenbedingungen, welche von einfachen mechanischen oder thermischen Belastungen (z. B. Zug, Druck oder Temperatur) über strömungsmechanische Kennzahlen bis zu komplexen Kinetikbeziehungen reichen können, wider. Folglich sind abhängig von den Zielen, die man mit einer konkreten Berechnung bzw. Simulation erreichen möchte, unterschiedliche Verfahren anzuwenden. So findet man im Bereich der technischen PE vorrangig Verfahren zur Durchführung von einfachen Modellanalysen, Festigkeitsberechnungen, Strömungs-, Fertigungs- oder Mehrkörpersimulationen sowie von Kollisionsbetrachtungen (z. B. Crashtests) [ES01, Con08c]. Während für einfache Berechnungen meist Tabellenkalkulationsprogramme verwendet werden, hat sich, wie bereits erwähnt, die FEM als numerisches Verfahren zur Durchführung von komplexeren Berechnungen bewährt. Sie ermöglicht die Anwendung von Differentialgleichungen für jedes finite Element des FE-Netzes, aus deren Lösungen schließlich Rückschlüsse auf die Zustän-

de und damit das Verhalten eines Produktes gezogen werden können [Rus08]. Hierzu muss der DMU neben dem FE-Netz die genannten Randbedingungen in Form von Lastfällen, Materialdaten und / oder kinematischen Beziehungen als Inputdaten für die Berechnungssoftware bereitstellen. Die auf seiner Basis ermittelten Ergebnisse sind anschließend in einem Bericht zu dokumentieren, welcher gemäß Rust aus Gründen der Qualitätssicherung¹⁰⁵ mindestens folgende acht Informationen enthalten sollte [Rus08]:

- das Berechnungsziel (Auslegung, konstruktive Verbesserung, Nachweis, Untersuchung von Schadensfällen oder Untersuchung der Maßhaltigkeit),
- die Quelle und verwendete Version der Geometriebeschreibung,
- eventuell vorgenommene Idealisierungen,
- Lasten und Randbedingungen,
- Berechnungsoptionen,
- im Nichtlinearen der Verlauf des Lösungsvorgangs,
- Ergebnisprüfungen (z. B. Gleichgewicht, Plausibilität, Kontinuität der Spannungsverläufe, . . .),
- Ergebnisse, abgeleitete Werte und Schlussfolgerungen.

In der industriellen Praxis werden solche Ergebnisberichte jedoch häufig nicht manuell, sondern automatisch von den verwendeten AwS erzeugt. Wie die beiden exemplarischen Ergebnisberichte in den Abbildungen 5.23 und 5.24 zeigen, handelt es sich hierbei in den meisten Fällen um relativ einfache Textdokumente, die entweder in Form unstrukturierter Textdateien (z. B. im .txt- oder .dat-Dateiformat) oder als strukturierte HTML-Dateien vorliegen. Beide Beispiele verdeutlichen auch, dass diese automatisch generierten Berichte häufig nicht die von Rust geforderten Mindestinformationen enthalten. Vielmehr sind sie auf die Ergebnisse selbst und die zugehörigen Rahmenbedingungen beschränkt. Dabei werden die Ergebnisse entweder durch eine einfache Auflistung der ermittelten Parameterwerte dargestellt, wie es das grün markierte Rechteck in Abbildung 5.23 verdeutlicht. Dies ermöglicht ein einfaches Auslesen der Werte mittels Methoden des Pattern Matching, welche schließlich zur Bildung produktspezifischer Attributfacetten — hier der metrischen Attributfacetten Volumen, Flächeninhalt und Masse — genutzt werden können. Der ebenfalls in diesem Beispiel mit aufgeführte Wert für die Dichte ist hingegen ein werkstoffspezifischer Parameter, wobei zu berücksichtigen ist, dass die Datei selbst keinerlei Informationen über den Werkstoff, aus dem das Produkt besteht, enthält. Folglich kann diese Information keinem Werkstoff zugeordnet und damit für eine Suche nicht weiter ge-

¹⁰⁵„Das Prinzip bei der Dokumentation lautet, dass der Stellvertreter sich anhand der Unterlagen zurechtfinden können muss.“ [Rus08, S. 595]

```

Produkt = Pleuel.prt

VOLUMEN = 3.0222562e+05 MMA3
FLÄCHENINHALT = 5.5751094e+04 MMA2
MASSE = 2.3634044e-03 TONNE
DICHT = 7.8200000e-09 TONNE / MMA3

SCHWERPUNKT in Bezug auf den Koordinatenrahmen _PLEUEL:
X Y Z 0.0000000e+00 1.0054769e+02 0.0000000e+00 MM

Trägheit in Bezug auf Koordinatensystem _PLEUEL (TONNE * MMA2)

TRÄGHEITSTENSOR:
Ixx Ixy Ixz 3.9408794e+01 0.0000000e+00 0.0000000e+00
Iyx Iyy Iyz 0.0000000e+00 1.4390825e+00 0.0000000e+00
Izx Izy Izz 0.0000000e+00 0.0000000e+00 4.0759248e+01

Trägheit im Schwerpunkt in Bezug auf Koordinatenrahmen _PLEUEL (TONNE * MMA2)

TRÄGHEITSTENSOR:
Ixx Ixy Ixz 1.5515157e+01 -3.6850347e-06 0.0000000e+00
Iyx Iyy Iyz -3.6850347e-06 1.4390825e+00 0.0000000e+00
Izx Izy Izz 0.0000000e+00 0.0000000e+00 1.6865612e+01

HAUPTTRÄGHEITSMOMENT: (TONNE * MMA2)
I1 I2 I3 1.4390825e+00 1.5515157e+01 1.6865612e+01

Rotationsmatrix von _PLEUEL Orientierung auf Hauptachsen:
0.00000 -1.00000 0.00000
1.00000 0.00000 0.00000
0.00000 0.00000 1.00000

Rotationswinkel von _PLEUEL Orientierung auf Hauptachsen (in Grad):
Winkel um x y z 0.000 0.000 90.000

Umlaufbewegungsradius in Bezug auf die Hauptachsen:
R1 R2 R3 2.4675947e+01 8.1023141e+01 8.4475746e+01 MM

```

Abbildung 5.23. — Beispiel eines automatisch generierten Berichts einer in Pro/E ausgeführten Modellanalyse, deren Ergebnis aus einer einfachen Auflistung von produktspezifischen Parameterwerten besteht (grün markiert)

nutzt werden. Anstelle einer textuellen Auflistung werden Ergebnisse aber in vielen Fällen auch nur mit Hilfe von Bildern dargestellt, wie es das zweite Beispiel in Abbildung 5.24 zeigt. Hierbei handelt es sich um zweidimensionale Bilder, welche die Ergebnisse der Berechnung bzw. Simulation mit Hilfe von Farben visualisieren, die wiederum bestimmte Wertebereiche des untersuchten Parameters (z. B. der Spannung oder der Temperatur) kodieren. Diese Bilder erfordern folglich immer eine Interpretation durch den Produktentwickler, so dass sie für eine Suche nicht brauchbar sind.

Wie die Ergebnisse werden auch die Rahmenbedingungen sehr unterschiedlich beschrieben und sind daher je nach Anwendungsfall zu analysieren. So setzen sich beispielsweise die Rahmenbedingungen im Ergebnisbericht aus Abbildung 5.24 sowohl aus Werkstoffdaten als auch aus Informationen zum betrachteten Lastfall zusammen. In diesem Beispiel können somit neben dem Produkt selbst, der für das Produkt verwendete Werkstoff (hier Stahl) und seine Kennzahlen als werkstoffspezifische Attributfacetten extrahiert werden. Allerdings ist zu beachten, dass die Angabe der Zeichenkette STEEL eigentlich keine eindeutige Identifikation des Werkstoffes zulässt; schließlich existieren diverse Stahlsorten, so dass hier eine qualitativ höherwertige Information für eine Indexierung erforderlich ist. Des Weiteren kann das Dokument selbst um eine Attributfacette **Lastfall** ergänzt werden, welche eine Selektion an-

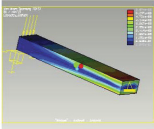
Mechanica Ergebnisbericht Inhaltsverzeichnis <ul style="list-style-type: none"> Window1 <ul style="list-style-type: none"> PRT0006.PRT <ul style="list-style-type: none"> Materialkennungen Lastset/Randbedingungen Lastset/Randbedingungssetze Materialien 																																																									
Window1  Abbildung 1: Window1																																																									
Verbindungen <ul style="list-style-type: none"> Analyse-Ergebnisdaten zeigen VS344: minmax 																																																									
PRT0006.PRT Materialkennungen <table border="1"> <tr> <th colspan="2">Materialkennung "MaterialAngul1"</th></tr> <tr> <td>Referenzen</td><td>Model (PRT0006.PRT)</td></tr> <tr> <td>Material</td><td>STEEL</td></tr> </table>		Materialkennung "MaterialAngul1"		Referenzen	Model (PRT0006.PRT)	Material	STEEL																																																		
Materialkennung "MaterialAngul1"																																																									
Referenzen	Model (PRT0006.PRT)																																																								
Material	STEEL																																																								
Lasten-Randbedingungen <table border="1"> <tr> <th colspan="2">Structural Last "Load1"</th></tr> <tr> <td>Referenzen</td><td>Laste (PRT0006.PRT)</td></tr> <tr> <td>Typ</td><td>Kraft/Moment</td></tr> <tr> <td>Koordinatensystem</td><td>WCS (PRT0006.PRT)</td></tr> <tr> <td>Verteilung</td><td>Gesamtlast</td></tr> <tr> <td>Räumliche Verteilung</td><td>Gleichmäßig</td></tr> <tr> <td>Kraft Komponenten</td><td>$F_x = 0, F_y = 0, F_z = 100 \text{ [N]}$</td></tr> <tr> <td>Moment Komponenten</td><td>$M_x = 0, M_y = 0, M_z = 0 \text{ [mm N]}$</td></tr> </table> <table border="1"> <tr> <th colspan="2">Structural Randbedingung "Constraint1"</th></tr> <tr> <td>Referenzen</td><td>Fläche (PRT0006.PRT)</td></tr> <tr> <td>Typ</td><td>Verschiebung</td></tr> <tr> <td>Koordinatensystem</td><td>WCS (PRT0006.PRT)</td></tr> <tr> <td>Verschiebung</td><td>Dx: Konstant Dy: Konstant Dz: Konstant</td></tr> <tr> <td>Rotation</td><td>Rx: Konstant Ry: Konstant Rz: Konstant</td></tr> </table> Lastset/Randbedingungssetze <table border="1"> <tr> <th colspan="2">Strukturlastset "LoadSet1"</th></tr> <tr> <td>Lasten</td><td>Load1</td></tr> </table> <table border="1"> <tr> <th colspan="2">Strukturmechanischer Randbedingungssetze "ConstraintSet1"</th></tr> <tr> <td>Randbedingungen</td><td>Constraint1</td></tr> </table> Materialien <table border="1"> <tr> <th colspan="2">Material "STEEL"</th></tr> <tr> <td>Typ</td><td>Isotrop</td></tr> <tr> <td>Dichte</td><td>$7.82708e-09 \text{ [tonne / mm}^3\text{]}$</td></tr> <tr> <td>Charakteristischer</td><td>$199948 \text{ [N / mm}^2\text{]}$</td></tr> <tr> <td>Querkontraktionszahl</td><td>0.27</td></tr> <tr> <td>Leitfähigkeit</td><td>$43.0125 \text{ [W / (sec C)]}$</td></tr> <tr> <td>Spezifische Wärme</td><td>$4.7344e+08 \text{ [mm}^2 \text{ / (sec}^2 \text{ C)]}$</td></tr> <tr> <td>Wärmeausdehnung</td><td>$1.17e-05 \text{ [C]}$</td></tr> <tr> <td>Schulterfestigkeit</td><td>$78719.7 \text{ [N / mm}^2\text{]}$</td></tr> <tr> <td>Vergewandertes</td><td>None</td></tr> </table>		Structural Last "Load1"		Referenzen	Laste (PRT0006.PRT)	Typ	Kraft/Moment	Koordinatensystem	WCS (PRT0006.PRT)	Verteilung	Gesamtlast	Räumliche Verteilung	Gleichmäßig	Kraft Komponenten	$F_x = 0, F_y = 0, F_z = 100 \text{ [N]}$	Moment Komponenten	$M_x = 0, M_y = 0, M_z = 0 \text{ [mm N]}$	Structural Randbedingung "Constraint1"		Referenzen	Fläche (PRT0006.PRT)	Typ	Verschiebung	Koordinatensystem	WCS (PRT0006.PRT)	Verschiebung	Dx: Konstant Dy: Konstant Dz: Konstant	Rotation	Rx: Konstant Ry: Konstant Rz: Konstant	Strukturlastset "LoadSet1"		Lasten	Load1	Strukturmechanischer Randbedingungssetze "ConstraintSet1"		Randbedingungen	Constraint1	Material "STEEL"		Typ	Isotrop	Dichte	$7.82708e-09 \text{ [tonne / mm}^3\text{]}$	Charakteristischer	$199948 \text{ [N / mm}^2\text{]}$	Querkontraktionszahl	0.27	Leitfähigkeit	$43.0125 \text{ [W / (sec C)]}$	Spezifische Wärme	$4.7344e+08 \text{ [mm}^2 \text{ / (sec}^2 \text{ C)]}$	Wärmeausdehnung	$1.17e-05 \text{ [C]}$	Schulterfestigkeit	$78719.7 \text{ [N / mm}^2\text{]}$	Vergewandertes	None
Structural Last "Load1"																																																									
Referenzen	Laste (PRT0006.PRT)																																																								
Typ	Kraft/Moment																																																								
Koordinatensystem	WCS (PRT0006.PRT)																																																								
Verteilung	Gesamtlast																																																								
Räumliche Verteilung	Gleichmäßig																																																								
Kraft Komponenten	$F_x = 0, F_y = 0, F_z = 100 \text{ [N]}$																																																								
Moment Komponenten	$M_x = 0, M_y = 0, M_z = 0 \text{ [mm N]}$																																																								
Structural Randbedingung "Constraint1"																																																									
Referenzen	Fläche (PRT0006.PRT)																																																								
Typ	Verschiebung																																																								
Koordinatensystem	WCS (PRT0006.PRT)																																																								
Verschiebung	Dx: Konstant Dy: Konstant Dz: Konstant																																																								
Rotation	Rx: Konstant Ry: Konstant Rz: Konstant																																																								
Strukturlastset "LoadSet1"																																																									
Lasten	Load1																																																								
Strukturmechanischer Randbedingungssetze "ConstraintSet1"																																																									
Randbedingungen	Constraint1																																																								
Material "STEEL"																																																									
Typ	Isotrop																																																								
Dichte	$7.82708e-09 \text{ [tonne / mm}^3\text{]}$																																																								
Charakteristischer	$199948 \text{ [N / mm}^2\text{]}$																																																								
Querkontraktionszahl	0.27																																																								
Leitfähigkeit	$43.0125 \text{ [W / (sec C)]}$																																																								
Spezifische Wärme	$4.7344e+08 \text{ [mm}^2 \text{ / (sec}^2 \text{ C)]}$																																																								
Wärmeausdehnung	$1.17e-05 \text{ [C]}$																																																								
Schulterfestigkeit	$78719.7 \text{ [N / mm}^2\text{]}$																																																								
Vergewandertes	None																																																								

Abbildung 5.24. — Beispiel eines automatisch generierten Ergebnisberichts zu einer in Pro/MECHANICA durchgeführten Simulation bei einer Belastung durch Verschiebung

hand der im Dokument betrachteten Lastfälle, d.h. der „(...)“ verschiedenen Arten der Änderung von Belastungsgrößen (...)“ [Con08c, S. 169], ermöglicht. Dazu sind die im Dokument enthaltenen Kennzahlen zur Beschreibung des Lastfalls (Referenzen = Kante (PRT0006.prt), Typ = Kraft/Moment, Verteilung = Gesamtlast, Räumliche Verteilung = Gleichmäßig, Kraft Komponenten = $F_x=0, F_y=0, F_z=100\text{[N]}$, ...) und seiner zugehörigen Randbedingungen (Typ = Verschiebung, Verschiebung = D_x : Konstant, D_y : Konstant, D_z : Konstant, ...) auszuwerten. Anhand dieser Werte lässt sich erkennen, dass es sich im vorliegenden Fall um eine mechanische Belastung durch Verschiebung bei einer Kraftauswirkung von 100 Newton handelt. Folglich könnte das Dokument prinzipiell um entsprechende, die einzelnen Kennzahlen wiedergebende Attributfacetten erweitert werden. Allerdings stellt sich die Frage, ob diese wirklich hilfreiche Kriterien für eine Suche darstellen. Häufig könnte es durchaus ausreichen, auf Basis dieser Informationen eine Einordnung des Lastfalls in die in der Praxis bekannten idealisierten Lastfälle (hier Lastfall I) vorzunehmen und diese als Filterkriterien heranzuziehen. Gemäß Conrad würden sich bei diesem Vorgehen folgende vier Ausprägungen für eine Filterung verwenden lassen [Con08c]:

- Lastfall I $\hat{=}$ ruhende Beanspruchung, d. h. die Spannung steigt zügig auf einen bestimmten Wert und behält diesen während einer bestimmten Zeit.
- Lastfall II $\hat{=}$ schwellende Beanspruchung, d. h. die Spannung steigt zunächst von 0 auf einen Höchstwert und sinkt dann wieder auf 0 ab.
- Lastfall III $\hat{=}$ wechselnde Beanspruchung, d. h. die Spannung schwankt ständig zwischen einem positiven und einem negativen Höchstwert.
- Lastfall I+III $\hat{=}$ schwingende Beanspruchung, d. h. die Spannung schwankt mit einer Ausschlagsspannung um einen Mittelwert.

Sollten weitere, der von Rust geforderten Mindestinformationen verfügbar sein, sollten auch diese auf eine Eignung als Filterkriterien untersucht werden. Anbieten würde sich hier beispielsweise die Information zum Berechnungsziel, welche als nominale Attributfacette mit einer der oben aufgeführten Ausprägungen dem Dokument hinzuzufügen wäre.

5.4.3.2. Testbericht: Informationsgehalt und Repräsentationsmöglichkeiten

Lassen sich bestimmte Produktanforderungen nicht mit Hilfe digitaler Berechnungsmethoden überprüfen oder sind im Pflichtenheft bestimmte Tests bzw. Prüfungen zwingend gefordert (vgl. Kapitel 5.4.1.3), so sind manuelle Versuche durchzuführen, die Aufschluss darüber geben, ob ein erwartetes Ergebnis erreicht wird oder nicht. Dabei hängt die Art des durchzuführenden Tests wiederum von der jeweiligen zu untersuchenden Anforderung ab, so dass grundsätzlich zwischen verschiedenen Testarten, wie beispielsweise Wärmewechselprüfungen, Dichtheitstest und anderen, zu differenzieren ist. Die dabei beobachteten oder gemessenen Ergebnisse sind ebenfalls in Ergebnisberichten zu dokumentieren, welche allerdings manuell zu erstellen sind und folglich meist in Form von Textdokumenten im DOC/DOCX- oder PDF-Format gespeichert werden. Darüber hinaus sollten in gewisser Weise natürlich auch hier die von Rust vorgeschlagenen acht Mindestinformationen (siehe Kapitel 5.4.3.1) in den Berichten enthalten sein, um sozusagen eine interne Qualitätssicherung gewährleisten zu können. Da jedoch die Gestaltung und auch der Inhalt von manuell erzeugten Testberichten zum einen von der Testart und zum anderen von den eventuell vorhandenen Dokumentvorlagen des Unternehmens oder zumindest vom Ersteller des Berichts abhängen, existiert auch hier wie im Fall von Lasten- und Pflichtenheften eine enorme Vielfalt. Daher sollen im Weiteren stellvertretend am Beispiel eines Prüfberichts aus der industriellen Praxis im DOC/DOCX-Dateiformat, welcher in Abbildung 5.25 dargestellt ist, mögliche Informationen und die daraus erstellbaren Repräsentationen für Testberichte aufgezeigt werden.

Wie hierbei zu erkennen ist, kann der in Abbildung 5.25 veranschaulichte Testbericht grundsätzlich in vier Abschnitte unterteilt werden, in denen organisatorische Daten,

046 / 04

Prüfbericht

046 / 04

Prüfbericht

Abbildung 5.25. — Beispiel eines im DOC/DOCX-Format gespeicherten Prüfberichts

Informationen zum Test selbst, das Testergebnis und schließlich sonstige Bemerkungen bzw. Informationen aufgeführt sind. Aus den organisatorischen Daten, welche zu Beginn des Dokumentes sowie am Ende jeder Seite verfügbar sind, können zum einen personenspezifische Informationen, wie die Person, die den Test in Auftrag gegeben hat (Auftraggeber) und die Person, die das Dokument erstellt, geändert bzw. freigegeben hat, ausgelesen werden. Zum anderen wird hier das getestete Produkt anhand seines Produktnamens (Teile-Bezeichnung) und seiner Sachnummer referenziert. Somit sollten zusätzlich zum Dokument weitere Artefakte (Personen und das Produkt) in Form eigenständiger **IndexableArtifact**-Objekte erstellt und um die jeweils zugehörigen Attribut- und entsprechenden Beziehungsfacetten ergänzt werden. Des Weiteren geben die organisatorischen Daten Auskunft darüber, wann ein Dokument erstellt, geändert bzw. freigegeben wurde, was die Bildung von dokumentspezifischen Attributfacetten erlaubt. Außerdem erfolgt hier auch die Angabe, welche Art von Test durchgeführt wurde, indem dieser anhand von Auswahlfeldern näher charakterisiert wird. Dazu werden drei Kategorisierungen vorgenommen, die dem Test die jeweils zugehörige Domäne (mechanisch, elektrisch, thermisch oder chemisch), den mit ihm verfol-

ten Zweck (Versuchsmuster, Erstmuster, Materialeingangsprüfung, Fertigungs-/ Ausgangsprüfung und Retourenprüfung) sowie eines der im betrachteten Unternehmen unterschiedenen Testverfahren (Dauerprüfung, Einzelprüfung und Röntgenprüfung) zuordnen. Das Problem hierbei ist allerdings, dass nur durch eine Interpretation der Auswahlfelder auf die mit ihnen verbundene Semantik geschlossen werden kann. Ferner muss vor einer eventuellen Facettenbildung ermittelt werden, welche der jeweiligen Auswahlfelder in einem konkreten Dokument tatsächlich ausgewählt sind. Während die Semantik im Rahmen der Extraktorkomponente fest kodiert werden müsste, kann die Information, welches Auswahlfeld ausgewählt ist, automatisch erfasst werden. Allerdings haben die im Rahmen der prototypischen Implementierung durchgeführten Versuche hierzu gezeigt, dass bei einer Verarbeitung der Datei mit der Softwarebibliothek Apache POI⁸¹ diese Auswahlfelder verloren gehen. Eine alternative Lösung wäre die Nutzung der ebenfalls frei verfügbaren Softwarebibliothek von OpenOffice.org¹⁰⁶, was allerdings eine Konvertierung der DOC/DOCX-Datei ins Open Document Format (.odt) und damit eine Installation des OpenOffice.org-Softwarepaketes voraussetzt. Das Open Document Format ist ein XML-basiertes Dateiformat, das für jedes Auswahlfeld ein `<form\>`-Element anlegt, in dem dessen Status als `checked` oder `not checked` definiert ist. Somit könnten diese Elemente mittels eines XML-Parsers ausgelesen und die entsprechenden Informationen extrahiert werden, um daraus dokumentenspezifische Attributfacetten für eine Selektion anhand dieser Testmerkmale zu bilden.

Im zweiten Abschnitt der den Test beschreibenden Informationen sind weitere Testmerkmale enthalten. So beschreibt beispielsweise der Punkt *Prüftitel/Prüfung*, um welche Art von Test es sich handelt (hier Dichtheitstest). Außerdem werden der *Prüfumfang* und der *Prüfablauf* näher beschrieben. Letzterer konkretisiert den Verlauf der Testdurchführung in natürlicher Sprache und geht dabei auch auf die Prüfparameter ein (hier: Wassersäule von 300 mm über eine Dauer von 1 Stunde), welche durchaus nützliche Suchkriterien darstellen würden. Daher sollten diese Textinformationen, ebenso wie Bemerkungen oder sonstige Angaben des vierten Abschnitts, in jedem Fall als Ausprägungen einer dokumentenspezifischen Ähnlichkeitsfacette zur Schlagwortsuche (QbT) gesetzt werden. Sind die Prüfparameter — wie es in manchen Dokumenten gegeben ist — hingegen separat aufgelistet, so könnte eine zusätzliche Attributfacette für diese dem Dokument hinzugefügt werden.

Der dritte Abschnitt enthält schließlich das Ergebnis des Tests, das in natürlicher Sprache formuliert zur Bildung eines Suchkriteriums nicht sinnvoll genutzt werden kann. Dies gilt jedoch nicht, wenn hier beispielsweise im Test erzielte Messwerte dokumentiert werden. Aus den gemessenen Parametern (z. B. Temperatur) könnte man zum einen auf die Prüfparameter schließen und somit die entsprechende Attributfacette erzeugen oder erweitern. Zum anderen wäre auch die Bildung einer Attributfacette

¹⁰⁶<http://api.openoffice.org>

für jeden einzelnen gemessenen Parameter und seinen Wert denkbar; jedoch wäre dann zu entscheiden, ob diese als dokumentspezifische Attributfacetten zur Verfeinerung von Testberichten dienen sollen oder ob diese, da sie eigentlich das Produkt näher beschreiben, als produktspezifische Attributfacetten zu erstellen sind.

Zusammenfassend ist festzuhalten, dass der Informationsgehalt von Testberichten und auch von Berechnungsergebnissen stark variiert, so dass keine allgemeingültigen Regeln zur Extraktion von Informationen gegeben werden können. Vielmehr sind die im Unternehmen existierenden Dokumente auf ihren Informationsgehalt und die daraus ermittelbaren Suchkriterien mit deren Relevanz bzw. Nützlichkeit hin zu analysieren. In jedem Fall sollte auch hier zwischen den verschiedenen, im Dokument adressierten Artefakten unterschieden werden, für die jeweils entsprechende Facetten (Beziehungs-, Ähnlichkeits- und /oder Attributfacetten) erzeugt werden. Für die Extraktion kann dabei entweder auf einfache Verfahren des Pattern Matching oder unter Umständen auf Methoden des NLP, wie sie in Kapitel 5.4.1.3 vorgestellt wurden, zurückgegriffen werden. Jedoch ist auch hier zu berücksichtigen, dass der Aufwand für die Extraktion einzelner Informationen sehr hoch sein kann. Aus diesem Grund sollte man sehr gut abwägen, bei welchen Kriterien dieser Aufwand in Kauf genommen werden soll und bei welchen bereits eine Ähnlichkeitsfacette zur Schlagwortsuche ausreichen würde.

5.4.4. Technische Zeichnung

Erfüllt das in der Entwurfsphase gestaltete Produkt alle Anforderungen, so können in der darauf folgenden Ausarbeitungsphase die zur Fertigung des Produktes benötigten Unterlagen erstellt werden. Obwohl heutzutage auch hier das 3D-Modell immer öfter für die Weitergabe von Fertigungsinformationen verwendet wird, stellt v. a. in KMU die technische Zeichnung noch immer ein wichtiges Kommunikationsmittel zwischen Konstruktions- und Fertigungsabteilung dar. Dabei ist der Begriff der technischen Zeichnung gemäß DIN-Norm 199 eigentlich als Sammelbegriff für alle zeichnerischen Unterlagen zu verstehen, welche sich durch die Art ihrer Darstellung, die Art ihrer Anfertigung, ihren Inhalt und ihren Zweck unterscheiden (z. B. Skizze, Original-, Entwurfs-, Gesamt-, Fertigungszeichnung, ...) [Deu62]. In der vorliegenden Arbeit wird dieser Begriff jedoch etwas enger gefasst, so dass mit einer technischen Zeichnung eine nach Normen erstellte und aus Linien bestehende, graphische Abbildung verbunden wird, welche das Produkt maßstabsgetreu in zwei Dimensionen veranschaulicht und alle notwendigen Angaben zu dessen Herstellung enthält [ABB⁺06, Con08e]. Abhängig von der Granularität des abzubildenden Produktes kann dabei zwischen Einzelteilzeichnungen und Gesamtzeichnungen differenziert werden. Eine Einzelteilzeichnung dient zur Beschreibung der mit Maßangaben versehenen Geometrie eines Einzelteils, welche in Form verschiedener Ansichten (siehe Abbildung 5.26 links) und /oder Schnitte dargestellt wird [VWSS94, ABB⁺06]. Eine Gesamtzeichnung hin-

gegen beschreibt eine Baugruppe entweder im zusammengebauten Zustand (vgl. Abbildung 5.26 in der Mitte) oder verdeutlicht deren Montagezustand und somit die Anordnung der einzelnen Komponenten sowie deren Beziehungen zueinander (siehe Abbildung 5.26 rechts) [EM91, ABB⁺06]. Da bei letzterer die Baugruppe derart dargestellt ist, wie sie bei einer Explosion zerbersten würde, spricht man hier auch häufig von einer sogenannten Explosionszeichnung. Trotz dieser Unterschiede folgt, wie in den drei Abbildungen zu erkennen ist, jede technische Zeichnung einem einheitlichen Aufbau, der sich aus zwei wesentlichen Bestandteilen — einer Zeichnungsfläche und einem Schriftfeld — zusammensetzt und auf den im Rahmen der weiteren Betrachtung noch eingegangen wird [Con08e].

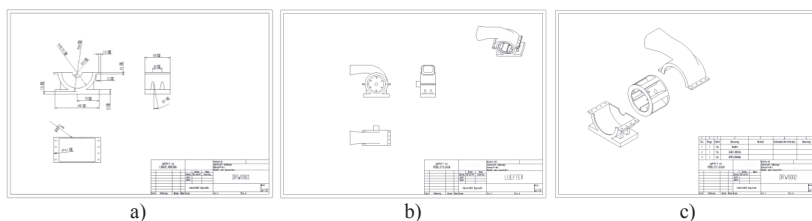


Abbildung 5.26. — Verschiedene Arten technischer Zeichnungen: a) Einzelteilzeichnung des Untergehäuses eines Lüfters; b) Gesamtzeichnung des Lüfters; c) Explosionszeichnung des Lüfters

In Anlehnung an Kapitel 5.4.2 wird im Folgenden zunächst der Prozess der Zeichnungserstellung zusammen mit den zur Speicherung verwendeten Dateiformaten näher betrachtet (vgl. Kapitel 5.4.4.1 und 5.4.4.2). Anschließend erfolgt eine Unterteilung des Informationsgehalts technischer Zeichnungen in drei Kategorien (Kapitel 5.4.4.3), für die Extraktions- und Repräsentationsmöglichkeiten zusammen mit den dabei zu berücksichtigenden Herausforderungen (Kapitel 5.4.4.4) aufgezeigt werden. Eine Darstellung der Evaluationsergebnisse von prototypisch umgesetzten Ähnlichkeitskonzepten für technische Zeichnungen (Kapitel 5.4.4.5) schließt diese Betrachtung.

5.4.4.1. Erstellung und Archivierung technischer Zeichnungen

Während noch vor etwa 20-30 Jahren eine technische Zeichnung am Zeichenbrett mit Lineal und Bleistift gezeichnet wurde [Sch87, VWSS94], erfolgt deren Erstellung heutzutage in nahezu allen Unternehmen in digitaler Form. Dazu werden entweder 2D-CAD-Systeme wie AutoCAD oder in den meisten Fällen die bereits in Kapitel 5.4.2.1 vorgestellten 3D-CAD-Systeme eingesetzt. Da diese auf Basis von 2D-Konstruktionsprogrammen entstanden sind, verfügt jedes dieser Systeme auch über Funktionalitäten zur zweidimensionalen Zeichnungserstellung. Diese Funktionalität besteht dabei vor-

rangig in einer automatischen Ableitung der Zeichnungsdaten aus den 3D-Daten, was nicht nur den Erstellungs-, sondern auch den Änderungsprozess technischer Zeichnungen wesentlich vereinfacht. Hierzu werden die 2D-Daten assoziativ mit den 3D-Daten verknüpft, wodurch sich Änderungen am 3D-Modell automatisch in der Zeichnung auswirken und umgekehrt [Con08e].

Die Speicherung erstellter Zeichnungen erfolgt dabei zum einen, wie bei CAD-Modellen, mit Hilfe proprietärer Datenformate, deren Dokumentationen nicht offen gelegt sind und auf die nur über teure Konverter zugegriffen werden kann. Als Beispiele werden im Folgenden für die in Kapitel 5.4.2.1 vorgestellten CAD-Systeme die zur Identifikation einer technischen Zeichnung verwendeten Dateierstendungen aufgelistet:

- AutoCAD bzw. AutoCAD Mechanical: .dwg
- Autodesk Inventor: .idw
- Pro/ENGINEER : .drw
- CATIA (ab Version V5): .CATDrawing
- SolidWorks: .slddrw
- Solid Edge: .dft

Zum anderen integrieren manche CAD-Systeme die Zeichnungsdaten aber auch in die erstellten 3D-Daten, wie dies beispielsweise bei den Systemen NX und CATIA V4 der Fall ist. Folglich werden hier keine separaten Dateien für Zeichnungen erstellt; vielmehr muss für eine Zeichnungsansicht immer erst das 3D-Modell (hier die .prt- bzw. .model-Datei) geöffnet und anschließend in das Funktionsmodul für die Zeichnungserstellung gewechselt werden [EM91, VWSS94].

Aufgrund dieser Vielfalt an unterschiedlichen Datenformaten und der damit verbundenen Probleme beim Datenaustausch (multiple teure Konverter, Datenverlust, manueller Nachbearbeitungsaufwand, . . .) wurden schon frühzeitig neutrale Formate für die Übertragung von technischen Zeichnungen entwickelt. Eines der ersten dieser Formate ist das im Jahr 1982 vom Unternehmen Autodesk, Inc. für das CAD-System AutoCAD entwickelte Drawing Interchange Format (DXF). Dieses hat sich trotz fehlender Standardisierung im Laufe der Jahre als Quasi-Standard für 2D-Daten in der Industrie etabliert und stellt heute im Vergleich zu den ebenfalls für Zeichnungen nutzbaren Formaten IGES, STEP und in jüngster Zeit PDF in diesem Bereich das am häufigsten genutzte Format dar. Aus diesem Grund wird im Weiteren etwas detaillierter auf dieses Austauschformat eingegangen.

5.4.4.2. DXF als Austauschformat

Das DXF-Format, welches manchmal fälschlicherweise auch als Drawing Exchange Format oder auch Drawing Exchange File bezeichnet wird, gehört zur Familie der vektorbasierten Graphikformate. Folglich definiert es eine zweidimensionale Darstellung mit Hilfe graphischer Primitive wie Punkte, Linien, Kreise und Polygone. Allerdings wird das Format mit jeder neuen Version von AutoCAD erweitert, so dass DXF seit dem Release von AutoCAD 2011 mittlerweile als Version AC1024 verfügbar ist. Trotz dieser häufigen Versionsänderungen gehört eine DXF-Unterstützung heutzutage für jedes CAD-System zu den Basisanforderungen, was v. a. dadurch begünstigt ist, dass Autodesk zum einen auf eine Kompatibilität der DXF-Versionen untereinander achtet und zum anderen die zugehörigen Dokumentationen (aktuell Version v.u.25.1.01) auf seiner Internetseite¹⁰⁷ kostenfrei zur Verfügung stellt. Gemäß dieser Spezifikation werden Zeichnungen entweder als ASCII-basierte Textdateien oder im Binärformat gespeichert, wobei für beide Varianten die Dateiendung .dxf verwendet wird. Obwohl es sich beim Binärformat um ein wesentlich kompakteres Dateiformat handelt, das etwa 25 % weniger Speicherplatz als die ASCII-basierte Version benötigt und von AutoCAD bis zu fünfmal schneller geschrieben und gelesen werden kann, wird es in der Praxis nur in sehr wenigen Anwendungsfällen genutzt. Der Großteil der Zeichnungen in der industriellen Praxis liegt in ASCII-Form vor und verfügt daher über einen sehr einfachen und leicht verständlichen Aufbau. Dieser setzt sich aus mehreren hintereinander aufgelisteten Datensätzen zusammen, wobei jeder Datensatz durch einen sogenannten Gruppencode eingeleitet wird. Dieser Gruppencode hat die Form einer Ganzzahl und dient zur eindeutigen Charakterisierung der zum Datensatz gehörenden Elemente bzw. Werte. Zusätzlich werden die Datensätze in insgesamt sieben Abschnitte unterteilt, wobei jeder Abschnitt nach der Form

```
0 SECTION 2 Abschnittsname Datensätze 0 ENDSEC
```

aufgebaut ist und zur Speicherung der folgenden Informationen dient:

HEADER Enthält allgemeine Informationen zu einer Zeichnung, wie beispielsweise die zugrunde liegende AutoCAD-Version, die DXF-Version und weitere Systemvariablen.

CLASSES Definiert die Objektklassen, deren Instanzen als Entitäten in den folgenden Abschnitten verwendet werden können.

TABLES Enthält Tabellen für die verschiedenen Linientypen, Textarten, Dimensionsstile usw.

BLOCKS Beschreibt verschiedene Blockelemente, die sich aus mehreren Entitäten und / oder Blockelementen zusammensetzen und im folgenden ENTITIES-Abschnitt mehrfach referenziert werden können.

¹⁰⁷<http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454>

ENTITIES Definiert die einzelnen Entitäten und Blockelemente zur Beschreibung der graphischen Objekte einer Zeichnung.

OBJECTS Dient zur Speicherung aller nicht-graphischen Objekte (Texte, Dimensionen, ...) einer Zeichnung.

THUMBNAILIMAGE Optionaler Abschnitt, der Bilddaten der Zeichnung zum Zweck einer Vorschau enthalten kann.

Zusätzlich stellt das DXF-Format die Anwendung der sogenannte Layertechnik zur Verfügung, welche als organisatorisches Hilfsmittel die Anfertigung einer Zeichnung auf mehreren Ebenen ermöglicht. Dabei können die einzelnen Aspekte einer Zeichnung, wie beispielsweise die geometrische Beschreibung eines Produktes oder seine Bemaßung, auf separaten Ebenen erstellt und schließlich zur Gesamtzeichnung zusammengefasst werden. Dieses Vorgehen entspricht in etwa der Erstellung einer Zeichnung auf mehreren transparenten Zeichenblättern, die übereinander gelegt werden können, so dass durch einfaches Wegnehmen oder Hinzufügen eines Blattes bzw. einer Ebene die graphische Darstellung an den jeweils benötigten Verwendungszweck angepasst werden kann. [Sch87, RSW00, Aut10a]

5.4.4.3. Informationsgehalt technischer Zeichnungen

Prinzipiell kann der Informationsgehalt jeder technischen Zeichnung in die drei Kategorien organisatorische Daten, geometrische Informationen und technologische Informationen unterteilt werden, deren Positionierung auf einer Zeichnung durch Normangaben fest vorgeschrieben sind [Sch87, ABB⁺06, Con08e, Wie08]. So dient das Schriftfeld in der rechten unteren Ecke einer Zeichnung zur Angabe organisatorischer Daten, welches einem in der DIN-Norm DIN EN ISO 7200 (früher DIN 6771-1) vorgeschriebenen Layout folgen muss. Danach setzt sich das Schriftfeld aus mehreren Grundschriftfeldern unterschiedlicher Größe zusammen, in denen die für die Auftragsabwicklung und interne Organisation benötigten dokument- und produktbezogenen Informationen einzutragen sind. Zu den dokumentbezogenen Informationen sind hierbei

- die Zeichnungsnummer (eventuell ergänzt um eine Blattnummer)
- das Zeichnungsformat (DIN A4, DIN A3, ...),
- der Maßstab der Zeichnung (1:1, 1:5, 1:10, ...),
- das Erstellungsdatum,
- der Autor der Zeichnung,
- Prüf- und Freigabevermerke,

- Änderungsinformationen,
- die Nummer des zur Zeichnung gehörenden Auftrags,
- Kundendaten sowie
- Angaben zu dem die Zeichnung erstellenden Unternehmen

zu zählen. Als produktbezogene Informationen hingegen sind der in der Zeichnung als Benennung oder Titel bezeichnete Produktname, die Sachnummer, eine Werkstoffangabe und eventuelle Hinweise auf beachtende Normen zu Allgemeintoleranzen und Oberflächenangaben im Schriftfeld zu definieren.

Die Angabe von geometrischen und technologischen Informationen hingegen erfolgt innerhalb der Zeichnungsfläche. Dabei handelt es sich bei den geometrischen Informationen um eine detaillierte Beschreibung der zweidimensionalen Produktgestalt in Form von mindestens einer, jedoch üblicherweise von drei (bei komplexeren Produkten auch sechs) Ansichten, die nach vorgeschriebenen Regeln auf der Zeichnung anzuordnen sind (z. B. bei drei Ansichten: Vorderansicht links oben, Seitenansicht rechts oben und Draufsicht links unten). Diese Ansichten werden üblicherweise durch eine Parallelprojektion des Produktes auf verschiedene, senkrecht zueinander stehende Ebenen erzeugt. Dabei ist grundsätzlich diejenige Ansicht als Vorder- und damit als Hauptansicht zu wählen, die alle wesentlichen Merkmale eines Produktes verdeutlicht und somit am aussagekräftigsten ist. Des Weiteren können diese Ansichten einerseits um eine axonometrische (üblicherweise eine isometrische) Ansicht und andererseits um Schnittdarstellungen ergänzt werden. Während erstere einen räumlichen Eindruck des Produktes vermitteln soll, dienen letztere zu einer übersichtlicheren und verständlicheren Beschreibung von inneren Komponenten eines Produktes. [Con08e, Wie08]

Zusätzlich zu diesen geometrischen Informationen muss eine technische Zeichnung aber auch sogenannte technologische Informationen beinhalten, die alle für die Fertigung des Produktes relevanten Angaben umfassen. Dazu gehören zum einen Maßangaben, welche die Abmessungen des Produktes eindeutig festlegen. Beispiele sind Längen-, Durchmesser-, Gewinde- oder Winkelmaße, für die zusätzlich mögliche Toleranzangaben eingetragen werden können. Zum anderen zählen auch werkstoffbezogene Daten (z. B. Gewicht, wichtige Werkstoffkennwerte), Angaben zur Oberflächenbeschaffenheit sowie zur Art der Oberflächenbehandlung und qualitätsbezogene Angaben zu dieser Kategorie. Letztere beziehen sich u. a. auf Fertigungs-, Abnahme-, Verpackungs- bzw. Transportbedingungen sowie Liefervereinbarungen. [ABB+06, Con08e, Wie08]

Obwohl sich prinzipiell jedes Unternehmen an diesen Normangaben und den mit ihnen verbundenen inhaltlichen Vorgaben orientiert, finden sich selbstverständlich in jedem Unternehmen auch Abweichungen hiervon. Demzufolge kann nicht von einer einheitlichen Gestaltung technischer Zeichnungen ausgegangen werden. Vielmehr basiert diese

zum einen auf unternehmensspezifischen Vorgaben und zum anderen auf der Arbeits- und Vorgehensweise des Zeichnererstellers. So können sich beispielsweise auch zusätzliche Schriftfelder oder wichtige Herstellungshinweise in Form graphischer Abbildungen innerhalb der Zeichnungsfläche befinden, wie es stellvertretend anhand einer aus der industriellen Praxis stammenden technischen Zeichnung in Abbildung 5.27 verdeutlicht ist.

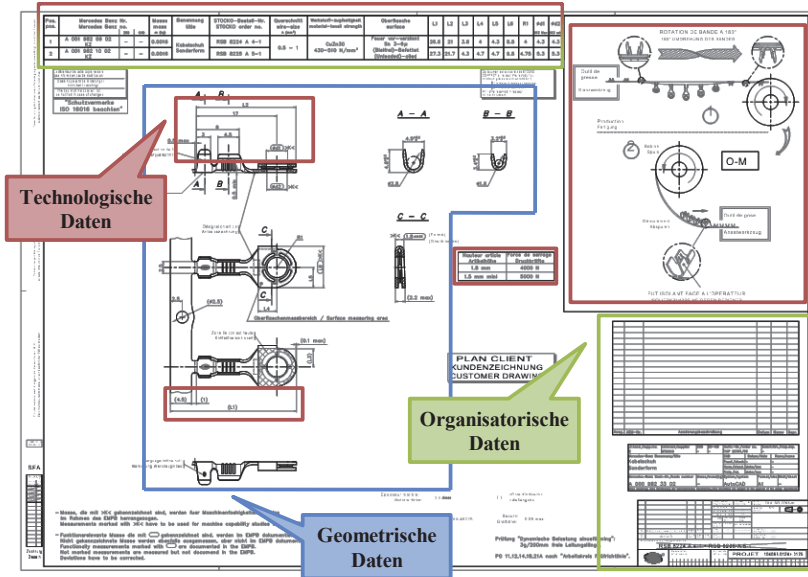


Abbildung 5.27. — Beispiel einer technischen Zeichnung aus der industriellen Praxis, deren Informationsgehalt in die drei Kategorien der organisatorischen, geometrischen und technischen Daten gruppiert wird

5.4.4.4. Extraktion und Repräsentation von Zeichnungsdaten

Wie die Analyse des Informationsgehalts technischer Zeichnungen verdeutlicht hat, ist auch dieser Dokumenttyp durch Daten zu unterschiedlichsten Artefakten charakterisiert. Folglich sollten auch hier zunächst die adressierten Artefakte identifiziert, als **IndexableArtifact**-Objekte erzeugt und mit den jeweils zugehörigen Beziehungsfacetten versehen werden. Wie Abbildung 5.28 zeigt, können somit neben der Polyrepräsentation für das Dokument selbst (Nummer 1) mindestens auch Polyrepräsentationen

für das Produkt (Nummer 2), für Personen (Nummer 3) und für den Werkstoff (Nummer 4) erstellt und mit den jeweils zugehörigen Attributfacetten erweitert werden. Zudem erlauben technische Zeichnungen aufgrund der in ihnen enthaltenen geometrischen Produktinformationen auch eine geometrische Beschreibung des Produktes. Diese ermöglicht als Ausprägung einer Ähnlichkeitsfacette einen Vergleich zweidimensionaler Geometriedaten und stellt daher sowohl für das Produkt als auch für das Dokument ein wichtiges Suchkriterium dar. Auf anwendbare Konzepte zur Generierung einer derartigen Beschreibung sowie auf die Evaluationsergebnisse ausgewählter Konzepte wird in Kapitel 5.4.4.5 noch ausführlich eingegangen. Ferner sollte insbesondere für den Zweck der Vorschaufunktion eine Thumbnailfacette für beide Artefakte (Produkt und Dokument) erzeugt werden. Wie erläutert, können derartige Daten in einer DXF-Datei in einem eigens hierfür definierten Dateiabchnitt zur Verfügung gestellt werden, der für diesen Zweck ausgelesen werden könnte. Allerdings wird diese Möglichkeit in der industriellen Praxis oftmals nicht genutzt, so dass ein Vorschaubild in Form eines Pixelbildes gesondert zu erstellen ist.

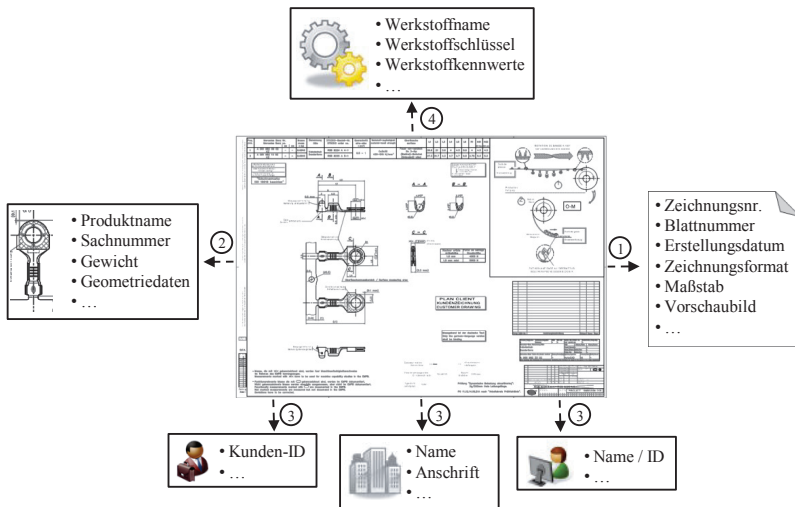


Abbildung 5.28. — In einer technischen Zeichnung adressierte Artefakte mit einigen ihrer zugehörigen Daten

Prinzipiell stehen für die Extraktion der zur Bildung der Artefakte und ihrer Facetten benötigten Informationen aus einer DXF-Zeichnung kostenfreie Java-Softwarebibliotheken zur Verfügung. Beispiele hierfür sind eine von einem Freiberufler namens Rammi entwickelte und in der prototypischen Implementierung verwendete DXF-Viewer-

Bibliothek¹⁰⁸ oder die bei Sourceforge.net erhältliche Kabeja-Bibliothek¹⁰⁹. Da die Informationsextraktion jedoch immer auch von der Art und Gestaltung der zu verarbeitenden Zeichnungen abhängt, ist eine unternehmensspezifische Realisierung der Extraktorkomponente notwendig. Dabei sind insbesondere zwei Probleme zu berücksichtigen, die im Folgenden etwas näher erläutert werden.

Reading HEADER... \$ACADVER=AC1015 \$DWGCODEPAGE=ANSI_1252 Reading CLASSES... Reading TABLES... Reading BLOCKS... Reading ENTITIES... Reading OBJECTS... End of file. === 104088-0124_0125_C.dxf === ===== ARC : 143 CIRCLE : 6 DIMENSION : 29 ELLIPSE : 44 HATCH : 2 INSERT : 45 LEADER : 7 LINE : 288 LWPOLYLINE : 44 MTEXT : 7 SPLINE : 9 TEXT : 8 Gesamt : 632 ===== DXF-Entitytyp : LWPOLYLINE Start line : 270384 Entity reference : 7628 Layer : 0 DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Scale of linetype : 1.0	Typ : Closed Line Nr of Points : 4 Position Point 1 : [572.9465, 273.0905] Position Point 2 : [831.0, 273.0905] Position Point 3 : [831.0, 586.0] Position Point 4 : [572.9465, 586.0] ===== DXF-Entitytyp : INSERT Start line : 270398 Entity reference : 7629 Layer : 0 DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Scale of linetype : 1.0 BLOCK-Name : ASC7F4D0C4E Insert position : [583.0277, 576.7610, 0.0] X-Scale : 1.5 Y-Scale : 1.5 Z-Scale : 1.5 Angle : 0.0 Nr of Lines : 1 Nr of Columns : 1 ===== ATTRIB : 2 BLOCK : 1 Gesamt : 3 ===== DXF-Entitytyp : BLOCK Start line : 174376 Entity reference : 3FC8	Layer : 0 DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Blockname : ASC7F4D0C4E External reference : Reference Point : [0.0, 0.0, 0.0] Switch : 2 ===== ARC : 102 ATTDEF : 2 CIRCLE : 18 HATCH : 1 INSERT : 3 LINE : 359 LWPOLYLINE : 176 POLYLINE : 47 SPLINE : 4 TEXT : 17 Gesamt : 729 ===== DXF-Entitytyp : ARC Start line : 174402 Entity reference : 3FC9 Layer : TRAITFIN DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Scale of linetype : 1.0 ===== Center : [99.4575, -102.2113, 0.0] Radius : 41.4218 Start angle : 184.2479 End angle : 215.2192 =====	DXF-Entitytyp : ARC Start line : 174428 Entity reference : 3FCA Layer : TRAITFIN DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Scale of linetype : 1.0 Center : [98.0802, -103.6033, 0.0] Radius : 40.2402 Start angle : 178.0175 End angle : 220.3958 ===== DXF-Entitytyp : TEXT Start line : 174454 Entity reference : 3FCB Layer : COTES DXF-Linetype : BYLAYER DXF-Colorindex : BYLAYER Space : Model space Visibility : VISIBLE Scale of linetype : 1.0 Text : Outil de pose Position : [105.6937, -135.9725, 0.0] Text height : 1.0 Aspect ratio : 178.0175 STYLE name : ITALIC8 Horizontal alignment : Left Vertical alignment : Baseline ===== DXF-Entitytyp : LINE Start line : 174480 Entity reference : 3FCC Layer : TRAITFIN
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Abbildung 5.29. — Beispiele für DXF-Objekte mit ihren jeweiligen Informationen, die aus der zu der in Abbildung 5.27 dargestellten technischen Zeichnung gehörenden DXF-Datei extrahiert werden können

Unstrukturiertheit der Daten Als primäres Kommunikationsformat ist DXF grundsätzlich so gestaltet, dass es sich nur für die Präsentation bzw. Darstellung von Zeichnungsdaten eignet. Demzufolge ist eine DXF-Datei gemäß Rudolph lediglich als ein Container zu betrachten, der beliebige Objekte ohne jeglichen Zusammenhang enthält [Rud00]. Dies spiegelte sich auch bei einer detaillierteren Analyse industrieller DXF-Testdaten wider. Zur Veranschaulichung zeigt Abbildung 5.29 einen Auszug der nacheinander ausgelesenen Objekte für die in Abbildung 5.27 dargestellte Zeichnung. Wie hier zu erkennen ist, wird jedes Objekt zwar durch seinen zugehörigen, anhand des Gruppencodes identifizierbaren Entitätentyp (LWPOLYLINE, BLOCK, ARC, TEXT, ...), seine Position auf der Zeichnung sowie eventuell weitere zugehörige geometrische oder textuelle Werte beschrieben (vgl. blau umrandete Informationen in Abbildung 5.29). Die

¹⁰⁸<http://www.caff.de/dxfviewer>

¹⁰⁹<http://kabeja.sourceforge.net>

tatsächliche Semantik hinter dem Objekt jedoch bleibt verborgen. Um diesen Sachverhalt zu verdeutlichen, stellen die in einer Zeichnung enthaltenen Texte ein gutes Beispiel dar. Diese werden in DXF mit Hilfe des `TEXT`- oder `MTEXT`-Entitätentyps definiert, welche über das Attribut `Text` verfügen. Obwohl das in Abbildung 5.29 grün markierte `TEXT`-Objekt die Extraktion des zugehörigen Textes (hier: `Outil de pose`) möglich macht, wird die Bedeutung dieses Textelementes nicht deutlich. Folglich kann man so zwar beispielsweise die Zeichenkette `A 000 982 33 02` auslesen; man weiß aber nicht, dass es sich hierbei um die Sachnummer des dargestellten Produktes handelt. Diese Information wiederum steht in einem separaten `TEXT`-Objekt, zu dem keine explizite Beziehung besteht. Diesbezüglich ist jedoch zu beachten, dass die beiden `TEXT`-Objekte auf der Zeichnung üblicherweise nahe beieinander positioniert sind, indem in den meisten Fällen die Zeichenkette für die Bedeutung in einem angrenzenden Schriftfeld oder sogar im selben Schriftfeld zu finden ist (vgl. Abbildung 5.30). Folglich kann die räumliche Nähe zwischen zwei `TEXT`-Objekten in der Zeichnung genutzt werden, um zu einer semantikbeschreibenden Zeichenkette die jeweils wertbeschreibende Zeichenkette ausfindig zu machen. Dazu sind unter Nutzung der Koordinatenangaben der einzelnen `TEXT`-Objekte die Distanzen zwischen ihnen zu bestimmen. Da sich aber durchaus mehrere `TEXT`-Objekte in der Nähe, d. h. in geringer Distanz zum betrachteten `TEXT`-Objekt befinden können, muss zusätzlich noch ein Pattern Matching durchgeführt werden, das sicherstellt, dass es sich bei dem gefundenen `TEXT`-Objekt bzw. seiner zugehörigen Zeichenkette in diesem Fall auch tatsächlich um die Sachnummer handelt. Folglich werden hierzu neben dem Synset für die semantikbeschreibenden Zeichenketten auch entsprechende reguläre Ausdrücke für die wertbeschreibenden Zeichenketten benötigt. Aus diesem Grund sollte eine Erweiterung des Facettenschemas dahingehend erfolgen, dass für jede Facette zusätzlich ein `<pattern/>`-Element zur Definition dieser regulären Ausdrücke angegeben werden kann.

LT.Aend./supp.rev. 3	Lieferant/supplier STOCKO	ZGS —	ED-KB —	Auftr.-Nr./order no. YAP 25361/06		federf.Abt./resp.dep. —	
Benennung/title Kabelschuh Sonderform				CAD	Datum/date	Name/name	
				Pruef./check	—	—	
				Norm/stand.	siehe/see	—	
				Freig./rel.	siehe/see	—	
Sach-Nr./basic number		Masse/mass(kg)	System/system		Format/size	Blatt/sheet	
A 000 982 33 02		—	AutoCAD		A1	—	
Keine Änderung ohne Zustimmung der federführenden Konstruktion./Any alterations are subject to the approval of the design department.							

Abbildung 5.30. — Exemplarisches Schriftfeld einer technischen Zeichnung

Fehlende Layer-Nutzung Wie bei Texten ist auch bei graphischen Objekten die eigentliche Bedeutung allein aus dem Dateinhalt nicht ersichtlich. Das bedeutet,

dass zwar die mit dem Objekt verbundene Semantik im weiteren Sinne bekannt ist, da der Entitätentyp den Zweck eines Objektes eindeutig definiert. Es lässt sich aber nicht erkennen, ob beispielsweise eine Linie Teil einer Schnittdarstellung, einer Ansicht (bzw. welcher Ansicht) oder vielleicht des Schriftfeldes ist. Diese Problematik der Zuordnung einzelner Objekte zu bestimmten Aspekten lässt sich theoretisch durch die Anwendung der in Kapitel 5.4.4.2 erläuterten Layertechnik vermeiden. Da das DXF-Format die Definition beliebig vieler Ebenen erlaubt, könnten für jeden einzelnen Aspekt eine oder eventuell auch mehrere Ebenen angelegt werden. Würde man folglich für jede einzelne Ansicht des Produktes sowie die Bemaßungen eine separate Ebene erzeugen, wäre eine Extraktion der zu einer bestimmten Ansicht gehörenden graphischen Objekte über eine Selektion anhand des für jedes Objekt verfügbaren Attributs **Layer** möglich (vgl. rot umrandetes Attribut in Abbildung 5.29). Allerdings hat die Analyse industrieller Testdaten gezeigt, dass diese Option in der Praxis nur sehr selten genutzt wird. In den meisten Fällen wird anstatt mehrerer Ebenen nur genau eine Ebene, die alle Informationen beinhaltet, definiert. Damit wird eine automatische Zuordnung bzw. Erkennung der einzelnen Produktansichten erschwert. Diese jedoch ist die Voraussetzung für die Realisierung eines Ähnlichkeitsvergleiches von 2D-Geometriedaten, da immer nur jeweils zwei Ansichten miteinander verglichen werden können. Diese beiden Ansichten müssen zudem noch im Hinblick auf den das Produkt beschreibenden Blickwinkel übereinstimmen, um vordefinierte Unterschiede ausschließen zu können. Schließlich würde ein Vergleich der Vorderansicht eines Produktes X mit der Seitenansicht eines Produktes Y zwangsläufig zu Unterschieden und damit zu schlechten Suchergebnissen führen. Demnach würde eine Definition der Ansichten auf verschiedenen Ebenen bereits während der Zeichnungserstellung insbesondere hierfür ein sinnvolles Vorgehen darstellen. Bei allen anderen Zeichnungen, bei denen dies nicht gegeben ist, müsste eine Segmentierung erfolgen, die als Ergebnis die einzelnen Ansichten liefert. Dazu sind aus der Menge der graphischen Elemente in einem ersten Schritt sämtliche Linien, Kreisbögen usw. zu identifizieren und zu eliminieren, die entweder dem Zweck der Dokumentgestaltung (z. B. Schriftfeld, Begrenzung Zeichnungsfläche usw.) oder zur Veranschaulichung der Maßangaben verwendet werden. Für mögliche Vorgehensweisen hierzu wird auf den Forschungsbereich zur Erkennung graphischer Objekte, insbesondere auf die Arbeiten von Ablameyko verwiesen, welcher sich in [Abl97] u. a. mit Methoden speziell für technische Zeichnungen beschäftigt. Anschließend muss im zweiten Schritt eine Zuordnung der verbleibenden graphischen Objekte zu den einzelnen Ansichten des Produktes erfolgen. Dies könnte beispielsweise durch ein Verfahren zur Polygonerkennung, wie es von Fonseca et al. in [FFJ05] beschrieben wird, erreicht werden, welches aus einer Menge von Liniensegmenten schließlich die die Ansichten bildenden Polygone identifiziert.

Ausgehend von diesen beiden Problemen ist eine Extraktorkomponente für technische Zeichnungen im DXF-Format demzufolge nach dem in Abbildung 5.31 dargestellten Aufbau zu realisieren. Dabei sind grundsätzlich zwei Prozesse zu unterstützen, welche parallel ausgeführt werden können. Während sich der eine Prozess mit der Identifikation der einzelnen Zeichnungsansichten durch das Entfernen von Layout- und Dimensionselementen sowie das Zerlegen der Geometrieelemente in die einzelnen Ansichten beschäftigt (vgl. Abbildung 5.31 links), ist der zweite Prozess für das Auslesen der Metadaten in Form der in der DXF-Datei enthaltenen **TEXT**-Objekte und deren Zuordnung zueinander zuständig (vgl. Abbildung 5.31 rechts). Auf Basis der dabei erzeugten Informationen können somit in beiden Prozessen entsprechende Facetten erzeugt und abhängig von der Anzahl der in der Zeichnung adressierten Artefakte zu einer oder mehreren Polyrepräsentationen zusammengefasst werden.

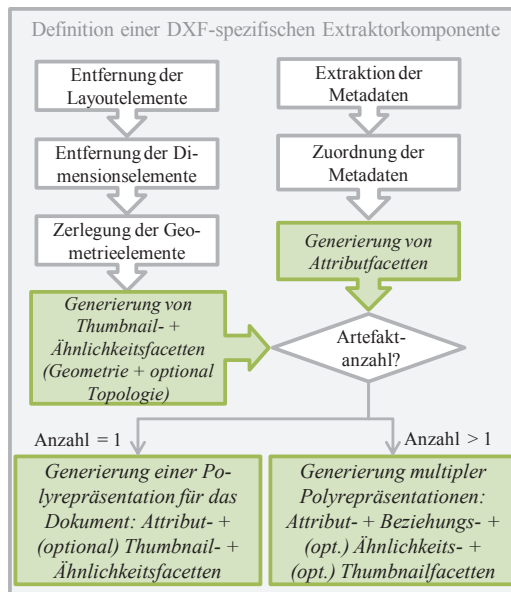


Abbildung 5.31. — Aufbau einer Extraktorkomponente für technische Zeichnungen im DXF-Format

5.4.4.5. Konzepte für einen 2D / 2D-Vergleich und Evaluationsergebnisse implementierter Ansätze

Obwohl sich der Forschungsbereich des Content Based Image Retrieval primär auf die inhaltsbasierte Suche nach Rastergraphiken auf Basis von Farben, Texturen oder Umrissen fokussiert [RHC97, DLW05], wurden hier auch einige wenige Konzepte für den Vergleich von technischen Zeichnungen als ein Spezialfall existierender Bilder entwickelt. Diese lassen sich abhängig davon, welches Format für eine Zeichnung bzw. ein Bild betrachtet wurde, in die beiden Kategorien der auf Rastergraphiken und Vektorgraphiken basierenden Ansätze unterteilen. In den folgenden zwei Abschnitten werden diese Konzepte zunächst anhand eines kurzen Überblicks vorgestellt, bevor im dritten Abschnitt auf Basis von Evaluationsergebnissen eine Empfehlung für die Verwendung eines der untersuchten Konzepte gegeben wird.

Rastergraphik-basierte Konzepte

Aufgrund der Tatsache, dass manuell erstellte Zeichnungen im Zuge der zunehmenden Digitalisierung eingescannt wurden, gingen die ersten entwickelten Konzepte zum Vergleich technischer Zeichnungen von der Existenz einer binären Rastergraphik aus, die lediglich aus schwarzen und weißen Bildpunkten (Pixeln) besteht. Gemäß Fränti et al. dienen die schwarzen Bildpunkte dabei der Darstellung der in der Zeichnung enthaltenen Liniensegmente, so dass letztere zunächst ermittelt werden müssen [FMKK00]. Dazu wird jeder schwarze Pixel (x, y) in den Parameterbereich überführt, indem dieser durch die Geradengleichung $d = x \cdot \cos \theta + y \cdot \sin \theta$ beschrieben wird. Dabei ist d die Distanz zwischen Ursprung und Gerade und θ der Winkel zwischen der x -Achse und der Normalen der Gerade. Beide Parameter können anschließend zur Bildung einer Matrix genutzt werden, deren Reihen den Distanzwerten und deren Spalten den Winkelwerten entsprechen. Durch Einordnung eines jeden schwarzen Bildpunktes in diese Matrix erhält man schließlich eine globale Beschreibung des Bildes auf Basis seiner Liniensegmente. Wird auf diese die sogenannte Hough-Transformation angewendet, lässt sich für jedes Bild ein FV erzeugen, der mit Hilfe eines Distanzmaßes mit anderen FV verglichen werden kann.

Im Gegensatz dazu beschreiben Tabbone et al. ein graphisches Objekt, d. h. eine Produktansicht, mit Hilfe sogenannter F-Signaturen [TWT03]. Unter einer F-Signatur ist dabei ein spezielles Kräftehistogramm zu verstehen, das durch die Ermittlung aller Kräfte, die zwischen den Pixeln eines Objektes herrschen, erzeugt wird. Dazu wird eine Funktion $\varphi_r(d) = 1/d^r$ definiert, deren Parameter d den Abstand zwischen zwei Pixeln des Objektes beschreibt. Abhängig vom Parameter r können anschließend unterschiedliche Kräfte, wie beispielsweise die Anziehungskraft ($r = 1$), zwischen allen Pixelpaaren ermittelt und in ein das Objekt beschreibende Histogramm eingeordnet werden.

Die Identifikation und Beschreibung graphischer Objekte ist auch Bestandteil des Konzeptes von Nabil et al., wobei hier im Gegensatz zum vorhergehenden Konzept die Annahme zugrunde gelegt wird, dass eine Zeichnung aus mehreren räumlich zueinander angeordneten graphischen Objekten besteht [NNS96]. Folglich können diese räumlichen Beziehungen genutzt werden, um zwischen jeweils zwei Objekten der Zeichnung eine als 2D Projection Interval Relationship (2D-PIR) bezeichnete Signatur zu erstellen. Basierend auf der von Chang et al. entwickelten Idee der 2D-String-Beschreibung von Bildobjekten [CSY87] und des Konzeptes der zeitlichen Intervalle nach Allen [All83], besteht diese Signatur aus einem symbolischen Tripel der Form $(dt, d, >)$. Dieses Tripel dient zur Beschreibung der Beziehungen zwischen zwei Objekten, wobei der erste Parameter die räumliche Beziehung (hier dt für disjunkt) und die beiden anderen Parameter die zeitlichen Beziehungen angeben. Für letztere sind die Objekte sowohl auf die x - als auch auf die y -Achse zu projizieren, so dass Aussagen darüber gemacht werden können, ob ein Objekt beispielsweise vor, nach (Symbol $>$) oder während (Symbol d) des anderen Objektes erscheint. Sind für alle Objektpaarungen 2D-PIR-Signaturen erstellt, kann ein gerichteter Graph mit den Objekten als Knoten und den Signaturen als Kanten erstellt werden, welcher letztendlich für einen Graphvergleich herangezogen werden kann.

Vektorgraphik-basierte Ansätze

Während auf Rastergraphiken basierende Methoden eher eine informationsarme Darstellung einer Zeichnung verwenden, versuchen die im Folgenden betrachteten Konzepte den tatsächlichen Inhalt, d. h. Text und v. a. Geometriedaten, aus einer Vektorgraphik zu erfassen. Ein Beispiel hierfür ist die von Love und Barton entwickelte und im kommerziellen CAD-IR-System CADFind sketch & search (vgl. Tabelle 2.1 in Kapitel 2.3) integrierte Repräsentation mit Hilfe sogenannter GT Codes [LB04]. Ein GT Code kann dabei in gewisser Weise als ein Klassifikationsschlüssel, wie er beispielsweise bei dem in Kapitel 2.1.2 erläuterten Opitz-Klassifikationssystem entsteht, gesehen werden; allerdings werden hier sämtliche geometrischen und fertigungsbezogenen Informationen eines Produktes zur Erstellung einer alphanumerischen Zeichenkette verwendet. Als Ergebnis eines nicht näher von den Autoren erläuterten Codierungsprozesses wird dieser Code für den Ähnlichkeitsvergleich von Produkten herangezogen.

Ein völlig anderes Vorgehen dieser inhaltsbasierten Suche schlagen Park und Um in [PU99] vor. Sie legen ihrem Konzept die Annahme zugrunde, dass sich jedes komplexe graphische 2D-Objekt rekursiv in dominante Geometrien und Hilfsgeometrien zerlegen lässt. Dabei verstehen sie unter dominanten Geometrien besonders auffallende Polygone, die von aufeinanderfolgenden Liniensegmenten gebildet werden und beispielsweise die Ansichten eines Produktes beschreiben. Weiterhin gehen sie von der Annahme aus, dass jedes dieser Polygone wiederum in eine Menge von vordefinierten

Primitiven wie beispielsweise Rechtecke oder Dreiecke unterteilt werden kann. Folglich ist es möglich, ausgehend von der dominanten Geometrie jede beliebige Form zu beschreiben, indem dieser Hilfsgeometrien zugefügt oder von ihr entfernt werden. Abbildung 5.32 a) zeigt dieses Vorgehen am Beispiel eines Objektes, das sich aus dem dominanten Rechteck R1 bilden lässt, indem man diesem ein kleineres Rechteck R2 hinzufügt sowie das Rechteck R3 und das Dreieck T1 aus diesem entfernt. Die auf diese Weise beschriebenen Objekte werden anschließend unter Nutzung von topologischen Informationen (speziell Nachbarschafts- und Inklusionsbeziehungen) in einem Topologiegraph, wie es Abbildung 5.32 unter b) verdeutlicht, dargestellt.

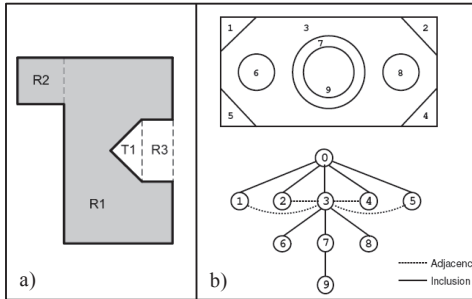


Abbildung 5.32. — a) Beschreibung eines komplexen Objektes anhand einer dominanten Geometrie (R1) und drei Hilfsgeometrien (R2, R3 und T1); b) Abbildung eines Objektes als Topologiegraph (nach [FFJ05, S. 9])

Da ein Graphvergleich allerdings nicht die effizienteste Lösung zur Suche nach ähnlichen Objekten darstellt, griffen Fonseca et al. diese Idee der dominanten Geometrien auf und erweiterten sie [FFJ05]. Anstelle eines direkten Vergleichs der nach der oben beschriebenen Methode erzeugten Graphen, stützten sie ihre Optimierung auf die spektrale Graphentheorie. Diese ermöglicht eine Beschreibung von „(...) topologische[n] Eigenschaften eines Graphen durch Spektraleigenschaften des diskreten Laplace-Operators auf dem Graphen (...)“ [Mül09, 1.Absatz], so dass unter Nutzung der Eigenwerte der Adjazenzmatrix ein FV erstellt werden kann. Dessen Dimension hängt jedoch von der Komplexität des Graphen ab, so dass ein Ähnlichkeitsmaß zum Vergleich derartiger FV in der Lage sein muss, mit unterschiedlichen Dimensionen umzugehen. Folglich kommen hierfür prinzipiell die von Rubner et al. entwickelte Earth Mover’s Distanz [RTG00] oder der von Fonseca et al. entwickelte NB-Baum (ein normierter B^+ -Baum) [Fon04] in Frage. Zusätzlich zu dieser auf Topologieinformationen der graphischen Objekte basierenden Beschreibung einer Zeichnung, schlägt Fonseca in seiner Dissertation eine Beschreibung auf Basis von Geometriedaten vor [Fon04]. Dazu werden für die identifizierten Polygone verschiedene geometrische Kennzahlen

wie beispielsweise das Verhältnis von Umfang zu Fläche der das Polygon umgebenden konvexen Hülle oder das Verhältnis von Höhe zu Breite des umgebenden Rechtecks berechnet, so dass aus diesen Kennzahlen ein FV fester Dimension gebildet werden kann.

Ein anderes, ebenfalls graphbasiertes Konzept zum Vergleich technischer Zeichnungen stammt von Liu et al. und nutzt sogenannte Attributgraphen zur Repräsentation technischer Zeichnungen [LBM04]. Dabei werden wichtige Primitive der Zeichnung, wie Linien oder Kurven, als Knoten des Graphen betrachtet, die aufgrund ihrer Positionsangaben über Kanten verbunden sind. Sowohl Knoten als auch Kanten werden anschließend mit Attributen näher charakterisiert, wobei Attribute eines Knotens dessen graphische Darstellung (kreisförmig, geradlinig, ...) und Attribute einer Kante die topologische Beziehung zwischen zwei Knoten (parallel, schneidend, ...) näher beschreiben. Ein Vergleich derartiger Graphen erfolgt schließlich durch die Ermittlung sowohl der Knoten- als auch der Kantenähnlichkeit.

Schließlich wurde die ursprünglich von Osada et al. für 3D-Modelle entwickelte histogrammbasierte Methode der Shape Distributions (vgl. Kapitel 5.4.2.5) unter der Bezeichnung 2D Shape Histograms von Pu und Ramani auf 2D-Modelle übertragen [PR05, PR06]. Der einzige Unterschied im Vorgehen besteht darin, dass die zufälligen Punkte hier nicht auf der Objektoberfläche, sondern auf den eine Zeichnung bildenden Liniensegmenten erzeugt werden. Prinzipiell ist außerdem die Rekonstruktion eines 3D-Modells aus den 2D-Ansichten eines Produktes möglich, wodurch ein 2D/3D-Ähnlichkeitsvergleich durchführbar wäre. Demzufolge könnten theoretisch zwei Ähnlichkeitsfacetten gebildet werden, von denen allerdings diejenige, die einen 2D/2D-Vergleich ermöglicht, die für die industrielle Praxis relevantere der beiden Facetten darstellt. Aus diesem Grund wird dieser 2D/2D-Vergleich im Weiteren näher untersucht, indem von den oben erläuterten Konzepten zwei Konzepte — ein Konzept auf Basis von Rastergraphiken und eines auf Basis von Vektorgraphiken — ausgewählt und evaluiert wurden.

Evaluation implementierter Konzepte

Aufgrund der im vorhergehenden Kapitel 5.4.4.4 erläuterten Zuordnungsproblematik von graphischen DXF-Objekten zu Produktansichten wurde für die Realisierung eines geometrischen Ähnlichkeitsvergleiches die Annahme zugrunde gelegt, dass die einzelnen Ansichten eines Produktes separat verfügbar sind. Folglich wurde die Evaluation ausgewählter Konzepte auf einer Menge von 140 einfachen technischen Zeichnungen im DXF-Format ausgeführt. Diese sind, wie die beiden Beispiele in Abbildung 5.33 verdeutlichen, inhaltlich auf eine Ansicht beschränkt.

Für die Evaluation selbst wurden aufgrund unvollständiger Konzeptbeschreibungen sowie der Ineffizienz graphbasierter Ansätze das auf Rastergraphiken basierende Kon-

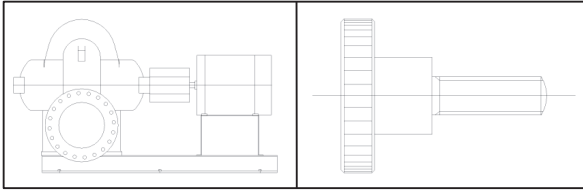


Abbildung 5.33. — Beispiele für die zur Evaluation der Konzepte verwendeten einfachen technischen Zeichnungen

zept von Fränti et al. sowie der von Vektorgraphiken ausgehende Ansatz von Fonseca et al. ausgewählt. Letzterer kann dabei wiederum in die beiden Varianten der rein geometrischen Beschreibung und der topologiebasierten Beschreibung differenziert werden, so dass schließlich drei Vorgehensweisen verglichen wurden. Diese wurden zunächst einzeln betrachtet, da jedes Konzept über mindestens einen Parameter verfügt, der in der Literatur nicht eindeutig spezifiziert wird. So ist beispielsweise für die Polygonerkennung im Rahmen des von Fonseca et al. entwickelten geometriebasierten Konzeptes der von Haperin und Pecker entwickelte Iterated Snap Rounding-Algorithmus [HP02] anzuwenden, der über einen Parameter zur Definition der Größe von Gitterzellen verfügt. Da für diesen Parameter in der Literatur kein spezieller Wert definiert wird, erfolgte für diesen eine Untersuchung der beiden Werte 5 und 10. Dabei zeigte sich, dass sich in allen durchgeführten Tests unter Nutzung der euklidischen Distanz bei Verwendung des Wertes 5 eine bessere Ergebnisqualität erzielen lässt. Folglich wurde diese Einstellung für den Gesamtvergleich verwendet.

Zur Implementierung des topologiebasierten Konzeptes von Fonseca et al. ist einerseits ebenfalls der Parameter für den Iterated Snap Rounding-Algorithmus zu definieren. Im Hinblick auf die Ergebnisqualität spielt andererseits gerade bei diesem Konzept, welches unterschiedliche FV-Dimensionen aufweist, auch das verwendete Ähnlichkeitsmaß eine Rolle. Demzufolge wurden beide Maße, die Earth Mover's Distanz und der NB-Baum, untersucht. Wie das Recall-Precision-Diagramm für die evaluierten Parameterkombinationen in Abbildung 5.34 zeigt, sollten bei Anwendung dieses Konzeptes als Gitterzellengröße der Wert 10 und als Ähnlichkeitsmaß der NB-Baum verwendet werden.

Im Hinblick auf das Konzept von Fränti et al. ist der wichtigste Parameter die Größe des für die technische Zeichnung zu erzeugenden Pixelbildes. Diesbezüglich wurden zwei Bildgrößen mit einem Seitenverhältnis von 1 : 1 (128x128 und 200x200) und zwei Bildgrößen mit einem Seitenverhältnis von 4 : 3 (100x75 und 200x150) evaluiert, wobei die Erzeugung von 128x128 Pixeln bei Vergleich der euklidischen Distanzen die besten Ergebnisse hervorbrachte. Folglich ist dieses mit den beiden, nach obigen Erläuterungen parametrisierten Konzepten zu vergleichen. Gemäß dem hieraus resultierenden,

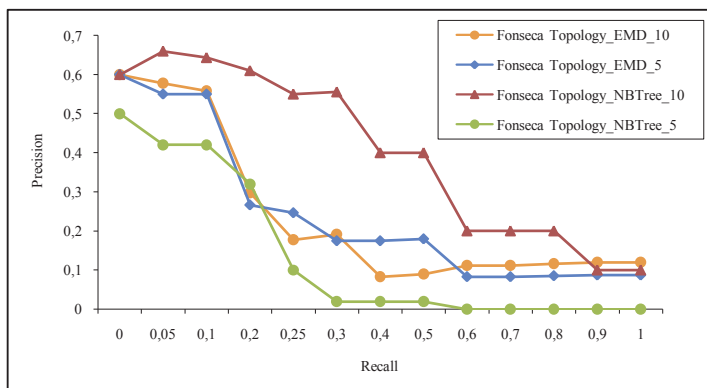


Abbildung 5.34. — Recall-Precision-Diagramm für das topologiebasierte Konzept von Fonseca et al.

in Abbildung 5.35 dargestellten Recall-Precision-Diagramm kann das von Fränti et al. entwickelte Konzept für einen 2D / 2D-Vergleich empfohlen werden. Jedoch bietet das Indexierungsframework, wie bereits in Kapitel 5.4.2.4 erwähnt, selbstverständlich auch für die Repräsentation der Geometriedaten von technischen Zeichnungen die Möglichkeit, andere Konzepte zu verwenden.

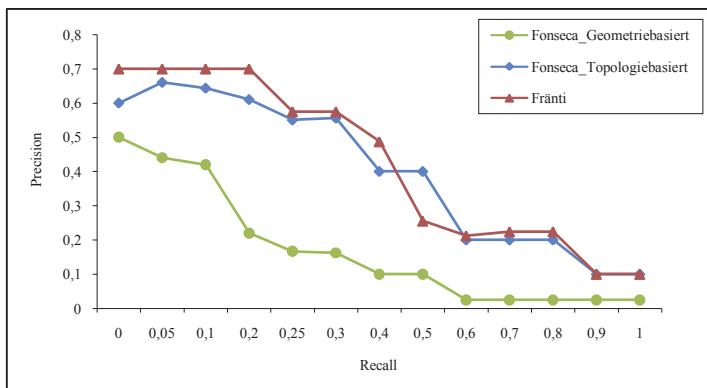


Abbildung 5.35. — Recall-Precision-Diagramm zum Vergleich der Ähnlichkeitskonzepte für technische Zeichnungen unter Berücksichtigung der jeweils besten Einzelergebnisse

5.4.5. Stückliste

Für die Herstellung eines Produktes sind neben technischen Zeichnungen aber auch noch andere Dokumenttypen von Bedeutung. Einer dieser Dokumenttypen ist die Stückliste, deren Zweck darin besteht, die Zuordnung der in einem Produkt enthaltenen Komponenten zueinander, d. h. die Produktstruktur, in kompakter und übersichtlicher Form zu verdeutlichen [ES01]. Demzufolge ist eine Stückliste gemäß Conrad als ein „(...) vollständiges, formal aufgebautes Verzeichnis (...), das alle Teile oder Gruppen mit Angabe von Benennung, Sachnummer, Menge und Einheit enthält“ [Con08b, S. 260] definiert. Somit sind Stücklisten prinzipiell der Gruppe der strukturierten Dokumente zuzuordnen, wobei dieses Verzeichnis abhängig davon, zu welchem Zweck es erstellt wird, unterschiedlich aufgebaut sein kann. Aus diesem Grund werden in der industriellen Praxis neben diversen Sonder- und Mischformen, wie beispielsweise Varianten-, Auswahl- oder Plus-Minus-Stücklisten, drei wesentliche Grundformen von Stücklisten differenziert [PBFG07, Con08b, Wie08]. Die einfachste Form hierbei ist die sogenannte Mengenübersichts- oder Mengenstückliste, in der jede Produktkomponente nur genau einmal und zwar mit der Angabe ihres Gesamtvorkommens im Produkt aufgelistet wird (vgl. Abbildung 5.36 links). Etwas konkreter hingegen ist die Baukastenstückliste, welche allerdings den Aufbau eines Produktes nur bis zur ersten Gliederungsstufe beschreibt (siehe Abbildung 5.36 rechts). Somit erfordert eine vollständige Beschreibung die Generierung mehrerer Baukastenstücklisten — nämlich eine für jede Baugruppe — und deren Zusammenfassung zu einem sogenannten Stücklistensatz. Die schließlich aussagekräftigste, aber bei einer großen Menge von Einzelkomponenten auch umfangreichste Form ist die sogenannte Strukturstückliste (vgl. Abbildung 5.36 Mitte). Sie gliedert jede Komponente durch Traversieren der Produktstruktur von oben nach unten bis zu ihrer höchsten Stufe auf und spiegelt somit den hierarchischen Aufbau eines Produktes aus den einzelnen Komponenten wider.

Was die Erzeugung von Stücklisten betrifft, so erfolgt diese ähnlich zur Generierung technischer Zeichnungen oftmals automatisch. Mit Hilfe von speziellen Stücklistenprogrammen bzw. -modulen lassen sich diese direkt aus dem CAD-System oder auch aus PDM- oder ERP-Systemen ableiten. Dazu werden Stücklisten einerseits in einem einfachen Textdokument mit der Endung .csv für Comma Separated Values (CSV) gespeichert, das jeden Datensatz in einer separaten Zeile beschreibt und dabei zur Trennung der einzelnen Datensatzelemente ein bestimmtes Zeichen — üblicherweise das Komma — verwendet. Zum anderen wird häufig auch das XLS/XLSX-Format zur Archivierung verwendet, da dieses die Struktur eines Produktes gleichzeitig tabellarisch visualisiert und diese dadurch schnell erfassbar macht. Für beide Formate stehen u. a. mit Java CSV¹¹⁰ und Apache POI⁸¹ kostenfreie Softwarebibliotheken zur Extraktion der darin enthaltenen Informationen zur Verfügung.

¹¹⁰<http://sourceforge.net/projects/javacsv>

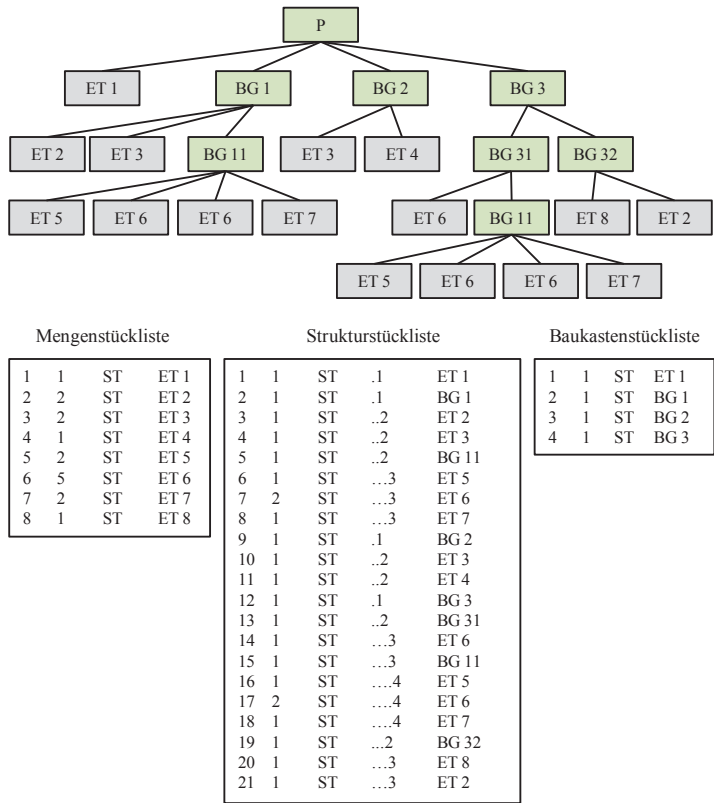


Abbildung 5.36. — Verdeutlichung der drei grundsätzlichen Stücklistenformen anhand der Beschreibung des aus Einzelteilen (grau) und Baugruppen (grün) zusammengesetzten Produktes P (in Anlehnung an [PBFG07, S. 554 - 564])

5.4.5.1. Informationsgehalt von Stücklisten

Da Stücklisten im Wesentlichen Auskunft darüber geben sollen, welche Komponenten in welchen Mengen in einem Produkt verbaut sind, setzen sie sich gemäß Conrad und Pahl et al. grundsätzlich aus einem Schriftfeld und einem tabellenartigen Stücklistenfeld zusammen [PBFG07, Con08b]. Obwohl beide Elemente abhängig von den unternehmensspezifischen Vorgaben sowie vom Verwendungszweck der Stückliste in einem

konkreten Dokument unterschiedlich positioniert und gestaltet sein können, dienen sie zur Übermittlung von drei wesentlichen, für die weitere Verarbeitung des Dokumentes erforderlichen Informationsgruppen. Zwei dieser Informationsgruppen sind dabei prinzipiell im Schriftfeld einer Stückliste vorzufinden. Hierbei handelt es sich zum einen um allgemeine bzw. organisatorische Daten, die der näheren Charakterisierung und Identifikation des Dokumentes dienen [Wie08]. Das bedeutet, dass hier dokumentbezogene Informationen, wie Angaben zum Dokumentautor, zum Erstellungs- und eventuell Änderungsdatum, sowie eventuelle Prüf- oder Bearbeitungsmerkmale für das Dokument eingetragen werden. Zum anderen muss das Schriftfeld sogenannte Bezugsdaten aufweisen, die für die Identifikation des Produktes, dessen Produktstruktur in der Stückliste beschrieben wird, benötigt werden [Wie08]. Die eigentliche Auflistung der die Struktur bildenden Komponenten folgt schließlich im Stücklistenfeld, das die sogenannten Folgedaten beinhaltet [Wie08]. Diese Folgedaten besitzen für jede einzelne Komponente eine eigene Tabellenzeile, die sich üblicherweise aus mindestens vier Angaben zusammen setzt. Neben einer fortlaufenden oder hierarchischen Positionsnummer sind hier eine Sachnummer zur Identifikation sowie eine Mengenangabe mit der für sie gültigen Mengeneinheit anzugeben [Con08b, Wie08].

Wie die Analyse von Testdaten aus der industriellen Praxis gezeigt hat, lassen sich diese drei Informationsgruppen grundsätzlich in jeder Stückliste identifizieren. So auch in dem in Abbildung 5.37 dargestellten Beispiel einer Strukturstückliste, die um zusätzliche Informationen, wie beispielsweise Lieferantendaten, ergänzt ist. Somit kann eine Stückliste abhängig davon, in welchem Unternehmen, zu welchem Zweck und für welches Produkt sie erstellt wurde, neben diesen Basisdaten noch diverse andere Informationen enthalten. Beispiele hierfür sind [Con05, Wie08]:

- die Stücklistenform (Mengen-, Baukasten- oder Strukturstückliste),
- eine Firmenbezeichnung,
- ein Änderungsindex aus Änderungsdatum und Änderungsgrund,
- die Stücklistenverwendung bzw. der Zweck ihrer Erstellung (Fertigung, Einkauf, ...),
- Hinweise auf bisher bzw. nicht mehr gültige Vorläuferstücklisten,
- werkstoffbezogene Angaben,
- Angaben zu Bezugsart (Eigen- oder Fremtteil) und Produktgruppe einer Komponente oder
- auftragsbezogene Daten wie Kostenträger, Auftragsnummer, Terminangaben, Auftragsmenge, Menge je Los, Losnummer (bei Aufteilung in Lose), Auftragsart (Vorrats-, Ersatz-, Betriebs-, Werksauftrag) sowie eventuelle Prüf- oder Liefervorschriften.

Part number	Product name	Product Lifecycle State	Author	Creation date
A2RFFAAACA	Product 1	approved	Engineer1	23/10/2008
Material code	Material density	Product group	Factories	Version
M0000	7.25	Z12CY	0501, 0800	V1

Position	Part number	Product name	Quantity	Product group	Supplier ID
0010	C2RRA22AFC	DS-SMDA22-FLA	1	ST	U24CC
0010	C2RCAR2A22	Socket SMDA22	5	ST	LCLCC
0020	C2RALAAAAC	Kappe SMDA22	20	ST	ECCAA
0030	C2RRC2F8CL	IS-Chip CMDA4RE	7	ST	AAAAA
0010	C2RYYYA8L2	Chip A.W.CMDA4REK	50	ST	AAAAA
0010	C2RAAA2L8A	SI-Plattek CMDA4RE	100	ST	AAAAA
0010	C2RAAA2RCR	SI-Platte CMDA4RE	500	ST	AAAAA
0010	C2RAAA2CRC	SI-Pi.K Rohwafer	5	ST	AY4CC
0010	C2RAAA2CR2	SI-Pi.K Rohwafer	5	ST	AY4CC
0020	C2RAAA2CR4	SI-Pi.K Rohwafer	5	ST	AY4CC
0020	C2RAAA2CRL	SI-Pi.K Rohwafer	5	ST	AY4CC
0020	C2RAAAYAFF	Glasplatte Bearb.	7	ST	LCLC2
0010	C2RAAAYAKR	Glasrohplatte	5	ST	AAAAA
0010	C2RAAAYAKR	Glasrohplatte	5	ST	AAAAA

Abbildung 5.37. — Beispiel einer Strukturstückliste aus der industriellen Praxis mit allgemeinen Daten und Bezugsdaten im Schriftfeld (oben) sowie Folgedaten im Stücklistenfeld (unten)

5.4.5.2. Extraktion und Repräsentation von Stücklisteninformationen

Der im vorhergehenden Kapitel erläuterte Informationsgehalt macht deutlich, dass auch in Stücklisten mehrere Artefakte in einem Dokument adressiert werden. Wie Abbildung 5.38 für die Strukturstückliste aus Abbildung 5.37 zeigt, lassen sich hier Polyrepräsentationen für unterschiedliche Artefakte erstellen. Dazu gehört neben dem durch Attribut- und Beziehungsfacetten beschriebenen Dokument v. a. das Produkt, dessen Produktstruktur in der Stückliste beschrieben wird (im Folgenden auch als Hauptprodukt bezeichnet). Aber auch jede einzelne, die Struktur bildende Komponente sollte als eigenständiges Artefakt repräsentiert werden, indem die in den einzelnen Zeilen des Stücklistenfeldes enthaltenen Daten zur Bildung von Attributfacetten einer zugehörigen Polyrepräsentation genutzt werden. Zudem lassen sich aus den Informationen zur Position der Komponenten in der Produktstruktur, sofern diese wie im Beispiel hierarchisch organisiert sind, automatisch Eltern-Kind-Beziehungen zwischen den Produkten erfassen. Dies wiederum ermöglicht einerseits die automatische Generierung einer Attributfacetten sowohl für das Hauptprodukt als auch für jede einzelne in ihm enthaltene Baugruppe, welche als Ausprägung die Anzahl der jeweiligen Komponenten besitzt. Andererseits lassen sich aus diesen Informationen grundsätzlich zwei Beziehungsfacetten für die Produkte ableiten, von denen die eine die Komponentenverwendung in Form von 'ist enthalten in'-Beziehungen und die andere die Komponentenaufösung im Sinne von 'besteht aus'-Beziehungen beschreibt [ES01]. Weiterhin kann primär für das Hauptprodukt — aber prinzipiell auch für die einzelnen in ihm enthaltenen Baugruppen — eine Ähnlichkeitsfacette gebildet werden, die für einen Vergleich von Produktstrukturen genutzt werden kann. Da mögliche

Repräsentationsformen hierfür bereits in Kapitel 5.4.2.4 erläutert wurden, wird diesbezüglich auf den entsprechenden Abschnitt auf Seite 162 verwiesen.

Allgemein ist zu beachten, dass insbesondere Strukturstücklisten ein und dieselbe Komponente durchaus auch mehrfach enthalten können. Im Beispiel in Abbildung 5.37 wird dies anhand der letzten beiden Positionen deutlich, welche zwar dieselbe Komponente referenzieren, dabei aber zwei unterschiedliche Lieferanten — *Vendor 03* und *Vendor 04* — für diese Komponente angeben. Diese Komponente kann demzufolge von zwei unterschiedlichen Lieferanten bezogen werden, was im Rahmen einer späteren Suche ersichtlich sein muss. Hierfür sind beide Komponenten zunächst als unabhängige Repräsentationen zu erzeugen, welche schließlich in der auf die Repräsentationserstellung folgenden Konsistenzprüfung hinsichtlich ihrer Gleichheit überprüft und bei positivem Prüfergebnis zu einem Indexeintrag kombiniert werden (siehe Kapitel 6).

Zusätzlich zu den Dokument- und Produktrepräsentationen können im Beispiel

- eine Werkstoffrepräsentation mit zwei aus den werkstoffbezogenen Daten erfassbaren Attributfacetten und einer Beziehungsfacette zum Hauptprodukt,
- eine Personenrepräsentation für den Dokumentersteller einschließlich einer Beziehungsfacette zum Dokument sowie
- jeweils eine Personenrepräsentation für jeden Lieferanten mit einer Beziehungsfacette zu dem von ihm gelieferten Produkt

gebildet werden. Zu berücksichtigen ist dabei, dass Beziehungsfacetten grundsätzlich immer in beide Richtungen definiert werden sollten. So impliziert beispielsweise die für den Werkstoff gültige Beziehungsfacette zum Hauptprodukt ('wird verwendet in'-Beziehung) gleichzeitig auch eine für das Hauptprodukt gültige Beziehungsfacette zum Werkstoff im Sinne einer 'besteht aus'-Beziehung (vgl. Beziehungsnetz in Kapitel 4.2).

5.4.6. Zusammenfassung

Auf Basis der zuvor erläuterten Extraktionsmöglichkeiten aus den einzelnen Dokumenttypen lässt sich zusammenfassend festhalten, dass grundsätzlich für alle Dokumente — unabhängig von ihrem jeweiligen Dokumenttyp — neben einer Polyrepräsentation für das Dokument immer auch Polyrepräsentationen für andere Artefakte erstellt werden können und sollten. Damit einhergehend sind nicht nur artefakttypspezifische Attribut-, Ähnlichkeits- und Thumbnailfacetten, sondern auch die Beziehungsfacetten zwischen den Artefakten zu generieren. Allerdings ist zu beachten, dass die Realisierung einer speziellen, für einen Dokumenttyp geeigneten Extraktorkomponente nicht immer trivial ist und mit einem hohen Aufwand verbunden sein kann. Aufgrund der enormen Vielfalt an relevanten Dokumenttypen im Bereich der technischen PE (vgl. Abbildung 4.4 und Anhang E) ist es daher nicht zielführend, jeden dieser

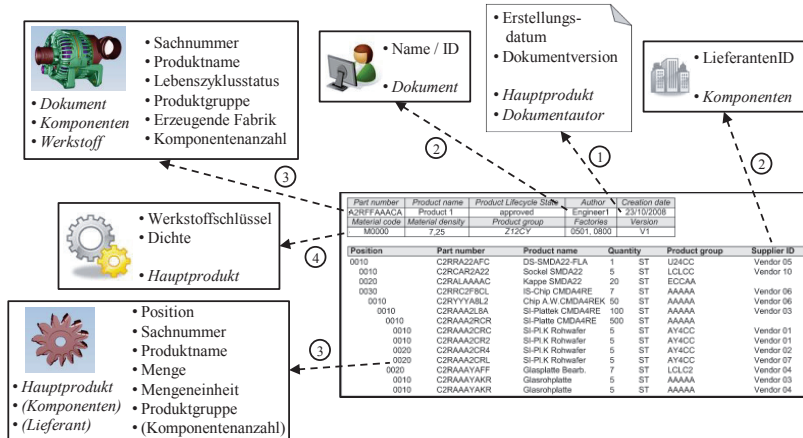


Abbildung 5.38. — Überblick über die aus der in Abbildung 5.37 dargestellten Stückliste extrahierbaren Artefakte mit ihren Attribut- (nicht kursiv) und Beziehungsfacetten (kursiv): Dokument (1), Dokumentautor und Lieferanten als Personen (2), Hauptprodukt und Komponenten als Produkte (3) und Werkstoff (4)

Dokumenttypen so detailliert zu berücksichtigen. Es darf jedoch auch nicht vergessen werden, dass insbesondere unstrukturierte oder semi-strukturierte Dokumenttypen ebenfalls wichtige Informationen enthalten können, die bei einer eventuellen Nichtberücksichtigung dieser Dokumenttypen verloren gehen. Aus diesem Grund wird für die Realisierung des LFRP- und damit auch des Indexierungsframeworks prinzipiell das Einbeziehen von so vielen Dokumenttypen wie möglich empfohlen. Hinsichtlich der Erstellung von Extraktorkomponenten jedoch sollten zwei Arten der Indexierungsmethode unterschieden werden. Letztere ist für jeden Dokumenttyp gesondert festzulegen und schreibt für diesen entweder eine feine oder eine grobe Indexierung vor. Wird beispielsweise für einen Dokumenttyp eine spezielle Extraktorkomponente auf Basis seines Dokumentinhaltes sowie seines Dateiformates benötigt, so ist für diesen die *feine Indexierungsmethode* anzuwenden. Für einen Dokumenttyp, der diesen Detaillierungsgrad nicht erfordert, ist hingegen die *grobe Indexierungsmethode* ausreichend. Indem hier lediglich eine Polyrepräsentation für das Dokument selbst erzeugt und mit aus dokumentspezifischen Metadaten erstellten Attributfacetten, einer optionalen Thumbnailfacette und einer Ähnlichkeitsfacette zur Schlagwortsuche (QbT), welche den gesamten textuellen Inhalt des Dokumentes als Ausprägung besitzt, erweitert wird, kann der erforderliche Aufwand zur Informationsextraktion in Grenzen gehalten und gleichzeitig ein Informationsverlust vermieden werden.

6. Konsistenzprüfung und Indexverwaltung

Nachdem die verschiedenen Artefakte in entsprechende Polyrepräsentationen überführt wurden, sind sie gemäß Abbildung 3.8 in Kapitel 3.3.3 an einen *Index Manager* zu übergeben. Dieser ist grundsätzlich für die Verwaltung des Index sowie das Hinzufügen der einzelnen Polyrepräsentationen zu diesem Index verantwortlich. Zusätzlich muss dieser *Index Manager* eine Überprüfung der erstellten Artefaktbeschreibungen hinsichtlich ihrer Konsistenz vornehmen. Der Grund hierfür resultiert aus der Tatsache, dass bei der Erzeugung von Polyrepräsentationen auf Basis von Informationen aus einem Drittsystem oder aus Dokumenten nicht notwendigerweise eine Eignung bzw. Korrektheit der Polyrepräsentation im Sinne einer für das LFRP-Framework erforderlichen Weise garantiert wird. Schließlich kann es durchaus vorkommen, dass nicht alle benötigten Informationen zu einem Artefakt verfügbar sind oder dass redundante oder auch widersprüchliche Informationen existieren. Folglich ist jede einzelne, von den jeweiligen **Indexern** erstellte Polyrepräsentation auf ihre Konsistenz zu prüfen. Dabei versteht man laut dem Gabler Online-Wirtschaftslexikon unter dem Begriff der Konsistenz bzw. unter dem Begriff des Konsistenzpostulats die „Forderung nach Widerspruchsfreiheit innerhalb der verschiedenen Aussagen eines theoretischen Systems (...)“ [Gab10, Abschnitt Kurzerklärung]. Überträgt man diese Bedeutung auf den vorliegenden Anwendungsfall, so wird hier mit Konsistenz die Widerspruchsfreiheit und Stimmigkeit einer zu indexierenden Polyrepräsentation verbunden. Allerdings kann diese Konsistenz allein durch Anwendung der Methoden im Rahmen der Repräsentationserstellung nicht gewährleistet werden.

Ein Grund dafür ist die Unvollständigkeit von Daten, wodurch Nullwerte (in Java durch die `null`-Referenz gekennzeichnet) entstehen und zu verarbeiten sind. Dabei sind gemäß Elmasri und Navathe Nullwerte grundsätzlich anhand der mit ihnen verbundenen Bedeutung in zwei Gruppen zu differenzieren [EN09]. So kann ein Nullwert dafür stehen, dass ein Attribut bzw. eine Facette für ein Artefakt nicht relevant ist. Da im vorliegenden Anwendungsfall jedoch gerade die Generierung artefaktspezifischer und somit nur relevanter Facetten im Fokus steht, wird diese Gruppe von Nullwerten nicht weiter betrachtet. Daher ist im Rahmen der Indexierung besonderes Augenmerk auf diejenigen Nullwerte zu legen, deren Bedeutung in der Unbekanntheit eines Attributwertes, d. h. einer Facettenausprägung, besteht. Der Begriff der Unbekanntheit

kann dabei entweder das Fehlen eines Wertes, dessen Existenz allerdings bekannt ist, oder aber die Unwissenheit über die Existenz des Wertes ausdrücken. Beide Fälle können im Rahmen der Informationsextraktion auftreten und müssen entsprechend bei einer Konsistenzprüfung — speziell im Hinblick auf obligatorische und fakultative Facetten für die Artefakte — berücksichtigt werden. Demzufolge beschäftigt sich Kapitel 6.1 speziell mit dem Vorgehen beim Vorliegen von unvollständigen Daten.

Ein weiterer Grund ist in der möglichen Existenz multipler Quellen für ein und dieselbe Artefaktinformation zu sehen. Aufgrund der Tatsache, dass im Bereich der technischen PE zum einen diverse IT-Systeme zur Verwaltung der Produktinformationen (v. a. PDM- und ERP-Systeme) eingesetzt werden und dass zum anderen diese Daten noch zusätzlich in unterschiedlichsten Dokumenten enthalten sind, kann eine Dateninkonsistenz durch die Berücksichtigung dieser verschiedenen Informationsquellen entstehen. Gemäß obiger Konsistenzdefinition sind diese Daten folglich widersprüchlich zueinander, obwohl jedes Datum für sich als gültig zu betrachten ist. Zur Verdeutlichung dieses Sachverhaltes kann der Produktname als Beispiel für eine derartige, möglicherweise inkonsistente Artefaktinformation betrachtet werden. Dieser kann üblicherweise aus einem PDMS gewonnen werden, wo er zusammen mit anderen Produktdaten zur möglichst vollständigen Beschreibung eines Produktes verwaltet wird. Zusätzlich wird aber auch häufig in Produktmodellen wie CAD-Modellen, technischen Zeichnungen oder Stücklisten der Name eines Produktes mit angegeben. Beide Produktnamen müssen dabei für ein und dasselbe Produkt nicht notwendigerweise übereinstimmen, weil beispielsweise auf eine detaillierte Angabe im Produktmodell aufgrund der Existenz der eindeutigen Sachnummer verzichtet wird. Damit werden im Rahmen der in Kapitel 5 beschriebenen Informationsextraktion für dasselbe Artefakt zwei Attributfacetten mit `facetId=“productName“` erstellt, die zwei unterschiedliche Ausprägungen (z. B. O-Ring 4711 aus dem PDMS und O-Ring aus dem Dokument) enthalten, obwohl nur eine Ausprägung für diese Facette sinnvoll und laut Facettenschema möglich ist. Folglich sind derartige Inkonsistenzen insbesondere im Rahmen des Hinzufügens der Polyrepräsentationen zum Index aufzudecken und in einen konsistenten Zustand zu überführen. Wie eine derartige Beseitigung von Inkonsistenzen erfolgen kann, wird in Kapitel 6.2 im Detail beschrieben.

6.1. Konsistenzprüfung zur Validierung extrahierter Artefaktbeschreibungen

Um die Konsistenz einer erstellten Polyrepräsentation zu erreichen, ist diese folglich vor dem tatsächlichen Hinzufügen zum Index zunächst auf ihre Gültigkeit für das jeweils zugehörige Artefakt zu prüfen. Dabei ist eine Polyrepräsentation in Anlehnung an das Wohlgeformtheitskriterium von XML-Dokumenten als gültig zu betrachten,

wenn sie den im Facettenschema definierten Vorgaben und Regeln entspricht. Hierzu wird ausgehend von der Superklasse *Artefakt* im Facettenschema (im Folgenden mit FaSe bezeichnet) für jede Subklasse festgestellt, ob die für die jeweilige Subklasse definierten Facetten in der Polyrepräsentation (im Folgenden mit PoRe bezeichnet) enthalten und ihre extrahierten Facettenausprägungen vom richtigen Facettentyp sind. Da zwischen obligatorischen und fakultativen Facetten unterschieden wird, sind diesbezüglich die folgenden sechs, in Pseudocode beschriebenen Szenarios zu betrachten:

1. (Facette == fakultativ) AND (Facettenausprägung == null)
2. (Facette == fakultativ) AND (Facettenausprägung != null) AND (*Facettentyp_{FaSe}* != *Facettentyp_{PoRe}*)
3. (Facette == fakultativ) AND (Facettenausprägung != null) AND (*Facettentyp_{FaSe}* == *Facettentyp_{PoRe}*)
4. (Facette == obligatorisch) AND (Facettenausprägung != null) AND (*Facettentyp_{FaSe}* == *Facettentyp_{PoRe}*)
5. (Facette == obligatorisch) AND (Facettenausprägung != null) AND (*Facettentyp_{FaSe}* != *Facettentyp_{PoRe}*)
6. (Facette == obligatorisch) AND (Facettenausprägung == null)

Da im ersten Szenario die Facette nicht notwendigerweise für ein Artefakt vorhanden sein muss, hat hier das Fehlen der zugehörigen Facettenausprägung keine Auswirkung auf die Funktionalitäten des LFRP-Frameworks. Anstelle die Facette jedoch nicht zu indexieren, empfiehlt sich die Verwendung einer Facettenausprägung, die entsprechend dem jeweiligen Facettentyp der Kategorie **Sonstiges** entspricht. Dadurch lässt sich gemäß Gaus zumindest formal die Forderung nach Vollständigkeit von Klassifikationssystemen, zu denen in gewisser Weise auch die vorgestellte facettierte Suche zu zählen ist, herstellen [Gau05]. Schließlich wird auf diese Weise nicht nur eine einheitliche Beschreibung aller Artefakte desselben Typs vorgenommen, sondern darüber hinaus dem Nutzer angezeigt, dass für manche Artefakte keine konkreten Facettenausprägungen vorhanden sind. Treten Nullwerte hingegen bei obligatorischen Facetten auf, wie es das sechste Szenario verdeutlicht, so kann eine Indexierung des Artefakts nur vorgenommen werden, wenn die tatsächliche Facettenausprägung bekannt ist. Daher sollten in diesem Fall die nicht-indexierbaren Daten für eine manuelle Nachbearbeitung in einer, für den Datenimport geeigneten XML-Datei protokolliert werden, wie dies auch bereits in Kapitel 5.4.1.3 auf Seite 136 beim Vorliegen von für die Indexierung nicht-ausreichenden Informationen erläutert wurde. Prinzipiell könnte man zwar auch hier eine Kategorie **Sonstiges** einführen; allerdings würden dadurch alle weiteren verfügbaren Facetten auf den unteren Ebenen der Artefaktyphierarchie verloren gehen. Zur Verdeutlichung wird der Fall betrachtet, dass die Produktgruppe eines Produktes nicht bekannt und damit nicht in der Polyrepräsentation enthal-

ten ist. Das Fehlen dieser Information führt dazu, dass das Artefakt lediglich der Ebene *Produkt* des Facettenschemas zugeordnet werden kann. Eventuell vorhandene Facetten wie beispielsweise der *Außenradius* sind allerdings auf dieser Ebene nicht zulässig, da sie nicht für alle Produkte gültig sind. Folglich kann diese Information nicht indexiert werden und geht verloren. Um dies zu vermeiden, sollte daher eine manuelle Nachbearbeitung der Daten erfolgen, so dass nur in Ausnahmen, d. h. wenn die Information wirklich nicht bekannt sein sollte, sozusagen als Notfalllösung auf eine Verwendung der Kategorie *Sonstiges* zurückgegriffen wird. Das Protokollieren von nicht-indexierbaren Informationen ist zudem auch in den noch übrigen Szenarios sinnvoll. So ist im zweiten und fünften Szenario bei Nichtübereinstimmung der Facettentypen diese Lösung ebenfalls anzuwenden, sofern keine automatische Transformation des einen Facettentyps in den anderen erfolgen kann¹¹¹. Trotz korrektem Facettentyp (vgl. drittes und viertes Szenario) kann es dennoch vorkommen, dass die extrahierte Facettenausprägung nicht gültig ist. Ein Grund hierfür kann beispielsweise die Existenz mehrerer Ausprägungen sein, obwohl gemäß Facettenschema nur eine Ausprägung für die Facette erlaubt ist. Folglich ist zusätzlich zum Facettentyp auch die Wertigkeit von Facetten zu überprüfen, so dass bei Nichtübereinstimmung auch hier das Protokollieren der Informationen sinnvoll ist, um durch eine manuelle Nachbearbeitung den richtigen Wert festzulegen. Eine Anwendung von Regeln, wie sie in Kapitel 6.2 zur Aktualisierung von Polyrepräsentationen vorgeschlagen werden, ist in diesem Fall nicht möglich, da zu diesem Zeitpunkt der Indexierung die Ausprägungen aus der gleichen Informationsquelle stammen und zur gleichen Zeit extrahiert wurden.

Zusammenfassend ist diese als Validierung bezeichnete Prüfung eine notwendige Voraussetzung dafür, dass nur für den Retrievalprozess geeignete Polyrepräsentationen im darauf folgenden Schritt dem Index hinzugefügt werden. Gleichzeitig ist sie notwendig, um bei erfolgreicher Prüfung die gemäß Kapitel 5.1 erforderliche Facette mit `facetId="artifactHierarchyPath"` für ein Artefakt setzen zu können, um während des Retrievalprozesses die geforderte, dynamische Facettenbereitstellung abhängig vom Artefakttyp und den bereits gewählten Facetten zu gewährleisten.

6.2. Indexverwaltung und Konsistenzprüfung zur Verifikation von Artefaktbeschreibungen

Nach einer erfolgreichen Validierung können die Polyrepräsentationen schließlich dem Index hinzugefügt werden. Dabei ist allerdings zu berücksichtigen, dass die unterschiedlichen Facettenarten auch unterschiedliche Anforderungen an einen Index stel-

¹¹¹Zahlenwerte lassen sich beispielsweise mittels statischer `valueOf()`-Methoden der Java-Klasse `String` in Zeichenketten transformieren (vgl. <http://download.oracle.com/javase/6/docs/api/>, Klasse `String`).

len. Während beispielsweise Attributfacetten eine schnelle Berechnung der für sie notwendigen Anfragevorschauen erfordern, muss bei Ähnlichkeitsfacetten eine Handhabung von unter Umständen sehr hohen Dimensionswerten gewährleistet werden. Somit ist prinzipiell für jede Facettenart ein entsprechend geeigneter Subindex bereitzustellen, der den jeweiligen Anforderungen gerecht wird. So lassen sich Thumbnails auch Beziehungsfacetten aufgrund ihrer relativ geringen Anforderungen beispielsweise unter Nutzung einer einfachen objektorientierten Datenbank indexieren, welche die Verwaltung von POJOs und damit einen einfachen objektbasierten Zugriff auf die im Rahmen des Retrievalprozesses erforderlichen Daten ermöglicht. Im Fall von Ähnlichkeitsfacetten kann eine solche objektorientierte DB zwar prinzipiell ebenfalls zum Einsatz kommen; allerdings lassen sich bei einer Indexierung von nur wenigen hundert CAD-Modellen nur Antwortzeiten von mehr als einer Minute realisieren, weil ein sequentieller Vergleich der Anfrage mit jeder einzelnen Ähnlichkeitsfacette im Index durchgeführt werden muss. Die für dieses lineare Vorgehen benötigte Zeit wächst proportional mit der Anzahl der Ähnlichkeitsfacetten im Index und macht sich direkt als Wartezeit beim Nutzer bemerkbar [Sch06]. Folglich würde die Akzeptanz des Systems bei einem Einsatz in der industriellen Praxis hierdurch deutlich beeinträchtigt werden, so dass auf Indexstrukturen zurückgegriffen werden sollte, die speziell zur Handhabung hoher Dimensionswerte entwickelt wurden. Obwohl die Nutzung einer effizienten Indexstruktur für hochdimensionale Daten als notwendig für den Erfolg des vorgestellten LFRP-Frameworks gesehen wird, wurde die Betrachtung dieses Aspektes in der vorliegenden Arbeit aufgrund der zahlreichen in der Literatur existierenden Veröffentlichungen hierzu explizit ausgeschlossen¹¹². Aus Gründen der Vollständigkeit jedoch werden im Folgenden zwei grundsätzliche Möglichkeiten zur Realisierung effizienter Indexstrukturen kurz aufgezeigt:

Reduktion der Vergleichsoperationen Die erste Möglichkeit basiert auf der Idee, dass so wenige Ähnlichkeitsberechnungen wie möglich zu einer Ergebnisbestimmung benötigt werden. Dies kann durch eine entsprechende Strukturierung der Ähnlichkeitsrepräsentationen im Index gewährleistet werden, bei der idealerweise nur die Teilstrukturen oder -bereiche durchsucht werden, in denen auch tatsächlich ähnliche Objekte vorkommen können. Die verschiedenen Ansätze in der Literatur nehmen hierzu üblicherweise eine topologische Unterteilung des n-dimensionalen Raumes vor, wobei zwischen flach strukturierten Indexstrukturen (z. B. Grid File [NHS84]) und hierarchisch strukturierten Indexstrukturen (z. B. verschiedene Varianten des R-Baumes [MNPT05], M-Baum [CPZ97], LSD-Baum [HSW89] usw.) differenziert wird. Allerdings haben Weber, Schek und Blott anhand ihrer Untersuchungen in [WSB98] gezeigt, dass diese Strukturen bereits ab einer Größe von mehr als 10 Dimensionen ineffizienter sind als lineare Verfahren, so dass letztendlich doch alle Ähnlichkeitsrepräsentationen

¹¹²Für einen Überblick über hochdimensionale Indexstrukturen wird auf die Veröffentlichungen von Samet [Sam06] und Yu [Yu02] verwiesen.

berücksichtigt werden müssen. Aus diesem Grund empfehlen die Autoren die Nutzung von Indexstrukturen, die eine im Anschluss erläuterte Dimensionsreduktion vornehmen.

Reduktion der Dimensionen Die Idee dieses Vorgehens besteht darin, die Effizienz und damit das Laufzeitverhalten durch eine Transformation der Repräsentationen aus dem n -dimensionalen Raum in den eindimensionalen Raum zu verbessern. Dabei wird die Annahme zugrunde gelegt, dass Punkte, die im n -dimensionalen Raum dicht beieinander liegen, dies auch im eindimensionalen Raum tun. Somit wurden diverse Transformationsverfahren entwickelt, zu denen u. a. das iDistance-Verfahren von Yu [Yu02] zu zählen ist. Auch das von Blott und Weber [BW97, WSB98] entwickelte Vector Approximation File nimmt in gewisser Weise eine Dimensionsreduktion vor, die allerdings genutzt wird, um die Anzahl der für die Ergebnisermittlung benötigten Vergleichsoperationen zu reduzieren. Dazu werden sämtliche Ähnlichkeitsrepräsentationen mit Hilfe von Signaturen approximiert. Im Rahmen der Ergebnisermittlung wird schließlich in einem ersten Schritt eine Kandidatenmenge auf Basis dieser Signaturen ermittelt, für die im zweiten Schritt Ähnlichkeitsberechnungen durchgeführt werden. Damit kombiniert das Vector Approximation File die Ideen beider Möglichkeiten und stellt somit einen guten Ansatz für die Realisierung einer Indexstruktur für Ähnlichkeitsfacetten dar.

Im Gegensatz zu Ähnlichkeitsfacetten erfordern Attributfacetten eine schnelle Ermittlung aller für eine Facette gültigen Facettenausprägungen inklusive ihrer Anfragevor-schauen, weshalb für diese Facettenart ein invertierter Index von Vorteil ist. Dieser wurde ursprünglich für den Zweck der Textindexierung entwickelt [BR⁺99, Gau05] und stellt für den vorliegenden Anwendungsfall im Prinzip ein Register dar, das für jede Facettenausprägung die IDs derjenigen Artefakte enthält, die diese Facettenausprägung besitzen. Ein derartiger invertierter Index bildet die Grundlage für das für eine facettenierte Suche kostenfrei zur Verfügung stehende Apache Solr (vgl. Kapitel 3.2.1), weshalb sich letzteres für die Bildung eines Subindex für die Attributfacetten empfiehlt. Hierzu ist lediglich für jedes Artefakt ein `SolrInputDocument`-Objekt zu erstellen, dass die artefaktspezifischen Attributfacetten als Felder in Form von Name-Wert-Paaren enthält. Der einzige Makel bei Verwendung von Apache Solr ist die zusätzliche Notwendigkeit einer Solr-Konfigurationsdatei namens *schema.xml*, welche die für ein `SolrInputDocument` möglichen Felder definiert und folglich mit den im Facettenschema definierten Facetten konform sein muss.

Somit hat der *Index Manager* neben der Konsistenzprüfung gemäß Abbildung 6.1 die Aufgabe die einzelnen Subindizes als Gesamtindex zu verwalten und abhängig von der jeweiligen Facettenart eine Zuordnung der in einer Polyrepräsentation enthaltenen Repräsentationen zu diesen Subindizes vorzunehmen. Dabei sollten alle zu indexierenden Repräsentationen nach dem Entwurfsmuster der Datenzugriffsobjekte

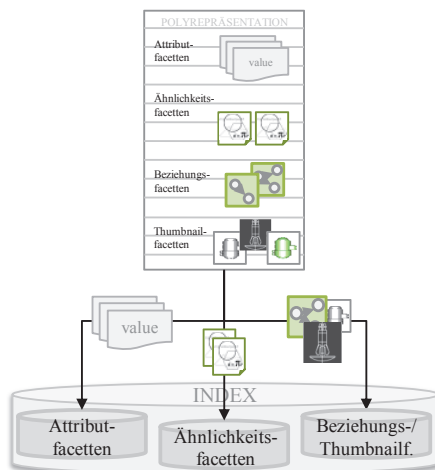


Abbildung 6.1. — Zuordnung der in einer Polyrepräsentation enthaltenen Repräsentationen auf mehrere Subindizes in Abhängigkeit von ihrer Facettenart

verwaltet werden, damit die Suchkomponente diese unabhängig von den tatsächlich zur Bildung des Gesamtindex genutzten Subindizes im Retrievalprozess nutzen kann. Des Weiteren ist beim Hinzufügen von Repräsentationen zu einem Index — unabhängig von ihrer Facettenart und dem jeweiligen Subindex — grundsätzlich immer zu berücksichtigen, dass ein Artefakt bereits im Index enthalten sein kann. Somit muss hier die Aktualität der Daten, die gemäß Tanenbaum und van Maarten ebenfalls einen wichtigen Aspekt der Konsistenz darstellt [Tv08], gewährleistet werden. Dies ist insbesondere dann problematisch, wenn die bereits im Index enthaltenen Facetten zu den zu indexierenden Facetten im Widerspruch stehen. Aus diesem Grund ist beim Hinzufügen eines Artefakts nochmals eine Konsistenzprüfung vorzunehmen, die dieses Mal allerdings der Verifikation der Artefaktbeschreibungen dient. Während somit im Fall der Validierung festgestellt wird, ob eine Polyrepräsentation prinzipiell für das Hinzufügen zum Index geeignet ist, wird im Fall der Verifikation geprüft, welche Daten bei Redundanz oder Widersprüchlichkeit zu verwenden sind. Dazu werden Regeln benötigt, die in Form von *wenn-dann*-Aussagen auf Basis der für jede Artefaktfacette verfügbaren Informationen zu dessen Informationsquelle und Extraktionszeitpunkt bestimmen, welche Information korrekt und damit zu indexieren ist. Diese Regeln sollten bei Initialisierung des LFRP-Frameworks mittels des Spring Frameworks innerhalb der Konfigurationsdatei *applicationContext.xml* dem Modul zur *Konsistenzprüfung* mit übergeben werden, um so eine möglichst hohe Flexibilität hinsichtlich

der Definition und Modifizierbarkeit der Regeln zu ermöglichen. Bei der Definition dieser Regeln ist allerdings genau zu analysieren, was unter Aktualität zu verstehen ist, da auch in diesem Fall verschiedene Szenarios zu differenzieren sind:

- Szenario 1: eine Facette darf nur eine Ausprägung besitzen
 $\Rightarrow \text{multiIndexable} = \text{„false“}$
- Szenario 2: eine Facette kann mehrere Ausprägungen aufweisen
 $\Rightarrow \text{multiIndexable} = \text{„true“}$

Im ersten Szenario ist eine bereits indexierte Facettenausprägung entweder durch einen aktualisierten Wert aus der gleichen Informationsquelle oder durch einen Wert aus einer anderen Informationsquelle zu ersetzen. Damit sind hier zunächst die Informationsquellen der Facettenausprägungen zu vergleichen. Stimmen diese — also die Informationsquelle der bereits indexierten Facettenausprägung und die der noch zu indexierenden — überein, so ist die bereits indexierte Facettenausprägung mit derjenigen zu ersetzen, die später extrahiert wurde und damit das jüngere Extraktionsdatum aufweist. Stimmen die Informationsquellen hingegen nicht überein, so ist mit Hilfe der definierten *wenn-dann*-Aussagen zu bestimmen, welcher Wert aus welcher Informationsquelle als gültig und korrekt zu betrachten ist. Kann beispielsweise eine eindeutige und von allen Beteiligten eingehaltene Charakterisierung von Artefakten im PDMS sichergestellt werden, so könnte man in diesem Fall eine Regel aufstellen, die besagt, dass Werte aus einem PDMS Werten aus einem Dokument oder aus anderen Informationsquellen vorzuziehen sind. Ähnlich gestaltet sich das Vorgehen im zweiten Szenario, das allerdings durch die mögliche Mehrwertigkeit einer Facetten verkompliziert wird. Schließlich kann es in diesem Fall auch möglich sein, dass eine der bereits indexierten Facettenausprägungen nicht mehr gültig und damit aus dem Index zu entfernen ist. Zur Verdeutlichung werden im Folgenden beispielhafte Aktualisierungsszenarios und das mit ihnen verbundene Verifikationsvorgehen aufgezeigt. Da letzteres davon abhängt, aus welchen Informationsquellen die bereits indexierte und die noch zu indexierende Facette stammen, ist eine Fallunterscheidung vorzunehmen, die anhand der Kapitel 6.2.1 und 6.2.2 erläutert wird. Für ein einfacheres Verständnis erfolgt die Beschreibung der diversen Fälle teilweise mit Hilfe mathematischer Gleichungen, deren Symbole die folgende Bedeutung haben:

- IQ ist die Menge der Informationsquellen mit $IQ = \{IQ1, IQ2, IQ3, IQ4, IQ5\}$,
- $F()$ beschreibt die Menge der Facetten mit $F() = \{Fi(), F^*()\}$ und $i = \{1, 2, \dots\}$,
- ai_{IQi} , bi_{IQi} und ci_{IQi} stellen aus der Informationsquelle IQi gewonnene Ausprägungen der Facette $Fi()$ dar,
- $|Fi()|$ beschreibt die Anzahl der Facettenausprägungen der Facette $Fi()$,
- $F^*()$ stellt die Facette mit ihren Ausprägungen ($a_{iQ_i}^*$, $b_{iQ_i}^*$ usw.) dar, die aus der Verifikation resultiert,

- Ed_x bezeichnet das Extraktionsdatum einer Facettenausprägung x ,
- die $>$ -Relation bei Datumswerten beschreibt, dass das als größer eingestufte Datum näher am Zeitpunkt der Konsistenzprüfung liegt und somit aktueller als der kleiner eingestufte Datumswert ist.

6.2.1. Aktualisierungsszenario 1: Identische Informationsquellen

Im ersten Aktualisierungsszenario stammen die für ein Artefakt bereits indexierte Facette $F1(a1_{IQ1}, b1_{IQ1})$ mit den Facettenausprägungen $a1_{IQ1}$ und $b1_{IQ1}$ sowie die für dieses Artefakt neu zu indexierende Facette $F2()$ aus der gleichen Informationsquelle $IQ1$. Da die neu zu indexierende Facette immer aktueller als die bereits indexierte Facette ist, weisen ihre Facettenausprägungen in jedem Fall ein jüngeres Extraktionsdatum auf und sind somit bei einem erforderlichen Vergleich der Datumswerte zu verwenden. Allerdings ist zu berücksichtigen, dass die beiden Facetten nicht notwendigerweise über dieselbe Anzahl an Facettenausprägungen verfügen müssen, so dass grundsätzlich drei Fälle zu unterscheiden sind. Diese sind im Weiteren differenzierter hinsichtlich der einzelnen Facettenausprägungen zu analysieren:

1. $|F2()| = |F1()|$, d. h. die neu zu indexierende Facette $F2(a2_{IQ1}, b2_{IQ1})$ besitzt genauso viele Ausprägungen wie die bereits indexierte Facette $F1(a1_{IQ1}, b1_{IQ1})$.
 - a) $a1_{IQ1} = a2_{IQ1}$ AND $b1_{IQ1} = b2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*, b_{IQ1}^*)$ mit $a_{IQ1}^* = a1_{IQ1} = a2_{IQ1}$ und $b_{IQ1}^* = b1_{IQ1} = b2_{IQ1}$
 - b) $a1_{IQ1} = a2_{IQ1}$ AND $b1_{IQ1} \neq b2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*, b_{IQ1}^*)$ mit $a_{IQ1}^* = a1_{IQ1} = a2_{IQ1}$ und $b_{IQ1}^* = b2_{IQ1}$, da $Ed_{b2_{IQ1}} > Ed_{b1_{IQ1}}$
 - c) $a1_{IQ1} \neq a2_{IQ1}$ AND $b1_{IQ1} = b2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*, b_{IQ1}^*)$ mit $a_{IQ1}^* = a2_{IQ1}$, da $Ed_{a2_{IQ1}} > Ed_{a1_{IQ1}}$ und $b_{IQ1}^* = b1_{IQ1} = b2_{IQ1}$
 - d) $a1_{IQ1} \neq a2_{IQ1}$ AND $b1_{IQ1} \neq b2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*, b_{IQ1}^*)$ mit $a_{IQ1}^* = a2_{IQ1}$, da $Ed_{a2_{IQ1}} > Ed_{a1_{IQ1}}$ und $b_{IQ1}^* = b2_{IQ1}$, da $Ed_{b2_{IQ1}} > Ed_{b1_{IQ1}}$
2. $|F2()| < |F1()|$, d. h. die neu zu indexierende Facette $F2(a2_{IQ1})$ besitzt weniger Ausprägungen als die bereits indexierte Facette $F1(a1_{IQ1}, b1_{IQ1})$. In diesem Fall ist $b1_{IQ1}$ als Ausprägung für die Facette aus dem Index zu löschen, so dass nur eine Facettenausprägung aktuell gültig ist.
 - a) $a1_{IQ1} = a2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*)$ mit $a_{IQ1}^* = a1_{IQ1} = a2_{IQ1}$

- b) $a1_{IQ1} \neq a2_{IQ1}$
 $\Rightarrow F^*(a_{IQ1}^*)$ mit $a_{IQ1}^* = a2_{IQ1}$, da $Ed_{a2_{IQ1}} > Ed_{a1_{IQ1}}$
3. $|F2()| > |F1()|$, d.h. die neu zu indexierende Facette $F2(a2_{IQ1}, b2_{IQ1}, c2_{IQ1})$ besitzt mehr Ausprägungen als die bereits indexierte Facette $F1(a1_{IQ1}, b1_{IQ1})$. In diesem Fall muss wie unter 1. vorgegangen werden, wobei die daraus resultierende Facette $F^*()$ in jedem Fall um die neue Facettenausprägung $c2_{IQ1}$ zu erweitern ist.

6.2.2. Aktualisierungsszenario 2: Unterschiedliche Informationsquellen

Das zweite Aktualisierungsszenario betrachtet den Fall, dass die für ein Artefakt bereits indexierte Facette $F4(a4_{IQ4}, b4_{IQ4})$ und die für dieses Artefakt neu zu indexierende Facette $F5()$ aus unterschiedlichen Informationsquellen stammen. Folglich sind die Informationen aus beiden Datenquellen im Index zu berücksichtigen, was bei einer Übereinstimmung von Facettenausprägungen dazu führt, dass für eine korrekte Verarbeitung in eventuell folgenden Aktualisierungen beide Informationsquellen für die jeweilige Facettenausprägung zu speichern sind. Im Fall einer Nichtübereinstimmung der Facettenausprägung sind die Ausprägungen beider Informationsquellen zu indexieren. Somit spielt hier im Gegensatz zum Aktualisierungsszenario 1 das Extraktionsdatum der einzelnen Facettenausprägungen keine Rolle; eine Differenzierung anhand der Anzahl der Facettenausprägungen muss jedoch ebenfalls vorgenommen werden:

1. $|F5()| = |F4()|$, d.h. die für ein Artefakt neu zu indexierende Facette $F5(a5_{IQ5}, b5_{IQ5})$ besitzt genauso viele Ausprägungen wie die bereits indexierte Facette $F4(a4_{IQ4}, b4_{IQ4})$, stammt jedoch nicht aus der Informationsquelle $IQ4$.
- a) $a4_{IQ4} = a5_{IQ5}$ AND $b4_{IQ4} = b5_{IQ5}$
 $\Rightarrow F^*(a_{IQ4, IQ5}^*, b_{IQ4, IQ5}^*)$ mit $a_{IQ4, IQ5}^* = a4_{IQ4} = a5_{IQ5}$ und $b_{IQ4, IQ5}^* = b4_{IQ4} = b5_{IQ5}$
- b) $a4_{IQ4} = a5_{IQ5}$ AND $b4_{IQ4} \neq b5_{IQ5}$
 $\Rightarrow F^*(a_{IQ4, IQ5}^*, b4_{IQ4}, b5_{IQ5})$ mit $a_{IQ4, IQ5}^* = a4_{IQ4} = a5_{IQ5}$
- c) $a4_{IQ4} \neq a5_{IQ5}$ AND $b4_{IQ4} = b5_{IQ5}$
 $\Rightarrow F^*(a4_{IQ4}, a5_{IQ5}, b_{IQ4, IQ5}^*)$ mit $b_{IQ4, IQ5}^* = b4_{IQ4} = b5_{IQ5}$
- d) $a4_{IQ4} \neq a5_{IQ5}$ AND $b4_{IQ4} \neq b5_{IQ5}$
 $\Rightarrow F^*(a4_{IQ4}, b4_{IQ4}, a5_{IQ5}, b5_{IQ5})$

2. $|F5()| < |F4()|$, d. h. die neu zu indexierende Facette $F5(a5_{IQ5})$ stammt nicht aus $IQ4$ und besitzt weniger Ausprägungen als die bereits indexierte Facette $F4(a4_{IQ4}, b4_{IQ4})$.
 - a) $a4_{IQ4} = a5_{IQ5}$
 $\Rightarrow F^*(a^*_{IQ4,IQ5}, b4_{IQ4})$ mit $a^*_{IQ4,IQ5} = a4_{IQ4} = a5_{IQ5}$
 - b) $a4_{IQ4} \neq a5_{IQ5}$
 $\Rightarrow F^*(a4_{IQ4}, b4_{IQ4}, a5_{IQ5})$
3. $|F5()| > |F4()|$, d. h. die neu zu indexierende Facette $F5(a5_{IQ5}, b5_{IQ5}, c5_{IQ5})$ besitzt mehr Ausprägungen als die bereits indexierte Facette $F4(a4_{IQ4}, b4_{IQ4})$, stammt aber nicht aus $IQ4$. In diesem Fall ist wie unter 1. beschrieben vorzugehen und die resultierende Facette $F^*()$ jeweils mit der Facettenausprägung $c5_{IQ5}$ zu erweitern.

Betrachtet man die Szenarios in Kapitel 6.2.1, so wird deutlich, dass alle Ausprägungen einer Facette immer durch dieselbe Informationsquelle charakterisiert sind. Im Vergleich dazu weisen die Ausprägungen der aus dem Aktualisierungsszenario 2 resultierenden Facetten verschiedene Informationsquellen auf, was insbesondere bei weiteren erforderlichen Aktualisierungen zu beachten ist und damit wiederum die Komplexität des Vorgehens erhöht. Anhand der folgenden beiden Szenarios soll dieses Vorgehen mit den darin durchzuführenden Prüfungen beispielhaft demonstriert werden:

4. Die unter Punkt 1d) entstandene und indexierte Facette $F^*(a4_{IQ4}, b4_{IQ4}, a5_{IQ5}, b5_{IQ5})$ ist im Anschluss mit der Facette $F6(a6_{IQ5})$ zu aktualisieren. Hier ist zu berücksichtigen, dass die Informationen aus der Informationsquelle $IQ5$ schon einmal indexiert wurden. Folglich sind alle Ausprägungen aus $IQ5$ gemäß dem unter Kapitel 6.2.1, Punkt 2 beschriebenen Vorgehen zu verifizieren. Des Weiteren ist zu beachten, dass die Facettenausprägungen aus $IQ4$ nicht einfach übernommen werden können. Schließlich besteht die Möglichkeit, dass Ausprägung $a4_{IQ4}$ mit der Ausprägung $a6_{IQ5}$ übereinstimmt, so dass diesbezüglich ebenfalls eine Fallunterscheidung vorzunehmen ist.
5. Die unter Punkt 2b) entstandene und indexierte Facette $F^*(a4_{IQ4}, b4_{IQ4}, a5_{IQ5})$ muss mit der Facette $F7(a7_{IQ5}, b7_{IQ5})$ aktualisiert werden. In diesem Fall sind wiederum zunächst die Facettenausprägungen aus $IQ5$ zu überprüfen. Dabei sind die Ausprägung $b7_{IQ5}$ und entweder beide (bei Nichtübereinstimmung) oder die aktuellere der beiden a -Ausprägungen (bei Übereinstimmung) für die Facette beizubehalten. Bevor diese jedoch indexiert werden, sind sie mit den Facettenausprägungen aus $IQ4$ zu vergleichen. Würde die Facette $F7$ dagegen nur $b7_{IQ5}$ als Ausprägung enthalten, wäre die Ausprägung $a5_{IQ5}$ zu entfernen und damit aus dem Index zu löschen.

6.2.3. Zusammenfassung

Wie die beispielhaften Szenarios in den vorhergehenden Kapiteln 6.2.1 und 6.2.2 verdeutlichen, ist die Herstellung konsistenter Facetten mit einem hohen Grad an Komplexität verbunden. Neben einer Validierung der Daten im Hinblick auf das gemäß Kapitel 5.1 definierte Facettenschema, ist zusätzlich eine Verifikation der Daten vorzunehmen, welche die Richtigkeit und Aktualität der Daten im Index gewährleistet. Dabei ist die Erfassung und Verwaltung von Herkunftsinformationen von entscheidender Bedeutung, so dass diese bei der Realisierung des LFRP-Frameworks — insbesondere für das in den Kapiteln 3 bis 6 vorgestellte Indexierungsframework — von Anfang an zu berücksichtigen sind.

Insgesamt hat sich in diesen Kapiteln 3 bis 6 somit gezeigt, dass die Realisierung eines geeigneten Indexierungsframeworks die Beherrschung eines hohen Komplexitätsgrades erfordert, der angefangen bei den zu unterstützenden Suchszenarios und den daraus resultierenden Artefakttypen bis hin zur Konsistenzprüfung der erstellten Repräsentationen den gesamten Prozess der Indexierung charakterisiert. Damit stellt sich die Frage, ob sich die Realisierung eines derartig komplexen Frameworks überhaupt lohnt, weshalb im folgenden Kapitel 7 mittels einer kritischen Reflexion eine Antwort auf diese Frage gegeben werden soll.

7. Kritische Reflexion

Die Realisierung des vorgestellten Indexierungsframeworks in einem konkreten unternehmerischen Umfeld ist ein Vorhaben, das aufgrund seiner Komplexität nicht ohne eine kritische Betrachtung aller damit verbundenen Aspekte durchgeführt werden sollte. Aus diesem Grund wird im Folgenden ein derartiges Vorgehen einer kritischen Reflexion unterzogen, welche sich auf die Erkenntnisse aus der prototypischen Implementierung des LFRP-Frameworks stützt. Dabei ist das Realisierungsvorhaben prinzipiell aus zwei Perspektiven zu analysieren und zu beurteilen. Der Fokus der ersten Perspektive liegt dabei auf der Realisierung des Frameworks an sich und damit auf dem Prozess bzw. den einzelnen Schritten, welche zur Umsetzung durchzuführen sind. Die zweite Perspektive stellt dagegen eher die Auswirkungen bestimmter Aspekte bei der Realisierung in den Vordergrund der Betrachtung und differenziert somit ergebnisorientiert zwischen dem Nutzen einerseits und den kritischen Erfolgsfaktoren andererseits. Beide Perspektiven werden in den folgenden Kapiteln 7.1 und 7.2 betrachtet.

7.1. Prozessorientierte Reflexion: Leitfaden für eine unternehmensspezifische Realisierung

Wie in den Kapiteln 4 und 5 gezeigt wurde, kann das vorgestellte Indexierungsframework im konkreten Realisierungsvorhaben abhängig von den berücksichtigten Artefakttypen und den für deren Informationserfassung genutzten Informationsquellen beliebig komplex gestaltet werden. Folglich bedingt diese gestalterische Komplexität auch die Komplexität des Vorgehens, das im Rahmen der jeweiligen Umsetzung durchzuführen ist. So lässt sich das Indexierungsframework beispielsweise realisieren, indem nur die im Unternehmen vorhandenen IT-Systeme als Informationsquellen für die Artefaktinformationen genutzt werden. Dadurch kann zwar bereits ein Großteil der notwendigen Funktionalitäten erreicht werden; auf Teilfunktionalitäten wie beispielsweise eine geometrische Ähnlichkeitssuche muss jedoch in diesem Fall verzichtet werden. Des Weiteren muss man sich auch des Informationsverlustes bewusst sein, der durch eine Nichtberücksichtigung von Dokumenten entsteht und einen wesentlichen Einfluss auf die Qualität der Suchergebnisse und damit auch die Akzeptanz des Systems hat. Demzufolge wird für eine Realisierung immer die Berücksichtigung

von beiden Informationsquellen — von IT-Systemen und Dokumenten — empfohlen, wobei selbstverständlich die Anzahl der zu berücksichtigenden und folglich zu verarbeitenden Dokumenttypen und ihrer Dateiformate als Kostentreiber zu sehen ist. Dieser muss in jedem Fall im Rahmen des konkreten Realisierungsvorhabens kritisch hinsichtlich der beiden Faktoren Nutzen und Aufwand analysiert werden, so dass schließlich entsprechende, auf diesen Analyseergebnissen und dem jeweiligen Anwendungsszenario basierende Einschränkungen bei der Repräsentationserstellung aus Dokumenttypen vorgenommen werden.

Aufgrund der aus diesen Überlegungen und Entscheidungen resultierenden Komplexität wurde ein Leitfaden erstellt, der das Vorgehen zur Realisierung in einem konkreten unternehmerischen Umfeld überblicksartig erläutert. Dazu werden die einzelnen, durchzuführenden Schritte als Vorgehensmodell in den Abbildungen 7.1 bis 7.3 visualisiert. Dieses VM ist als eine Art Orientierungshilfe für diejenige(n) Person(en) in einem Unternehmen gedacht, welche die Realisierung des Indexierungsframeworks bzw. des LFRP-Frameworks zu verantworten hat (haben).

Insgesamt betrachtet besteht dieses VM aus fünf Hauptaktivitäten, welche in den Abbildungen 7.1 bis 7.3 durch graue Kästen dargestellt sind. Dabei sind vor der eigentlichen Realisierung des Frameworks zunächst die domänenspezifischen Informationsbedürfnisse zu analysieren, die selbstverständlich von der Ziel-Nutzergruppe des LFRP-Frameworks abhängen (vgl. Kapitel 1.1.2 für die Analyse der Ziel-Nutzergruppe *Produktentwickler*). Parallel dazu sind die in der jeweiligen Domäne vorkommenden Artefakte zu identifizieren und in Artefakttypen zu differenzieren. Da jedoch eine Unterstützung aller möglichen Artefakttypen nicht zielführend ist, sind ausgehend von diesen Analyseergebnissen die für die Informationsbedürfnisse der Ziel-Nutzergruppe relevanten Artefakttypen auszuwählen und hinsichtlich ihrer artefakttypspezifischen Suchkriterien zu untersuchen (vgl. hierzu Kapitel 4.1 und 4.2). Im Anschluss daran kann die Auswahl der für das Framework benötigten Suchkriterien erfolgen, welche als Facetten zusammen mit der im Framework zu unterstützenden Artefakttyphierarchie im Facettenschema festgelegt werden (vgl. Kapitel 5.1). Nach der prinzipiellen Definition, welche artefakttypspezifischen Suchkriterien durch welche Facetten im Framework repräsentiert werden sollen, müssen die dafür benötigten Artefaktinformationen aus einer Quelle gewonnen werden. Dazu sind spätestens im nächsten Schritt sowohl die IT-System- als auch die Dokumentenlandschaft im jeweiligen Unternehmen oder Unternehmensbereich zu untersuchen und geeignete Quellen auszuwählen (siehe Kapitel 4.3). Im Fall anzubindender IT-Systeme sind — sofern keine Dokumente in den IT-Systemen verwaltet werden — deren Informationen anschließend in Form von Name-Wert-Paaren und gegebenenfalls Assoziationen zu extrahieren und mit Hilfe einer XML-Importdatei in das Indexierungsframework zu importieren (vgl. Kapitel 4.4). Dort sind sie so zu verarbeiten, dass für jedes Artefakt eine Polyrepräsentation aus Attribut- und gegebenenfalls Beziehungs- und / oder Thumbnailfacetten erstellt wird (siehe Ausführungen in Kapitel 5.2). In dem Fall, in dem (auch) Doku-

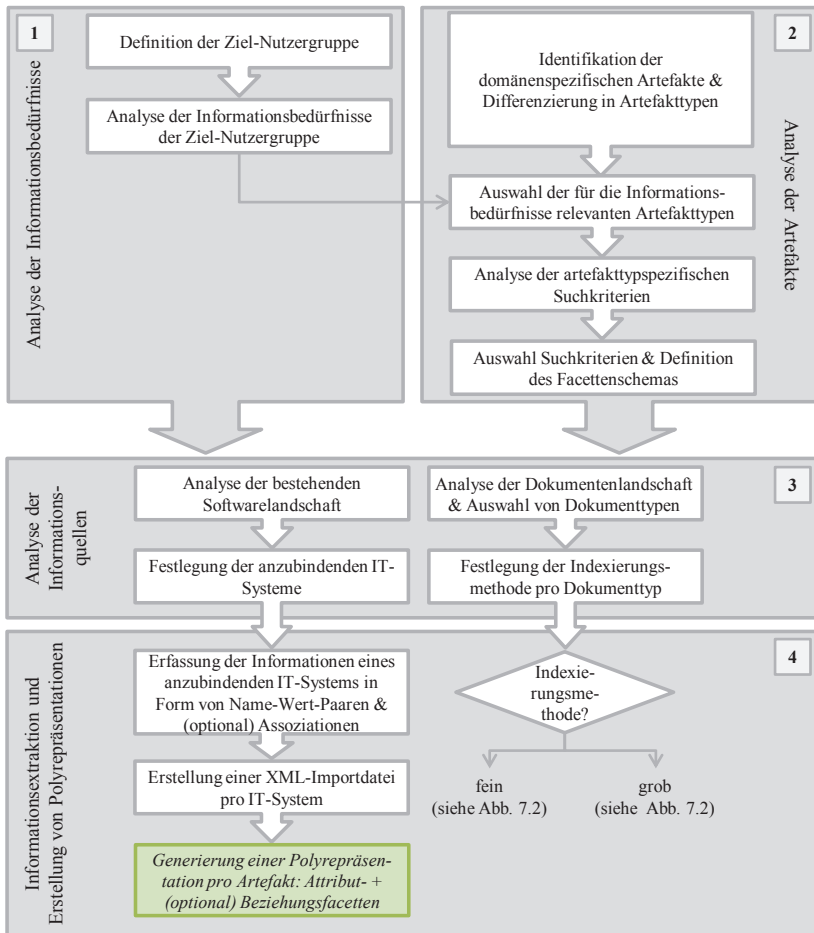


Abbildung 7.1. — Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks

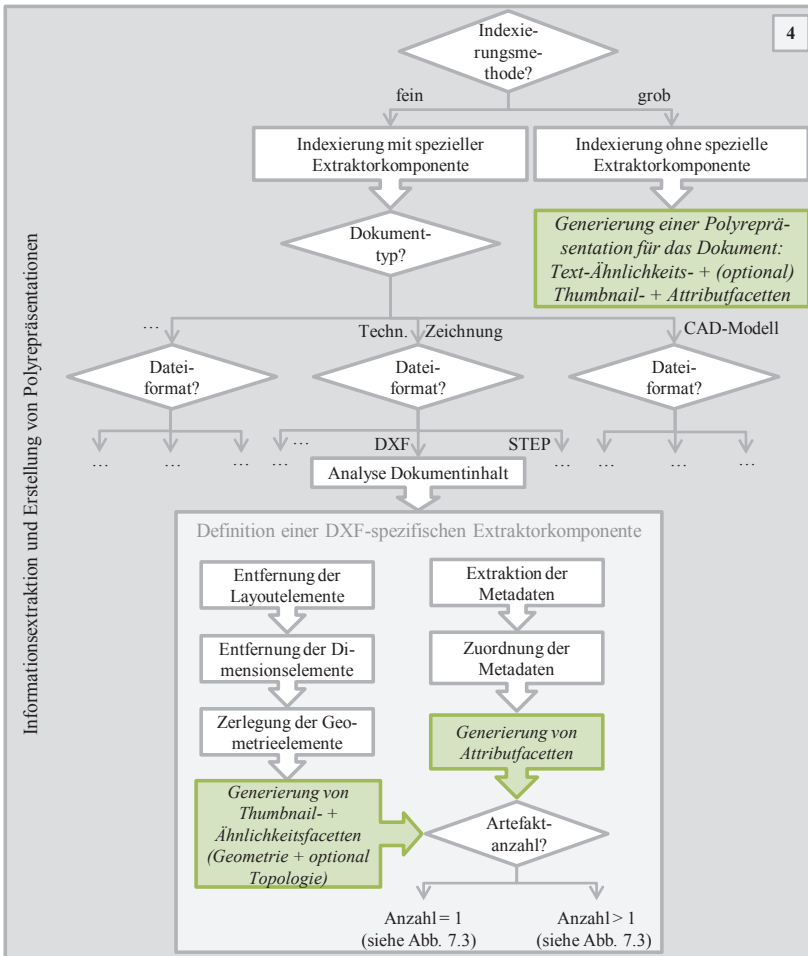


Abbildung 7.2. — Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks (Fortsetzung)

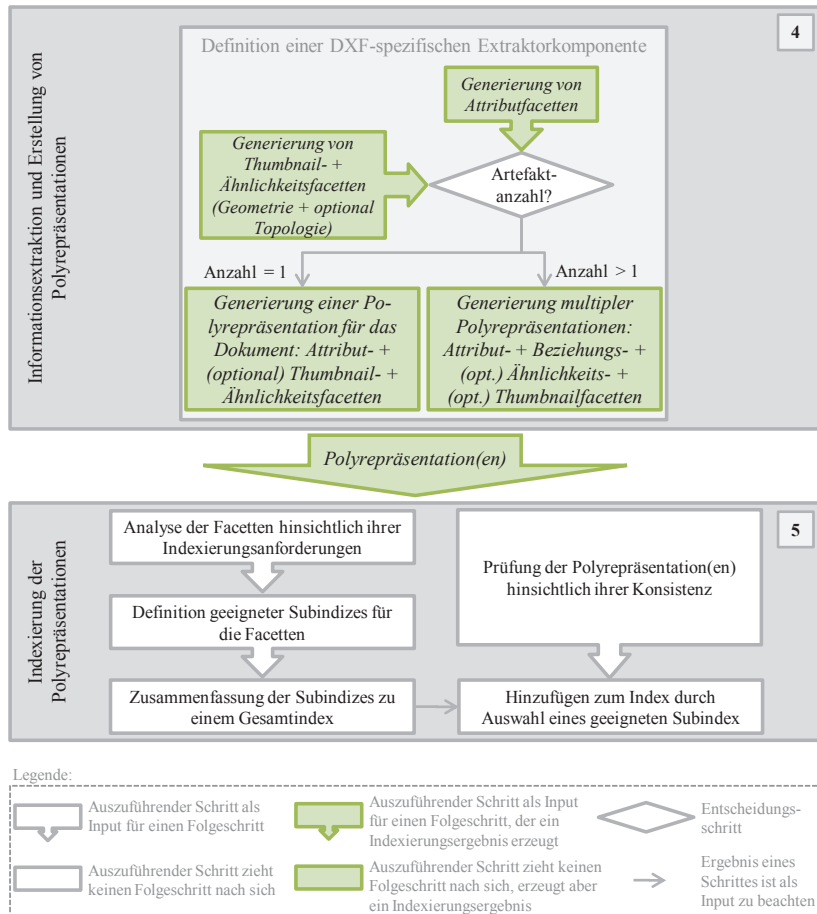


Abbildung 7.3. — Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks (Fortsetzung)

mente berücksichtigt werden, muss jeder Dokumenttyp detailliert analysiert und für diesen eine Indexierungsmethode festgelegt werden (vgl. hierzu Kapitel 5.3 und 5.4). Dabei ist grundsätzlich zwischen einer feinen und einer groben Indexierungsmethode zu differenzieren (siehe Kapitel 5.4.6). Wird eine feine Indexierungsmethode für einen Dokumenttyp festgelegt, so ist für diesen eine spezielle Extraktorkomponente auf Basis seines Dokumentinhaltes und seines Dateiformates zu erstellen. Zur Verdeutlichung zeigt Abbildung 7.2 als Beispiel die Definition der in Kapitel 5.4.4.4 vorgestellten Extraktorkomponente für technische Zeichnungen im DXF-Format, die Attribut-, Ähnlichkeits-, Thumbnail- und / oder Beziehungsfacetten erzeugt und diese abhängig von den im Dokument betrachteten Artefakten zu einer oder mehreren Polyrepräsentationen zusammenfasst. Soll hingegen eine grobe Indexierungsmethode für einen Dokumenttyp verwendet werden, so ist keine spezielle Extraktorkomponente zu realisieren. In diesem Fall ist vielmehr eine Polyrepräsentation für das Dokument selbst zu erzeugen, die aus dokumentspezifischen Metadaten erstellte Attributfacetten, eine optionale Thumbnailfacette und eine Ähnlichkeitsfacette für den gesamten textuellen Inhalt des Dokumentes (QbT) enthält. Abschließend ist jede in der Hauptaktivität *Informationsextraktion und Erstellung von Polyrepräsentationen* generierte Polyrepräsentation zu indexieren. Dazu ist ein Index zu definieren, der abhängig von den Indexierungsanforderungen der benötigten Facetten aus verschiedenen Subindizes besteht. Nach einer erfolgreich durchgeführten Konsistenzprüfung sind diesen letztlich die jeweiligen Facetten hinzuzufügen (siehe Kapitel 6).

7.2. Ergebnisorientierte Reflexion: Nutzen und kritische Erfolgsfaktoren

Nach einer prozessorientierten Betrachtung der Realisierung des vorgestellten Indexierungsframeworks wird diese nun im Weiteren einer eher ergebnisorientierten Reflexion unterzogen. Dazu ist der Nutzen zu analysieren, der sich durch die Umsetzung des Indexierungsframeworks realisieren lässt. Per definitionem beschreibt der Nutzen „(...) die Fähigkeit eines Gutes, ein bestimmtes Bedürfnis (...) befriedigen zu können;“ [Gab11b, Abschnitt „Kurzerklärung“]. Damit spiegelt der Nutzen in gewisser Weise den Vorteil oder auch Ertrag bzw. Gewinn wider, den man durch die Befriedigung dieses Bedürfnisses erzielt. Aus diesem Grund wird häufig auch vom wirtschaftlichen Wert eines Gutes gesprochen. Allerdings lässt sich dieser Vorteil nur dann realisieren, wenn bestimmte Faktoren die Erreichung des Vorteils ermöglichen. Diese Faktoren werden als kritische Erfolgsfaktoren bezeichnet, da sie für den Erfolg und damit den Nutzen von entscheidender Bedeutung sind [Gab11a, Abschnitt „Kurzerklärung“]. Stellt man sich folglich die Frage nach dem Nutzen, so ist damit zwangsläufig auch immer eine Betrachtung der kritischen Erfolgsfaktoren verbunden.

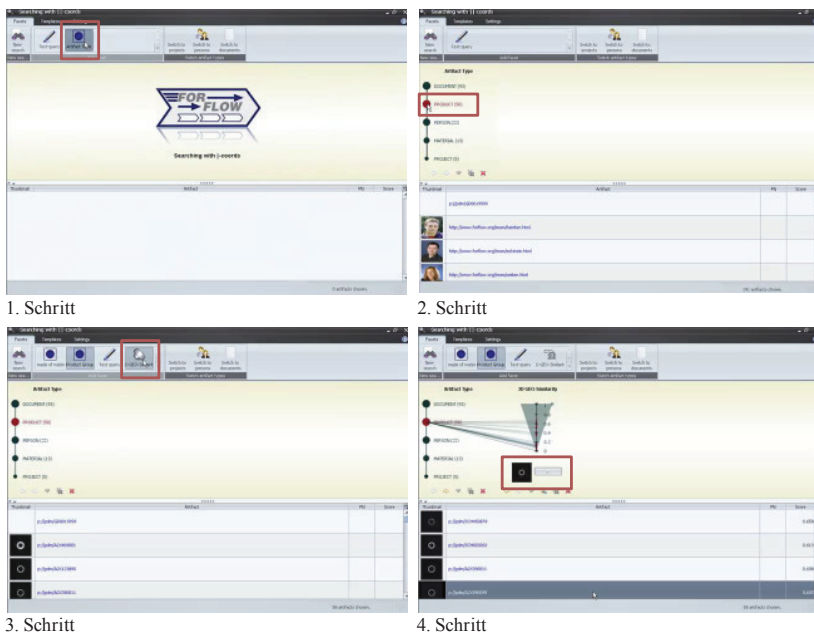


Abbildung 7.4. — Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 1. bis 4. Suchschritt

Bezogen auf die vorliegende Fragestellung führt die Analyse des Nutzens in erster Linie zu den Anforderungen zurück, aufgrund derer das vorgestellte Indexierungsframeworks entwickelt wurde. Ziel dieser Anforderungen war die Bereitstellung einer übergreifenden Suchfunktionalität, die sowohl unerfahrenen als auch erfahrenen Produktentwicklern in komplexen Suchsituationen Unterstützung beim (Wieder-)Auffinden von Informationen bietet und dadurch u. a. eine Verringerung der Suchzeit, eine Entlastung bei Routineaufgaben, eine Erhöhung des Wiederverwendungsgrades bereits existierender Artefakte und damit letztendlich eine Reduktion der Entwicklungskosten bewirkt. Die Möglichkeiten, die das hierzu erforderliche facettierte Suchvorgehen mit Hilfe der in Kapitel 3.3.2 vorgestellten Suchoberfläche den Nutzern bietet, wird zur Verdeutlichung in den Abbildungen 7.4 bis 7.6 dargestellt. Dazu wird das Suchszenario D aus Kapitel 1.1.2 (Seite 15) aufgegriffen, in dem ein Produktentwickler das CAD-Modell einer ringförmigen Dichtung benötigt. Um dieses Dokument zu finden, wählt der Produktentwickler im vorliegenden Beispiel als erstes den ihn interessierenden Artefakttyp (hier *Produkt*) mit Hilfe der Facette zum Artefakttyp aus (Abbildung 7.4:

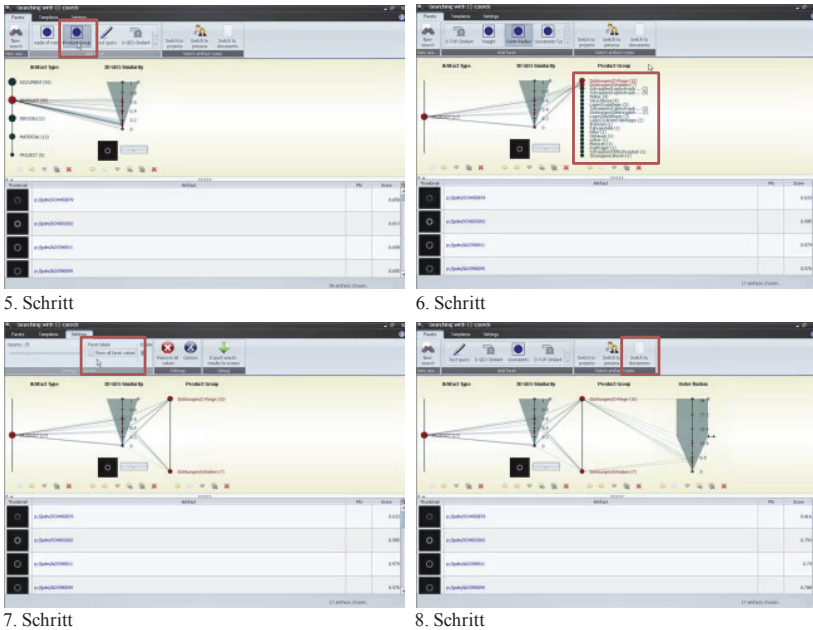
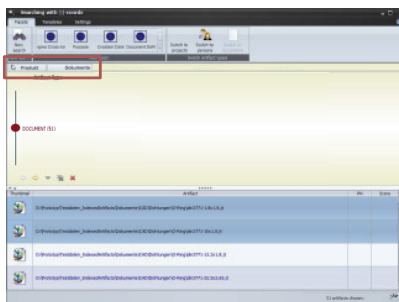


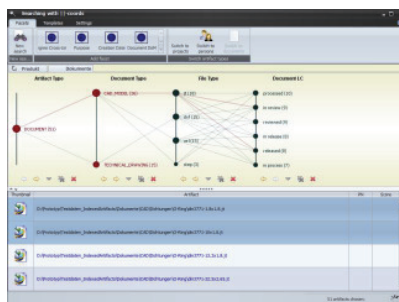
Abbildung 7.5. — Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 5. bis 8. Suchschritt

1. und 2. Schritt). Anschließend fügt er seiner Anfrage die Facette zur geometrischen Ähnlichkeitssuche hinzu und übergibt dieser eine 3D-Beschreibung eines ringförmigen Objektes. Auf diese Weise schränkt der Produktentwickler die Ergebnismenge auf Produkte ein, die eine ähnliche, also ebenfalls ringförmige Form besitzen (Abbildung 7.4: 3. und 4. Schritt). Da jedoch die Vielfalt an ringförmigen Produkten in der Suchergebnisliste groß ist und der Produktentwickler in diesem speziellen Fall weiß, dass er eine Dichtung benötigt, wendet er im nächsten Schritt die Facette zur Produktgruppe als Filterkriterium an (Abbildung 7.5: 5. Schritt). Hier wählt er die für ihn relevanten Facettenausprägungen aus und blendet alle nicht-relevanten Ausprägungen zur besseren Übersichtlichkeit aus (Abbildung 7.5: 6. und 7. Schritt). Daraufhin filtert der Produktentwickler aus der Menge der gefundenen Dichtungen diejenigen heraus, deren Wert für den Außenradius innerhalb eines bestimmten, mittels der Präferenzfunktion definierten Intervalls liegt (Abbildung 7.5: 8. Schritt). Aus diesen wählt der Produktentwickler die für seine Aufgabe passende Dichtung aus und wechselt im nächsten Schritt auf die Ebene der Dokumente, um das zugehörige CAD-Modell zu

finden (Abbildung 7.5: 8. Schritt). Hier schränkt der Produktentwickler die Menge der für die ausgewählte Dichtung relevanten Dokumente solange anhand von Facetten ein, bis er das gesuchte Dokument gefunden hat (Abbildung 7.6: 9. und 10. Schritt).



9. Schritt



10. Schritt

Abbildung 7.6. — Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 9. und 10. Suchschritt

Obwohl das hier dargestellte Beispiel nur eine Möglichkeit des Suchens beschreibt, macht es bereits den wesentlichen Nutzen des LFRP-Frameworks und damit auch des Indexierungsframeworks deutlich: Trotz vieler Freiheitsgrade beim Suchen werden Nutzer gerade in schwierigen, nicht einfach zu handhabenden Suchsituationen durch den Suchprozess und damit zu einem Ergebnis geführt, was in einem Erfolgserlebnis für den Nutzer resultiert und seine Einstellung und Akzeptanz gegenüber dem System wesentlich beeinflusst. Dabei ist zu berücksichtigen, dass dieser Erfolg zum einen natürlich von der Suchoberfläche — v. a. von ihrer Gestaltung und den von ihr angebotenen Suchfunktionalitäten — abhängt. Zum anderen trägt aber auch das Indexierungsframework einen wesentlichen Anteil hierzu bei. Schließlich müssen die im Unternehmen existierenden Informationen in einer ganz bestimmten Art und Weise aufbereitet werden, um diese Suchfunktionalitäten überhaupt anbieten zu können. Dazu sind gemäß Kapitel 7.1 fünf Hauptaktivitäten durchzuführen, wobei einer der wesentlichsten Aspekte, der berücksichtigt werden muss, die Nutzer des LFRP-Frameworks betrifft. Um ihre Akzeptanz zu gewinnen, müssen diese sowohl Freude am Umgang mit dem System haben als auch von der Ergebnislieferung des Systems positiv überzeugt sein. Aus diesem Grund muss zum einen die Zeit, während der ein Nutzer auf das Suchergebnis wartet, möglichst kurz sein. Damit ist die Verwendung geeigneter Indexstrukturen als kritischer Erfolgsfaktor zu sehen, da diese wie in Kapitel 6.2 erläutert, diese Wartezeit wesentlich beeinflussen. Zum anderen muss aber auch eine hohe Qualität und Aktualität der im System enthaltenen Informationen angestrebt werden. Hierfür sind neben Informationen aus wichtigen IT-Systemen auch Dokumente zu indexieren, die darüber hinaus möglichst detailliert zu analysieren und zu verarbei-

ten sind. Dazu werden wiederum spezielle Extraktorkomponenten benötigt, wodurch der Aufwand für die erstmalige Implementierung des Indexierungsframeworks steigt. Auch die Anzahl der betrachteten Informationsquellen an sich stellt einen kritischen Erfolgsfaktor dar. Zum einen wächst mit dieser Anzahl auch die Menge und Vielfalt an Informationen, die zu verarbeiten und zu indexieren ist, was wiederum in steigenden Speicheranforderungen resultiert. Zum anderen steigt damit aber auch die Gefahr einer Informationsüberflutung, da gegebenenfalls kein sinnvolles Maß bei der Begrenzung der Suchkriterien gefunden wird. Zwar kann davon ausgegangen werden, dass mit einer zunehmenden Anzahl von Suchkriterien bzw. Facetten der Strukturierungsgrad einer Suche steigt und dadurch die Wahrscheinlichkeit einer positiven Zielerreichung erhöht wird. Gleichzeitig ist aber zu bedenken, dass eine vollständige Abdeckung aller möglichen Suchszenarios nur im Idealfall zu erreichen ist und damit der mit dem Indexierungsframework angestrebte Komplexitätsgrad ebenfalls als kritischer Erfolgsfaktor zu sehen ist. Ist das LFRP-Framework allerdings einmal implementiert, so ist sein primärer Nutzen in der größtenteils automatischen Aufbereitung der Informationen zu sehen, in deren Rahmen der manuelle Indexierungsaufwand — abhängig von der Datenqualität und Datenvollständigkeit in einem Unternehmen — deutlich reduziert ist. Aufwändiger kann sich hingegen die Wartung des IRS gestalten, wenn das System an neue Anforderungen des Unternehmens oder an Veränderungen der IT-Infrastruktur anzupassen ist. Folglich ist mit der Umsetzung und Wartung des Indexierungsframeworks zwar ein gewisser, nicht zu unterschätzender Aufwand verbunden; es sollte allerdings bedacht werden, dass nur auf diese Art und Weise sinnvolle und zielführende Suchfunktionalitäten mit einer hohen Ergebnisqualität bereitgestellt werden können.

8. Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit ist die Konzeption eines Indexierungsframeworks, das den Indexierungsprozess eines IRS abwickelt und damit als Grundvoraussetzung die Realisierung eines interaktiven Retrievalmodells ermöglicht. Dieses Retrievalmodell wird benötigt, um durch die Bereitstellung geeigneter Suchfunktionalitäten Produktentwickler bei ihrer Suche nach Informationen während des PEP zu unterstützen und durch die Verringerung der Suchzeit sowie durch die gezielte Wiederverwendung bereits existierender Daten eine Reduktion der Entwicklungskosten zu bewirken. Da sich dieses als LFRP-Framework bezeichnete Retrievalmodell auf die vier, aus den Bereichen Modellierung, IR und Visualisierung bekannten Konzepte der multiplen Ebenen, der facettierten Suche, des Rankings und der parallelen Koordinaten stützt, erfordert es eine geeignete facettierte Aufbereitung und Indexierung der suchbaren Informationen. Dazu war es zunächst notwendig, die betrachtete Domäne der technischen PE näher zu untersuchen, indem neben dem objektiven und dem subjektiven Informationsbedarf der Ziel-Nutzergruppe auch die in dieser Domäne auftretenden Suchsituationen anhand von Beispielszenarios analysiert wurden. Auf Basis dieser Analyseergebnisse konnte ein hoher Komplexitätsgrad der in dieser Domäne vorkommenden Suchsituationen identifiziert werden, der primär durch die Vielfalt der zu befriedigenden Informationsbedürfnisse und der damit verbundenen Vielzahl an nachgefragten Suchobjekten geprägt ist. Die Suchobjekte selbst werden dabei in Form von Artefakten unterschiedlichen Artefakttyps (Produkt, Dokument, Person usw.) und häufig in Zusammenhang mit bestimmten Eigenschaften oder Merkmalen in den Informationsbedürfnissen adressiert. Demzufolge werden für die jeweiligen Artefakttypen spezifische Suchkriterien von den Produktentwicklern benötigt, die im Rahmen der zu entwickelnden Suchfunktionalitäten zur Verfügung zu stellen sind. Um allerdings an die notwendigen Informationen zur Bereitstellung dieser als Facetten realisierten Suchkriterien zu gelangen, sind unterschiedliche im Produktentwicklungsbereich existierende Informationsquellen zu berücksichtigen. Neben IT-Systemen enthalten insbesondere entwicklungsspezifische Dokumente wertvolle Informationen, die mit Hilfe vom jeweiligen Dokumenttyp und dessen Dateiformat abhängiger Extraktorkomponenten zu erfassen sind. Folglich resultiert die vom Indexierungsframework zu beherrschende Gesamtkomplexität gemäß Abbildung 8.1 genau genommen aus drei Teilkomplexitäten, welche die Ergebnisse der drei in dieser Arbeit untersuchten Forschungsfragen widerspiegeln:



Abbildung 8.1. — Darstellung der aus drei Teilkomplexitäten bestehenden und durch das Indexierungs- und damit auch das LFRP-Framework zu beherrschenden Gesamtkomplexität

- Forschungsfrage 1: *Was?*
⇒ Komplexität der Informationsbedürfnisse und der daraus resultierenden Suchobjekte
- Forschungsfrage 2: *Woher?*
⇒ Komplexität der Informationsquellen
- Forschungsfrage 3: *Wie?*
⇒ Komplexität der Informationsextraktion (insbesondere aus Dokumenten)

Das in den Kapiteln 3 bis 6 vorgestellte Indexierungsframework weist zwar einen hohen Komplexitätsgrad auf und ist damit nicht ohne einen gewissen Aufwand zu realisieren; jedoch ist seine Realisierung insgesamt als lohnenswert und gewinnbringend zu beurteilen. Wie W. R. Ashby bereits im Jahr 1957 feststellte, muss die Varietät eines lenkenden Systems mindestens so groß sein wie die Varietät des gelenkten Systems (Gesetz von Ashby) [Ash57]. Somit kann eine gegebene Komplexität nur mit Hilfe eines mindestens ebenso komplexen Systems beherrscht werden, was impliziert, dass im Fall der Realisierung eines Indexierungsframeworks geringerer Komplexität immer Einschränkungen bei der Erfüllung der Anforderungen in Kauf zu nehmen sind. Allerdings ist auch zu berücksichtigen, dass Informationen prinzipiell nur dann (wieder) gefunden werden können, wenn sie auch dokumentiert sind. Das bedeutet: Sind die notwendigen Informationen nicht verfügbar, scheitern selbst die besten Extraktionsmethoden und damit die Indexierungskomponente und schließlich auch das darauf basierende Suchsystem.

Des Weiteren ist das vorgestellte Indexierungsframework in der vorliegenden Arbeit zwar speziell für den Bereich der technischen (v. a. mechanischen) PE konzipiert; es

ist jedoch insgesamt als generisches Framework zu betrachten. Aufgrund der Tatsache, dass Artefakte in den Mittelpunkt der Indexierung gestellt werden, lässt sich das Indexierungsframework durchaus auch für andere Domänen nutzen. Hierfür kommen prinzipiell alle Anwendungsbereiche in Frage, in denen komplexe Informationsbedürfnisse zu befriedigen sind und somit das Erforschen von Daten und Informationen eine wichtige Rolle spielt. Beispiele sind u. a. die Entwicklung elektronischer bzw. elektrotechnischer Komponenten, die Softwareentwicklung, die Medizintechnik oder der Multimediabereich. Eine Anpassung für einen dieser Anwendungsbereiche lässt sich, wie auch anhand des Leitfadens in Kapitel 7 zu erkennen ist, zielgerichtet umsetzen, indem grundsätzlich die für die jeweilige Domäne und Informationsbedürfnisse relevanten Artefakttypen zu definieren sowie deren Artefakte mit Hilfe geeigneter Extraktorkomponenten zu erfassen und aufzubereiten sind. Wird das Konzept im Bereich der technischen PE eingesetzt, so empfiehlt sich die Integration des gesamten LFRP-Frameworks entweder in einen Prozessnavigator, wie er im Forschungsverbund FORFLOW entwickelt wurde [EFH⁺09], oder in ein PDMS. Da es beispielsweise Ziel eines PDMS ist, möglichst alle für die Entwicklung eines Produktes benötigten Informationen zu archivieren, werden hier bereits viele Informationen manuell gepflegt oder aus anderen Systemen — primär aus ERP- und CAD-Systemen — importiert. Damit würden sich nicht nur die erforderlichen Aktivitäten für den Informationsimport reduzieren sowie die bisherigen Suchfunktionalitäten eines PDMS verbessern lassen; vielmehr müssten Produktentwickler durch eine derartige Integration während ihrer Arbeit nicht noch ein weiteres, separates Werkzeug für die Suche benutzen, was die Akzeptanz des Systems fördert.

Ausblick auf weiteren Forschungsbedarf

Da im Rahmen der vorliegenden Arbeit nicht alle möglichen Aspekte und Probleme betrachtet werden konnten, lassen sich selbstverständlich immer noch Verbesserungspotentiale für das vorgestellte Indexierungsframework finden. Eines dieser Potentiale liegt beispielsweise im Bereich der Suche nach geometrisch ähnlichen Produkten. Während im vorgestellten Indexierungsframework die Gesamtähnlichkeit von Produkten in den Fokus der Untersuchung gestellt wurde, rückt die Ermittlung partieller Ähnlichkeiten insbesondere in der 3D-Forschung immer weiter in den Mittelpunkt der Betrachtung. Ziel der hierfür entwickelten Ansätze ist es, durch geeignete Repräsentations- und Vergleichsoperationen Objekte zu finden, die sich nicht gesamtheitlich, sondern lediglich in einzelnen Bestandteilen ähneln. Ein mögliches Szenario in der betrachteten Anwendungsdomäne, für das diese Art der Suche eine Rolle spielt, ist beispielsweise der Bedarf nach einem bestimmten Befestigungsclip für einen Kabelkanal. Zur Verdeutlichung zeigt Abbildung 8.2 unter a) ein Positivbeispiel eines Kabelkanals, das den besagten Befestigungsclip (rot umrandet) besitzt und unter b) und c) zwei Negativbeispiele. Durch die Indexierung und Bereitstellung einer Facette zur partiellen Ähnlichkeitssuche, die ein Konzept, wie es beispielsweise von Gal und

Cohen-Or [GC06] oder von Tierny et al. [TVD09] entwickelt wurde, nutzt, könnte der in Abbildung 8.2 unter a) dargestellte Kabelkanal gefunden und hierdurch der Wiederverwendungsgrad von Produkten bzw. Produktbestandteilen deutlich gesteigert werden.

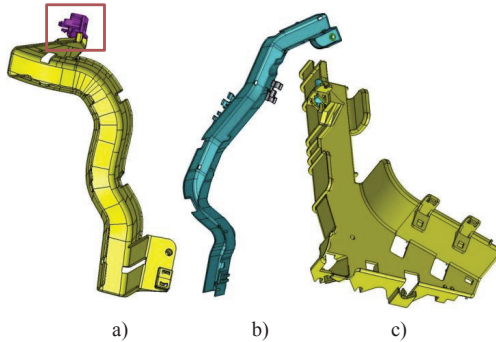


Abbildung 8.2. — Kabelkanäle mit a) einem bestimmten Befestigungsclip (rot umrandet) und b) bzw. c) jeweils ohne diesen speziellen Befestigungsclip

Damit verbunden besteht eine denkbare Erweiterung des LFRP-Frameworks in der Art und Weise, wie ein Nutzer eine Ähnlichkeitsanfrage stellen kann. Da jede Ähnlichkeitsanfrage ebenfalls in eine, für den Ähnlichkeitsvergleich geeignete Repräsentation zu transformieren ist, wirkt sich diese Erweiterung auch auf das Indexierungsframework aus. Manche seiner Module werden für diesen Zweck im Retrievalprozess des LFRP-Frameworks genutzt, so dass diese Module entsprechend anzupassen sind. Um eine Ähnlichkeitsanfrage stellen zu können, muss der Nutzer im vorgestellten Framework bisher bereits über ein CAD-Modell oder eine technische Zeichnung des Anfrageproduktes verfügen. In dem Fall, in dem der Nutzer nicht über ein solches Dokument verfügt, könnte er natürlich das Produkt beispielsweise in seinem CAD-System grob modellieren. Schneller und einfacher wäre es jedoch, wenn der Nutzer direkt beim Suchen die Möglichkeit hätte, das gesuchte Produkt oder die gesuchte Produktkomponente zu skizzieren. Demzufolge stellt die Integration eines Skizzierungstools, wie es beispielsweise von Min et al. [MCF02] für die in Kapitel 2.5.1 vorgestellte 3D Model Search Engine, von Igarashi et al. [IMT99] oder auch von Masry und Lipson [ML05] entwickelt wurde, eine sinnvolle und v. a. eine den vom Nutzer zu erbringenden Aufwand reduzierende Erweiterung dar.

Des Weiteren sind Verbesserungspotentiale auch im Hinblick auf die zu berücksichtigenden Dokumenttypen zu sehen, was insbesondere im Rahmen der Entwicklung mechatronischer Produkte von Bedeutung ist. Hier sind wichtige und wertvolle Informationen v. a. in Dokumenten aus dem Elektronikbereich wie Schaltplänen oder Lei-

terplattenlayouts enthalten, so dass diese mit ihren jeweiligen Dateiformaten genauer zu analysieren und in einer für das LFRP-Framework geeigneten Weise aufzubereiten und zu indexieren sind. Ein weiterer Aspekt, der dabei im Zusammenhang mit Dokumenten oder allgemein mit Informationen im vorgestellten Indexierungsframework nicht explizit betrachtet wurde, ist der Umgang mit mehreren Sprachen in einem Unternehmen. Um beispielsweise im Kontext einer heutzutage üblichen, verteilten PE die weltweite Zusammenarbeit zu fördern, sollte es möglich sein, standortübergreifend — eventuell auch länderübergreifend — nach Informationen suchen zu können. Demzufolge besteht eine notwendige Weiterentwicklung des LFRP- und damit auch des Indexierungsframeworks darin, der Forderung nach Mehrsprachenfähigkeit gerecht zu werden. Da allerdings die Betrachtung von Ansätzen und Möglichkeiten zur Realisierung dieser Forderung im Rahmen der hier vorgenommen Betrachtung zu umfangreich ist, soll lediglich auf die Notwendigkeit und auf den speziell auf diese Thematik fokussierten Forschungsbereich des *Cross Language Information Retrieval* verwiesen werden [ATO05].

Schließlich sind für den tatsächlichen Einsatz des vorgestellten Indexierungsframeworks in der unternehmerischen Praxis in jedem Fall die zur Facettenspeicherung benötigten Indexstrukturen detaillierter zu betrachten. Wie in Kapitel 6.2 angedeutet, ist dies insbesondere für die Ähnlichkeitsfacetten von Relevanz. Da die für sie verwendete Indexstruktur wesentlich die Suchzeit und damit die Wartezeit des Nutzers beeinflusst, ist sie somit als kritischer Erfolgsfaktor derjenige Optimierungsaspekt, dessen Umsetzung als erstes angestrebt werden sollte. Dazu sind verschiedene Indexstrukturen zu implementieren, zu evaluieren und daraus schließlich die am besten geeignete Indexstruktur auszuwählen.

Wie die hier exemplarisch betrachteten Themen an möglichen Verbesserungspotentialen zeigen, hält das LFRP-Framework mit seinem Indexierungs- und seinem Retrievalprozess auch für die Zukunft noch weitere interessante Forschungsthemen bereit.

A. Vergleich von Vorgehensmodellen

Die Tabellen A.1 und A.2 geben einen Überblick über gängige, in der PE angewendete Vorgehensmodelle zur Strukturierung des PEP, wobei die dargestellten Prozessphasen nicht notwendigerweise in starrer, sequentieller Reihenfolge ablaufen müssen [PBF07]. Obwohl sich die einzelnen VM in der Anzahl und im Detaillierungsgrad ihrer Phasen unterscheiden, beschreiben doch alle im Prinzip dasselbe Vorgehen, indem vom Groben zum Detail entwickelt wird.

Tabelle A.1. — Vergleich von Vorgehensmodellen der technischen PE durch Gegenüberstellen der zugehörigen Prozessphasen

VDI 2221-2223 [Ver86, Ver97, Ver99]	Pahl et al. [PBF07]	Koller [Kol98]	Münchner VM [Lin05]
Klären & Präzisieren der Aufgabenstellung	Planen & Klären der Aufgabe	Aufgabenstel- lung erarbeiten	Ziel planen
Ermitteln v. Funktionen & deren Strukturen	Konzipieren	Funktions- -synthese	Ziel analysieren
Suchen nach Lösungsprinzi- pien & deren Strukturen		Effektsynthese	Problem strukturieren
Gliedern in realisierbare Module			Lösungsideen
		Effektträger-	ermitteln
Gestalten der maßgebenden Module	Entwerfen	synthese	Eigenschaften ermitteln
Gestalten des gesamten Produkts		Qualitative Gestaltsynthese	Entscheidungen herbeiführen
Ausarbeiten der Aus- führungs- & Nutzungs- angaben	Ausarbeiten	Quantitative Gestaltsynthese	Zielerreichung absichern

Tabelle A.2. — Vergleich von Vorgehensmodellen der technischen PE durch Gegenüberstellen der zugehörigen Prozessphasen - Fortsetzung

FORFLOW [KEL ⁺ 09, EFH ⁺ 09]	VDI 2206 [Ver04, GR05]	Tjalve [Tja78] nach [And05b]	Roth [Rot00]
Präzisieren der Aufgabenstellung	Anforderungen	Problem- analyse	Aufgaben- formulierung
Ermitteln von Funktionen & deren Strukturen	Systementwurf	Hauptfunktionen ermitteln	Funktionelle Phase
Suchen nach Lösungsprin- zipien & deren Strukturen	Domänenspezifischer Entwurf	Teilfunktionen & Funktionsträger	Prinzipielle
	System- integration	Prinzipielle Struktur	Phase
Gesamtkonzept gestalten	Eigenschafts- absicherung	Quantitative Struktur	Gestaltende Phase
System- gestaltung	Modellbildung & -analyse	Form der Elemente	
Produktionsan- lauf & -betreuung	Produkt	Form des Gesamtsystems	

B. Attribute der Artefakttypen

B.1. Produkt

Alphabetische Auflistung der in der vorliegenden Arbeit erwähnten Produktattribute zusammen mit einer Definition dieser und der Angabe beispielhafter Ausprägungen in den Fällen, in denen mögliche Werte nicht aus dem Attribut selbst ersichtlich sind. Verwendete Quellen: [Sch02a, ABB⁺06, BS07, DK07, MWJV07, EFH⁺09, KMW09, BS10, Fer08, ACI99, Wik10c]

Attribut	Definition : <i>mögliche Ausprägungen</i>
2D-Geometrie- beschreibung	Beschreibung der 2D-Produktgeometrie aus techn. Zeichnung
2D-Topologie- beschreibung	Beschreibung der 2D-Produkttopologie aus techn. Zeichnung
3D-Geometrie- beschreibung	Beschreibung der 3D-Produktgeometrie aus CAD-Modell
Abgabemenge	Menge, in der ein Produkt verkauft wird : <i>1 Stück, 5 Meter</i>
Antriebsart	Beschreibt die Art der Energieumformung, die den Motor bewegt : <i>Verbrennung, Elektronik, Hybrid, Hydraulik</i>
Anzahl-BG	Anzahl der in einer Baugruppe enthaltenen Baugruppen : <i>5, 12, 26, 321</i>
Anzahl-ET	Anzahl der in einer Baugruppe enthaltenen Einzelteile : <i>1, 7, 47</i>
Außen-Ø	Wert für den Außendurchmesser
Baugröße	Definiert den Abstand von der Welle zur Auflage eines Elektromotors
Baujahr	Jahr, in dem ein Motor erbaut wurde
Beschaffungsart	Definiert, wie ein Produkt beschafft wird, d. h. ob eigen gefertigt oder eingekauft : <i>Eigenteil, Fremdteil</i>
Bestellmenge	Menge, zu der ein Produkt eingekauft werden kann : <i>1 Stück, 12 Stück</i>

Produktattribut	Definition : <i>mögliche Ausprägungen</i>
Betriebsdauer	Kennzahl, die beschreibt, wie lange ein Motor durchschnittlich funktioniert (Angabe z. B. in Stunden oder Kilometer)
Breite	Wert für die Breite des Bauteils
Bruchdehnung	Indikator für die Verformbarkeit der Schraube; sie gibt an um wie viel sich die Schraube im Zugversuch bis zum Bruch plastisch verlängert : <i>0.4 %</i> , <i>2 %</i>
CO2-Emission	Wert für die während des Motorbetriebes abgegebene Kohlenstoffdioxidmenge
DFX-Aspekte	Definieren die Gerechtheiten, die im Rahmen der Produktentwicklung und -fertigung zu beachten sind : <i>Montage-, Ergonomie-, Fertigungs-, Recycling-, Transportgerechtheit, ...</i>
Drehzahl	Anzahl der Umdrehungen der Welle
Eckmaß	Abstand zwischen den Enden zweier gegenüberliegender Angriffsflächen, an denen das Schraubwerkzeug ansetzt
Einbaulänge	Länge vom Ventilsitz bis zur Mitte des Einstiches am Ventilschaft
Einkaufspreis	Preis, zu dem das Produkt von einem Lieferanten bezogen werden kann : <i>0.55 €</i> , <i>51.13 €</i>
Einsatzzweck	Zweck, für den der Motor eingesetzt wird : <i>Land-, Wasser-, Luftfahrzeuge, Landwirtschaft, Industrieanlagen</i>
Einschraublänge	Länge, bis zu der die Schraube eingeschraubt werden muss, um tragfähig zu sein
Einspritzart	Art der Einspritzung : <i>direkte Einspritzung, indirekte Einspritzung, Saugrohreinspritzung</i>
Einspritzdüsenzahl	Anzahl der Einspritzdüsen für den Kraftstoff
Energiespeicher	Beschreibt, welches Bauteil zur Energiespeicherung verwendet wird : <i>Batterie, Akkumulator</i>
Fertigungszeit	Dauer der Produktion einer Mengeneinheit des Produktes (z. B. in Minuten)
Fertigungsverfahren	Art des zur Herstellung verwendeten Fertigungsverfahrens : <i>Urformen, Umformen, Trennen, Beschichten</i>
Festigkeitsklasse	Dient zur Beschreibung der Zugfestigkeit und Streckgrenze von Schrauben : <i>3.6, 4.6, 8.8, 10.9</i>
Funktion	Die von einem Produkt zu erfüllende(n) Funktion(en) : <i>steuern, antreiben, leiten, dichten, erwärmen, trennen, speichern, wandeln</i>
Gemischbildungs-ort	Ort, an dem die Gemischbildung stattfindet : <i>Einlasssystem, Arbeitsraum</i>
Gewicht	Gewicht eines Produktes (z. B. in Gramm)

Produktattribut	Definition : <i>mögliche Ausprägungen</i>
Gewindeausführung	Profilform des Gewindes, d. h. der Einkerbung, die längs der um einen Zylinder gewundenen Schraubenlinie verläuft : <i>Metrisches ISO-Gewinde, Whitworth-, Rohr-, Trapezgewinde</i>
Gewindelänge	Länge des Gewindes ohne Abschluss (z. B. ohne Kopf)
Gewinderichtung	Definiert die Drehrichtung des Gewindes : <i>rechts, links</i>
Herstellungskosten	Kosten der Herstellung von einer Mengeneinheit des Produktes
Höhe	Wert für die Höhe des Bauteils
Hubvolumen	Volumen des Raumes, den der Kolben bei einem Kolbenhub vom unteren Totpunkt bis zum oberen Totpunkt durchläuft : <i>1800 cm³, 1950 cm³</i>
Innen-Ø	Wert für den Innendurchmesser
Kinematik	Beschreibt die verwendete kinematische Version : <i>Drehkolben, Kreiskolben</i>
Kolbenbewegung	Definiert die Bewegungsart der Kolben : <i>Hub, Rotation</i>
Kopf-Außenprofil	Beschreibt die durch die Art des Kraftangriffs des Schraubwerkzeuges bedingte Außenform des Kopfes : <i>Sechskant, Zylinderkopf, Senkkopf, Flügelkopf</i>
Kopf-Ø	Durchmesser des Schraubenkopfes
Kopfhöhe	Höhe des Schraubenkopfes
Kopf-Innenprofil	Beschreibt die durch die Art des Kraftangriffs des Schraubwerkzeuges bedingte Innenform des Kopfes (auch als Antriebsart bezeichnet) : <i>Innensechskant, Innenzölzahn, Kreuzschlitz, Schlitz, Torx</i>
Kraftbewegungsart	Definiert die Bewegungsart der Kraftwirkung : <i>Drehbewegung, Längsbewegung</i>
Kraftstoff	Definiert den verwendeten Kraftstoff : <i>Biogas, Erdgas, Benzin, Diesel, Kohlenstaub</i>
Kraftstoffverbrauch	Wert für den durchschnittlichen Verbrauch an Kraftstoff : <i>6.2 l, 210 g/kWh</i>
Kühlungsart	Art der Kühlung, die zum Schutz der Bauteile erfolgt : <i>Luftkühlung, Flüssigkeitskühlung</i>
Lagerstatus	Gibt an, ob ein Produkt im Lager vorrätig ist oder nicht : <i>lagernd, nicht-lagernd</i>
Länge	Wert für die Länge des Bauteils
Läuferausführung	Definiert den Typ des eingesetzten Läufers bzw. Rotors : <i>Schenkelpol-, Vollpol-, Schleifring-, Kurzschlussläufer</i>
Läuferbewegung	Bewegungsart des Läufers bzw. Rotors: <i>synchron, asynchron</i>

Produktattribut	Definition : <i>mögliche Ausprägungen</i>
Leerlaufdrehzahl	Anzahl der Umdrehungen pro Minute eines Motors im Leerlauf : <i>700, 750, 900</i>
Lieferzeit	Dauer von der Bestellung bis zur Lieferung eines Produktes (z. B. in Tagen)
Lösbarkeit	Definiert, ob eine Dichtung lösbar ist : <i>lösbar, unlösbar, bedingt lösbar</i>
Magnetfeld- erzeugung	Beschreibt die Art der Magnetfelderzeugung : <i>Haupt-, Neben-, Compoundschluss, Permanenterregung</i>
Maulweite	Breite der Ringöffnung im nicht-eingebauten Zustand
Max. Zylinder- volumen	Maximales Zylindervolumen, das vorliegt, wenn sich der Kolben im unteren Totpunkt befindet
Nenn-Ø	Theoretisches Nennmaß für den Außendurchmesser zwischen den Gewindespitzen : <i>1.2, 1.6, 2, 8, 10</i>
Nenn-Ø Welle	Theoretisches Nennmaß für den Durchmesser der Welle
Nenn-drehzahl	Anzahl der Umdrehungen pro Minute, bei der ein Motor unter Vollast die größtmögliche Nennleistung erbringt : <i>1000, 2300, 4500</i>
Nenninnen-Ø	Theoretisches Nennmaß für den Innendurchmesser eines Zylinders
Nennleistung	Theoretisches Nennmaß für die von einem Motor abgegebene Leistung : <i>60 kW, 180 kW</i>
Nennstrom	Definiert den vom Motor aufgenommenen Strom bei Versorgung mit der Netzspannung
Netzspannung	Definiert, mit welcher Netzspannung der Motor betrieben werden kann
Nutzwirkungsgrad	Theoretisches Nennmaß für den Wirkungsgrad eines Motors : <i>27 %, 54 %, 85 %</i>
Ort der Dicht- pressung	Beschreibt an welchem Ort bzw. von wo die Dichtpressung erfolgt : <i>außen, innen</i>
Polpaarzahl	Anzahl der magnetischen Polpaare : <i>2, 4</i>
Produktgranu- larität	Beschreibt die Granularität eines Produktes : <i>Einzelteil, Baugruppe</i>
Produktgruppe	Gruppe, der ein Produkt angehört : <i>Dichtung, Feder, Motor, Pkw</i>
Produktlebens- zyklusstatus	Beschreibt, in welcher Prozessphase sich das Produkt befindet : <i>Entwurf, Geprüft, in Freigabe, Freigegeben, Gesperrt</i>
Produktname	Bezeichnung eines Produktes : <i>O-Ring4711, Schafschraube-XY</i>
Produktreifegrad	Beschreibt den zu einem bestimmten Zeitpunkt erfassten Zustand eines Produktes hinsichtlich zu erfüllender Anforderungen : <i>50 %, 80 %, 100 %</i>

Produktattribut	Definition : <i>mögliche Ausprägungen</i>
Produktversion	Definiert einen dokumentierten und jederzeit wiederherstellbaren Entwicklungsstand eines Produktes : <i>V1, V2, V3</i>
Rändelart	Art der Rändelung, d. h. der zur besseren Haptik an runden Teilen angebrachten Längsverzahnung : <i>Linksrändel, Rechtsrändel, Rändel mit achsparallelen Riefen</i>
Relativbewegungstyp	Definiert die Relativbewegung der abzudichtenden Gegenfläche als ruhend oder bewegt : <i>statisch, dynamisch</i>
Ringdicke	Dicke des Rings
Ringstoß	Breite der Ringöffnung im eingebauten Zustand
Sachnummer	Eindeutige Identifikationsnummer, die je nach verwendetem Nummernsystem auch der Klassifikation dient : <i>H16-28-03-028, A2A135007</i>
Schaft-Ø	Bei Schafschrauben Durchmesser des gewindefreien Schaftes; Bei Ventilen Durchmesser des zylindrischen Teils des Ventils (= Ventilschaft)
Schlüsselweite	Abstandsmaß der parallelen Angriffsflächen, an denen das Schraubwerkzeug ansetzt
Schraubenform	Beschreibt die Form der Schraube : <i>Kopf-, Schaft-, Stift-, Stockschraube</i>
Schraubenlänge	Gesamtlänge der Schraube, d. h. Gewindelänge plus Abschluss (z. B. Kopf)
Schraubennorm	Norm, in der die Eigenschaften der Schraube festgelegt sind : <i>DIN EN ISO 898-1, DIN ISO 4759, ISO 4762</i>
Schutzart	Kennzeichnet den Schutz eines Elektromotors gegen das Eindringen von Fremdkörpern (Wasser, Staub, ...) in Form sogenannter IP-Codes : <i>IP 12, IP 56</i>
Sitzwinkel	Kegelwinkel des Ventilsitzes, d. h. der Dichtfläche zwischen Kanal und Brennraum : <i>30°, 45°</i>
Streckgrenze	Gibt die Grenze der elastischen Dehnung einer Schraube an; bei Überschreiten längt sich die Schraube und geht schließlich zu Bruch
Stromart	Art des eingesetzten Stromes : <i>Gleich-, Wechsel-, Drehstrom</i>
Teller-Ø	Durchmesser der Unterseite des Ventilkopfes (= Ventilteller)
Totvolumen	Minimales Zylindervolumen, das vorliegt, wenn sich der Kolben im oberen Totpunkt befindet
Ventilanzahl	Anzahl der Auslass- und Einlassventile
Ventilwinkel	Bezeichnet den Winkel, den die Ventilachse mit der Zylinderachse einschließt : <i>18°, 20°, 70°</i>

Produktattribut	Definition : <i>mögliche Ausprägungen</i>
Ventilzahl-Brennraum	Anzahl der Ventile je Brennraum (mindestens zwei)
Verbindungstechnik	Definiert die Art der unlösbaren Dichtung : <i>Schweiß-, Klebe-, Lötverbindung</i>
Verbindungstyp	Beschreibt die Art der Verbindung, d. h. ob diese passgenau ist oder Luftspalten aufweist : <i>Passschraube, Durchsteckschraube</i>
Verkaufspreis	Preis, zu dem das Produkt am Markt bzw. den Kunden angeboten wird
Wirkprinzip	Beschreibt das Wirkprinzip der Dichtpressung : <i>Berührung, Berührungslos</i>
Zugfestigkeit	Definiert ab welcher Zugspannung eine Schraube brechen darf
Zündkerzenzahl	Anzahl der Zündkerzen
Zündungsart	Art der Zündung des Kraftstoff-Luft-Gemisches im Motor : <i>Selbstzündung, Fremdzündung</i>
Zylinderabstand	Abstand zwischen den Mitten von zwei benachbarten Zylindern
Zylinderachslage	Lage der Zylinderachse : <i>stehend, liegend, hängend</i>
Zylinderanordnung	Beschreibt die Anordnung der Zylinder : <i>Reihenmotor, V-Motor, Boxer-Motor, X-Motor, Sternmotor</i>
Zylinderhubvolumen	Hubraum eines Zylinders
Zylinderversatz	Abstand zwischen den Mitten von zwei sich in benachbarten Zylinderbänken gegenüberliegenden Zylindern
Zylinderzahl	Anzahl der Zylinder : <i>1, 2, 8, 12, 48</i>

B.2. Dokument

Alphabetische Auflistung der in der vorliegenden Arbeit erwähnten Dokumentattribute zusammen mit einer Definition dieser und der Angabe beispielhafter Ausprägungen in den Fällen, in denen mögliche Werte nicht aus dem Attribut selbst ersichtlich sind. Verwendete Quellen: [Sut96, ABB⁺04, EFH⁺09]

Attribut	Definition : <i>mögliche Ausprägungen</i>
Archivierungssystem	Ort der Aufbewahrung des Dokumentes : <i>Aktenschrank 01, Aktenschrank 02, PDMS, DMS</i>
AwS	System bzw. IT-Werkzeug, mit dessen Hilfe das Dokument erstellt wurde : <i>MS Word, MS Powerpoint, AutoCAD, Catia V4, Pro/ENGINEER</i>
Dateiformat	Format, in dem das Dokument gespeichert wurde : <i>pdf, txt, xml, jt, step, iges, dxf</i>
Dateigröße	Größe eines elektronischen Dokuments : <i>10 KB, 165 KB, 1367 KB</i>
Dokumentbezeichnung	Name oder Titel des Dokuments : <i>Pflichtenheft_Motor_V8, Schraube-830-01</i>
DokumentID	Eindeutige Identifikationsnummer für ein Dokument
Dokument-kategorie	Kategorie, der ein Dokument aufgrund seines Inhaltes zugeordnet werden kann : <i>Produktmodell, Projektdokument, Allgemeines Dokument</i>
Dokumentlebenszyklusstatus	Beschreibt den im Rahmen der Workflowsteuerung aktuell gültigen Lebenszyklusstatus eines Dokuments : <i>in Bearbeitung, in Änderung, in Prüfung, in Freigabe...</i>
Dokumentsprache	Sprache, in der der Dokumentinhalt verfasst ist : <i>Deutsch, Englisch, Spanisch</i>
Dokumenttyp	Definiert den Typ eines Dokumentes : <i>Pflichtenheft, CAD-Modell, Stückliste</i>
Dokumentversion	Dokumentiert den Änderungszustand : <i>V1, V2, V3</i>
Einrichtung	Organisation, Institut oder sonstige Einrichtung, die das Dokument verabschiedet oder als Standard festgelegt hat : <i>DIN, ISO, VDA, VDI, VDE,...</i>
Entwicklungsstand	Beschreibt den aktuellen Entwicklungsstand eines Produktes in Bezug auf die fertige Entwicklung : <i>0-20 %, 21-40 %, 41-60 %, 61-80 %, 81-100 %</i>

Attribut	Definition : <i>mögliche Ausprägungen</i>
Erstellungsdatum	Datum der Erstellung des Dokumentes : <i>05-06-2010, 23-02-1997</i>
Erscheinungsform	Definiert die Form des Dokumentes : <i>Papier, Elektronisch</i>
Erscheinungsjahr	Jahr, in dem das Dokument erschienen ist : <i>1990, 2007, 2009, 2010</i>
Inhaltsart	Beschreibt die Art des Inhaltes eines Produktmodells : <i>Bewertung/Berechnung, Anforderung, Geometrische Darstellung, Eigenschaftsbeschreibung, Lösungsideen</i>
Konkretisierungsgrad	Beschreibt das Entwicklungsstadium des beschriebenen Produktes : <i>Aufgabenklärung, Konzept, Entwurf, Detail, Freigabe</i>
ProjektID	Identifikationsnummer oder -name des zugehörigen Projekts : <i>Projekt-Motor_V8, Projekt-Bremssystem_2764</i>
Prozessphase	Phase des PEP, in der das Dokument erstellt wurde : <i>Ermitteln von Funktionen & Strukturen, Gliedern in Module, Modulgestaltung, Produktgestaltung, ...</i>
Sachverhalt	Der im Dokument adressierte Sachverhalt : <i>Sicherheit, Ergonomie, Umweltschutz, Transport, Qualität</i>
Vernetzungsgrad	Beschreibt den Grad der Informationsvernetzung zu anderen Produktmodellen : <i>1 $\hat{=}$ keine Vernetzung, 2, 3, 4, 5 $\hat{=}$ sehr hohe Vernetzung</i>
Verwendungszweck	Definiert den Zweck, dem ein Produktmodell im PEP dient : <i>Fertigung, Frontloading, Absicherung, Eigenschaftsermittlung, Lösungssuche</i>
Wirksamkeitsdatum	Datum, ab dem ein Dokument wirksam, d. h. gültig ist : <i>25-01-1993, 01-01-2009</i>
Zugriffsdatum	Datum des letzten Zugriffs auf das Dokument : <i>03-07-2010, 11-10-2002</i>

B.3. Werkstoff

Alphabetische Auflistung der in der vorliegenden Arbeit erwähnten Werkstoffattribute zusammen mit einer Definition dieser und der Angabe beispielhafter Ausprägungen in den Fällen, in denen mögliche Werte nicht aus dem Attribut selbst ersichtlich sind. Verwendete Quellen: [SP07, Con08c, MT08, MRB⁺08]

Attribut	Definition : <i>mögliche Ausprägungen</i>
Fertigungsverfahren	Definiert die Fertigungsverfahren, die für einen Werkstoff angewendet werden können : <i>Gießen, Schweißen, Löten, Walzen, Fräsen, . . .</i>
Verwendungszweck	Beschreibt den Zweck bzw. die Einsatzfelder, für die ein Werkstoff typischerweise verwendet wird : <i>Federring, Schraubensicherung, Federplatte, Spannmittel für den Oberbau</i>
Werkstoffgruppe	Die Werkstoffgruppe, der ein Werkstoff aufgrund seiner Eigenschaften zugeordnet wird : <i>Metall, Polymer, Elastomer, Verbundwerkstoff, Glas, Keramik</i>
Werkstoffname	Der Name des Werkstoffs : <i>Federstahl, Aluminiumoxid, Titan, Polyethylen</i>
Werkstoffnorm	Die Kurzbezeichnung der Norm, in der der Werkstoff standardisiert ist : <i>EN 10089, EN 10027, DIN 59145</i>
Werkstoffnummer	Dient der Klassifikation eines Werkstoffes : <i>1.5023, 2.1816, 3.7032</i>
Werkstoffschlüssel	Ein Kurzzeichen zur eindeutigen Identifikation eines Werkstoffes : <i>38Si7, ZrO2, SSiC, BeO, Mg</i>

B.4. Person

Alphabetische Auflistung der in der vorliegenden Arbeit erwähnten Personenattribute zusammen mit einer Definition dieser und der Angabe beispielhafter Ausprägungen in den Fällen, in denen mögliche Werte nicht aus dem Attribut selbst ersichtlich sind.

Attribut	Definition : <i>mögliche Ausprägungen</i>
Abteilung	Beschreibt die Abteilung, zu der ein Mitarbeiter im Unternehmen gehört
Anschrift-Land	Angabe des Ländernamens
Anschrift-Ort	Angabe des Ortsnamens
Anschrift-PLZ	Angabe der Postleitzahl des Ortes
Anschrift-Straße	Angabe des Straßennamens
Attraktivitätstyp	Beschreibt die Attraktivität eines Kunden hinsichtlich des durch ihn erzielten Umsatzanteils : <i>A-Kunde, B-Kunde, C-Kunde</i>
Aufträge	Enthält die bisher vom Kunden erteilten Aufträge : <i>Auftrag-011, Auftrag-027, Auftrag-138</i>
Ehem. Aufgaben	Beschreibt die von einer Person bereits in der Vergangenheit erfüllten Aufgaben
Branche	Branche, in der die juristische Person tätig ist : <i>Automobil, Halbleiterelektronik</i>
E-Mail-Adresse	E-Mail-Adresse der Person
Erfahrungsbereiche	Gibt die Bereiche an, in denen eine Person bereits Erfahrung besitzt
Kenntnisse	Beschreibt die Kenntnisse einer Person
Konditionen	Definiert die Zahlungskonditionen, die bei der Beschaffung einzuhalten sind : <i>14 Tage netto, 30 Tage netto</i>
Kundenklasse	Definiert den Status eines Kunden in Bezug auf die Häufigkeit seiner Auftragserteilung : <i>Neukunde, Stammkunde, Altkunde</i>
Kundennummer	Nummer zur eindeutigen Identifikation eines Kunden
Kundensegment	Teilgruppe der Kunden, die das gleiche Kaufverhalten zeigt : <i>Privatkunde, Geschäftskunde</i>
Lieferantennummer	Nummer zur eindeutigen Identifikation eines Lieferanten
MitarbeiterID	Eindeutige Identifikationsnummer für einen Mitarbeiter
Nachname	Nachname der Person

Attribut	Definition : <i>mögliche Ausprägungen</i>
Name	Bezeichnung der juristischen Person : <i>Robert Bosch GmbH, Siemens AG</i>
Positionierung	Beschreibt die Position eines Lieferanten innerhalb des Zulieferernetzwerks: <i>Rohstoff-, Tier-1-, Tier-2-Lieferant</i>
Rechtssubjektgruppe	Einordnung der Person in die Gruppe der natürlichen und juristischen Personen
Rolle	Rolle, die in Mitarbeiter im Unternehmen einnimmt : <i>Konstrukteur, Tester, Fertigungsplaner</i>
SC-Rolle	Rolle, die eine Person in der (einfachen) Lieferkette einnimmt : <i>Lieferant, Unternehmensmitarbeiter, Kunde</i>
Telefonnummer	Telefonnummer der Person
Vorname	Vorname der Person

B.5. Projekt

Alphabetische Auflistung der in der vorliegenden Arbeit erwähnten Projektattribute zusammen mit einer Definition dieser und der Angabe beispielhafter Ausprägungen in den Fällen, in denen mögliche Werte nicht aus dem Attribut selbst ersichtlich sind. Verwendete Quellen: [HHMS06, Stö09]

Attribut	Definition : <i>mögliche Ausprägungen</i>
Beschreibung	Detaillierte Beschreibung des Projektes, d. h. seiner Inhalte, Aufgaben usw.
Budget	Menge an Geld, die für die Projektdurchführung zur Verfügung steht
DfX-Strategie	Definiert die für ein Projekt zu verfolgende DfX-Strategie : <i>Leichtbau, geringe Kosten, hohe Qualität</i>
Endtermin	Termin des (voraussichtlichen) Projektendes : <i>30-06-2009, 31-12-2002</i>
Kosten	Menge an Geld, die im Lauf der Projektdurchführung tatsächlich benötigt wurde
Kurzbezeichnung	Kurzbezeichnung für ein Projekt : <i>Pr0001, Pr0002, Pr4711</i>
Projektart	Definiert, um welche Konstruktionsart bzw. -aufgabe es sich im Rahmen des Projektes handelt : <i>Neu-, Anpassungs-, Variantenkonstruktion</i>
Projektfortschritt	Kennzahl, die angibt, wie weit fortgeschritten ein Projekt ist bzw. wie weit es das Soll erfüllt : <i>30 %, 60 %, 75 %</i>
Projektname	Name bzw. Titel eines Projektes : <i>Projekt-V8-Motor, Projekt-Chassis, Projekt-Fahrgestell</i>
Starttermin	Termin, an dem das Projekt startet oder gestartet ist : <i>01-06-2009, 15-02-1999</i>
Ziele	Beschreibt das bzw. die mit der Projektdurchführung angestrebten Ziele : <i>Entwicklung Hybridmotor, Max. 7 Monate, Max. 280 Manntage</i>
Zeitdauer	Beschreibt die Dauer eines Projektes als Zeitspanne zwischen End- und Starttermin (z. B. in Tagen) : <i>30, 84, 122</i>

C. Exemplarische XML-Datei für den Datenimport

Beispiel einer XML-Datei für den Datenimport von Artefaktinformationen aus einem PDMS mit der ID „S001“. Der dargestellte Ausschnitt verdeutlicht die Definition von drei Artefakten — einem Produkt, einem Werkstoff und einem Dokument — gemäß der in Kapitel 4.4 in Listing 4.1 beschriebenen DTD.

beispielimport.xml

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE artefaktimport SYSTEM ".../artefaktimport.dtd">
<Artefaktimport>
  <!-- Organisatorische Daten -->
    <Informationsquelle quelleID="S001" quelletyp="PDM">
      <Zeitstempel>2010-06-21T22:19:05</Zeitstempel>
    </Informationsquelle>
  <!-- Auflistung der Artefakte -->
    <Artefakte>
      <!-- Definition eines Produktes -->
        <Artefakt artefakttyp="Produkt" artefaktID="A2A800358"
          ssIdent="S001-A2A800358" artefaktURI="p://
            S001/Produkte/S001-A2A800358">
          <Metadatum>
            <Name>Produktname</Name>
            <Wert>O-Ring01</Wert>
          </Metadatum>
          <Metadatum>
            <Name>Produktgruppe</Name>
            <Wert>Dichtung</Wert>
          </Metadatum>
          <Metadatum>
            <Name>Dichtungstyp</Name>
            <Wert>O-Ring</Wert>
          </Metadatum>
          <Metadatum>

```

```

        <Name>Beschaffungsart </Name>
        <Wert>Eigenteil </Wert>
    </Metadatum>
    <Metadatum>
        <Name>Produktgranularität </Name>
        <Wert>Einzelteil </Wert>
    </Metadatum>
    <Metadatum>
        <Name>Funktion </Name>
        <Wert>dichten </Wert>
    </Metadatum>
    <Metadatum>
        <Name>Außenradius </Name>
        <Wert>4.5 </Wert>
    </Metadatum>
</Artefakt>
<!-- Definition eines Werkstoffs -->
<Artefakt artefakttyp="Werkstoff" artefaktID="PE-LD"
    ssIdent="S001-PE-LD" artefaktURI="p://S001/
    Werkstoffe/S001-PE-LD">
    <Metadatum>
        <Name>Werkstoffname </Name>
        <Wert>Polyethylen </Wert>
    </Metadatum>
    <Metadatum>
        <Name>Werkstoffgruppe </Name>
        <Wert>Polymer </Wert>
    </Metadatum>
    <Metadatum>
        <Name>Dichte </Name>
        <Wert>0.92 </Wert>
    </Metadatum>
    <Metadatum>
        <Name>E-Modul </Name>
        <Wert>1000 </Wert>
    </Metadatum>
    <Assoziation assoziationstyp="Werkstoff-verwendet-
        in-Produkt">
        <Assoziationselement artefaktID="A2A800358"/>
    </Assoziation>
</Artefakt>
<!-- Definition eines Dokumentes -->
<Artefakt artefakttyp="Dokument" artefaktID="D-CM-0001">

```

```

        ssIdent="S001-D-CM-0001" artefaktURI="p://
        S001/Dokumente/O-Ring-Carrlane-R3335SW_2.wrl
    ">
    <Metadatum>
        <Name>Dokumenttyp</Name>
        <Wert>CAD-Modell</Wert>
    </Metadatum>
    <Metadatum>
        <Name>Dateiformat</Name>
        <Wert>wrl</Wert>
    </Metadatum>
    <Metadatum>
        <Name>Erstellungsdatum</Name>
        <Wert>2009-02-17T10:44:20</Wert>
    </Metadatum>
    <Metadatum>
        <Name>Dokumentlebenszyklusstatus</Name>
        <Wert>freigegeben</Wert>
    </Metadatum>
    <Assoziation assoziationstyp="Dokument-beschreibt -
        Produkt">
        <Assoziationselement artefaktID="A2A800358"/>
    </Assoziation>
    <Assoziation assoziationstyp="Dokument-erstellt -
        in-Projekt">
        <Assoziationselement artefaktID="Projekt_Chassis"/>
    </Assoziation>
    <Assoziation assoziationstyp="Dokument-erstellt -
        von-Person">
        <Assoziationselement artefaktID="NW-BA"/>
    </Assoziation>
    </Artefakt>
</Artefakte>
</Artefaktimport>

```


D. XML-Schema für das Facettenschema

facetSchema.xsd

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.forflow.org"
  xmlns="http://www.forflow.org"
  elementFormDefault="qualified">

  <xs:annotation>
    <xs:documentation xml:lang="en">
      FORFLOW facet schema for index and query framework.
    </xs:documentation>
  </xs:annotation>

  <xs:element name="facetdescription" type="
    FacetDescriptionType" />

  <xs:complexType name="FacetDescriptionType">
    <xs:sequence>
      <xs:element name="facets" type="FacetsType" />
      <xs:element name="entities" type="EntitiesType" />
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="FacetsType">
    <xs:sequence>
      <xs:element name="facet" type="FacetDefinitionType"
        maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="FacetDefinitionType">
    <xs:sequence>
      <xs:element name="visual" type="VisualType"
```



```

        minOccurs="0" maxOccurs="1" />
    <xs:element name="synset" type="ComparableTerms"
        minOccurs="0" maxOccurs="1" />
    <xs:element name="rule" type="CalculationRule"
        minOccurs="0" maxOccurs="1" />
</xs:sequence>
<xs:attribute name="facetId" type="xs:string"
    use="required" />
<xs:attribute name="facetType" type="FacetType"
    use="required" />
<xs:attribute name="multiIndexable" type="xs:boolean"
    use="required"/>
<xs:attribute name="similarityType" type="SimilarityType"
    use="optional" />
<xs:attribute name="documentType" type="DocumentType"
    use="optional" />
<xs:attribute name="weightingEnabled" type="xs:boolean"
    />
</xs:complexType>

<xs:complexType name="CalculationRule">
    <xs:sequence>
        <xs:element name="term" type="CalculationTerm"
            minOccurs="1" maxOccurs="3"/>
    </xs:sequence>
    <xs:attribute name="operator" type="OperatorType"
        use="required" />
</xs:complexType>

<xs:simpleType name="OperatorType">
    <xs:restriction base="xs:string">
        <xs:enumeration value="DIFFERENCE" />
        <xs:enumeration value="SUM" />
        <xs:enumeration value="PRODUCT" />
        <xs:enumeration value="RATIO" />
        <xs:enumeration value="IF-THEN-ELSE" />
        <xs:enumeration value="IDENTITY" />
    </xs:restriction>
</xs:simpleType>

<xs:complexType name="CalculationTerm">
    <xs:choice>

```

```
<xs:element name="rule" type="CalculationRule"
  maxOccurs="1"/>
<xs:element name="reference" type="ReferenceType"
  maxOccurs="1" />
<xs:element name="argumentDigit" type="xs:double"
  maxOccurs="1" />
<xs:element name="argumentText" type="xs:string"
  maxOccurs="1" />
</xs:choice>
</xs:complexType>

<xs:complexType name="ReferenceType">
  <xs:sequence>
    <xs:element name="constraint" type="ConstraintType"
      maxOccurs="1"/>
  </xs:sequence>
  <xs:attribute name="facetId" type="xs:string"
    use="required" />
  <xs:attribute name="entityId" type="xs:string"
    use="optional" />
</xs:complexType>

<xs:complexType name="ConstraintType">
  <xs:attribute name="constraintOperator"
    type="ConstraintOpType" use="required" />
  <xs:choice>
    <xs:element name="regExpression" type="xs:string"
      maxOccurs="1" />
    <xs:element name="reference" type="ReferenceType"
      maxOccurs="1" />
    <xs:element name="argumentDigit" type="xs:double"
      maxOccurs="1" />
    <xs:element name="argumentText" type="xs:string"
      maxOccurs="1" />
  </xs:choice>
</xs:complexType>

<xs:simpleType name="ConstraintOpType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="EQUAL" />
    <xs:enumeration value="GREATER" />
    <xs:enumeration value="SMALLER" />
```

```

    <xs:enumeration value="UNEQUAL" />
  </xs:restriction>
</xs:simpleType>

<xs:complexType name="ComparableTerms">
  <xs:sequence>
    <xs:element name="compTerm" type="xs:string"
      minOccurs="1" maxOccurs="unbounded" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="VisualType">
  <xs:sequence>
    <xs:element name="comparators" type="ComparatorsType"
      />
    <xs:element name="visualizationtypes"
      type="VisualizationTypes" />
  </xs:sequence>
  <xs:attribute name="name" type="xs:string"
    use="required" />
  <xs:attribute name="iconPath" type="xs:string" />
  <xs:attribute name="multiValued" type="xs:boolean"
    use="required"/>
  <xs:attribute name="multiValueOp">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="AND" />
        <xs:enumeration value="OR" />
        <xs:enumeration value="AND_OR" />
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>

<xs:complexType name="ComparatorsType">
  <xs:sequence>
    <xs:element name="comparator" type="ComparatorType"
      maxOccurs="unbounded" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="ComparatorType">

```

```
<xs:attribute name="default" type="xs:boolean" />
<xs:attribute name="classname" type="xs:string"
  use="required"/>
</xs:complexType>

<xs:complexType name="VisualizationTypes">
  <xs:sequence minOccurs="1" maxOccurs="unbounded">
    <xs:element name="visualizationtype"
      type="VisualizationType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="EntitiesType">
  <xs:sequence>
    <xs:element name="entity" type="EntityType"
      maxOccurs="unbounded" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="EntityType">
  <xs:sequence>
    <xs:element name="facet" type="FacetRefType"
      minOccurs="0" maxOccurs="unbounded" />
    <xs:element name="relationfacet"
      type="RelationFacetType"
      minOccurs="0" maxOccurs="unbounded" />
    <xs:element name="dependency" type="DependencyType"
      minOccurs="0"/>
    <xs:element name="parent-entity">
      <xs:complexType>
        <xs:attribute name="entityId" type="xs:string"
          use="required" />
      </xs:complexType>
    </xs:element>
  </xs:sequence>
  <xs:attribute name="entityId" type="xs:string"
    use="required" />
  <xs:attribute name="name" type="xs:string"
    use="required" />
  <!-- Declares this entity as root element of the
  hierarchy. -->
  <!-- Only one entity should be root -->
```

```

    <xs:attribute name="root" type="xs:boolean"
                  default="false" />
</xs:complexType>

<xs:complexType name="FacetRefType">
  <xs:attribute name="required" type="xs:boolean"
                use="required"/>
  <xs:attribute name="facetId" type="xs:string"
                use="required"/>
</xs:complexType>

<xs:complexType name="RelationFacetType" >
  <xs:complexContent>
    <xs:extension base="FacetRefType">
      <xs:attribute name="toEntityId" type="xs:string"/>
      <xs:attribute name="toEntityFacetId"
                    type="xs:string"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

<xs:complexType name="DependencyType">
  <xs:sequence>
    <xs:element name="from">
      <xs:complexType>
        <xs:attribute name="facetId" type="xs:string"
                      use="required"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="mapping" type="DependencyMappingType"
                maxOccurs="unbounded" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="DependencyMappingType">
  <xs:sequence>
    <xs:element name="value" type="xs:string" />
    <xs:element name="to" type="xs:string" />
  </xs:sequence>
</xs:complexType>

<xs:simpleType name="FacetType">

```

```
<xs:restriction base="xs:string">
  <xs:enumeration value="NOMINAL" />
  <xs:enumeration value="ORDINAL" />
  <xs:enumeration value="FUNCTION_INTERVAL" />
  <xs:enumeration value="DATE" />
  <xs:enumeration value="TEXT_QUERY" />
  <xs:enumeration value="QUERY_BY_EXAMPLE" />
</xs:restriction>
</xs:simpleType>

<xs:simpleType name="DocumentType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="TEXT" />
    <xs:enumeration value="CAD_MODEL" />
    <xs:enumeration value="TECHNICAL_DRAWING" />
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="SimilarityType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="TEXT" />
    <xs:enumeration value="GEOMETRY_2D" />
    <xs:enumeration value="GEOMETRY_3D" />
    <xs:enumeration value="TOPOLOGY_2D" />
    <xs:enumeration value="TOPOLOGY_3D" />
  </xs:restriction>
</xs:simpleType>

<xs:complexType name="VisualizationType">
  <xs:attribute name="facetType" type="VisualFacetType" />
  <xs:attribute name="default" type="xs:boolean"
    default="false" />
</xs:complexType>

<xs:simpleType name="VisualFacetType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="NOMINAL" />
    <xs:enumeration value="ORDINAL" />
    <xs:enumeration value="FUNCTION_INTERVAL" />
    <xs:enumeration value="DATE" />
  </xs:restriction>
</xs:simpleType>
```

```
</xs: schema>
```

E. Ergebnisse der Dokumenttypanalyse

Sammlung und Analyse der während des PEP erstellten und benötigten Dokumenttypen hinsichtlich ihrer Zuordnung zu den einzelnen Entwicklungsphasen sowie ihres Informationsgehalts.

Die Abbildungen E.1 bis E.6 veranschaulichen die Zuordnung der verschiedenen Dokumenttypen zu den Phasen des im Forschungsverbund entwickelten FORFLOW-Prozessmodells [EFH⁺09, KEL⁺09], welches das vierstufige Phasenmodell von Pahl et al. [PBFG07] detailliert und um diverse Aspekte, wie beispielsweise eine situationsspezifische Prozessplanung, die Unterstützung von DfX-Aspekten oder eine Absicherung auf mehreren Ebenen, erweitert. Im Rahmen dieser Zuordnung erfolgte gleichzeitig eine Charakterisierung der Dokumenttypen in Bezug auf die Rolle, die diese für den PEP spielen, was durch die in Tabelle E.1 näher erläuterten Kurzbezeichnungen erfolgt. Diese geben Aufschluss darüber, in welcher Phase ein Dokument eines bestimmten Typs erstellt werden sollte oder spätestens zu erstellen ist, wann es geändert werden kann bzw. muss¹¹³ oder in welcher Phase es eine hilfreiche Informationsquelle darstellt. Darüber hinaus stellen die in den Abbildungen E.1 bis E.6 grau hinterlegten Dokumenttypen zwar keine spezifischen Dokumenttypen der PE dar, werden aber auch hier für diverse Aufgaben verwendet.

In den Abbildungen E.7 bis E.10 wird ein Überblick über die in den verschiedenen Dokumenttypen enthaltenen Informationen gegeben. Dabei sind die grau hinterlegten Dokumenttypen als allgemeine Dokumenttypen zu verstehen, die abhängig von ihrem Verwendungszweck unterschiedliche Informationen enthalten können. Folglich wird eine spezifische Einordnung in die in den Abbildungen betrachteten Kategorien als nicht sinnvoll erachtet. Des Weiteren muss ein bestimmter Dokumenttyp nicht notwendigerweise alle Kategorien enthalten; vielmehr verdeutlicht ein in Klammern stehendes x , dass Daten der jeweiligen Kategorie prinzipiell enthalten sein können.

¹¹³Gekennzeichnet dadurch, dass ein Dokument in einer Phase sowohl Input- als auch Outputdokument sein kann.

Tabelle E.1. — Bei der Zuordnung von Dokumenttypen zu Prozessphasen verwendete Kurzbezeichnungen und ihre Erklärung

Kurzbezeichnung	Erklärung
Im	Muss-Input aus dem aktuellen Projekt ($\hat{=}$ Muss-Output einer vorangegangenen Phase, Muss-Input intern)
ImH	Muss-Input aus vorangegangenen Projekten oder von Stellen außerhalb der PE-Abteilung, Strategieabteilung, Lieferant, ... ($\hat{=}$ Muss-Input extern)
I	Optionaler Input aus dem aktuellen Projekt ($\hat{=}$ Kann-Input intern)
IH	Optionaler Input aus vorangegangenen Projekten oder von Stellen außerhalb der PE-Abteilung ($\hat{=}$ Kann-Input extern)
Om	Muss-Output im aktuellen Projekt ($\hat{=}$ Muss-Output intern)
O	Optionaler Output im aktuellen Projekt ($\hat{=}$ Kann-Output intern)

FORFLOW - Prozessphase		Schritt 1	Schritt 2	Schritt 3	Schritt 4: Gesamtkonzept gestalten		
Artefakt	Klären & Präzisieren der Aufgabenstellung	Ermitteln v. Funktionen & Strukturen	Suche Lösungsprinzipien & Strukturen	Modularisieren	Großgestaltung	Eigenschaftsabsicherung / -analyse	Konzeptauswahl
Änderungsdokumente	O	O	O	O	O	O	O
Arbeitsanweisungen	ImH	ImH	ImH	ImH	ImH	ImH	ImH
Baugruppenbeschreibungen				IH, Om	IH, O, I	IH, O, I	IH, O, I
Berechnungsergebnisse				IH, O	Om, I, IH	O, I, IH	IH, I, O
Betriebsdaten (geforderte, tatsächliche)	IH, Om	IH, Im		IH, Im	Im	Im	Im
Bewertungsergebnisse	IH, O	IH, I, O		IH, I, Om	Im, IH	I, IH, Om	IH, Im, Om
CAD-Modell (v.a. 3D)				IH	IH, O	I	I, IH
Checklisten	IH	IH	I, IH, O	I, IH, O	I, IH, O	I, IH, O	I, IH, O
Einflussmatrix				IH, O	I	I, O	I
Fachliteratur / -zeitschriften (+ Internetrecherchen)	IH	IH		IH	IH	IH	IH
Fertigungsprozessplan					IH	IH	IH
Technische Zeichnung: Fertigungszeichnung							
Technische Zeichnung: Zusammenstellungszeichnung					IH	IH	
FMEA-Analyse-Dokument					IH, O	IH, O	I
Funktionsmodell / -struktur		IH, O bzw. Om	I bzw. Im, O	I bzw. Im	I bzw. Im	I bzw. Im	I
Gestaltstudie				IH, O	I, O	I	I
Ideenformular				IH, O	I		
Konstruktions-/Funktionskataloge		IH		IH	IH		
Konzeptbeschreibung / -heft				IH, O	I	I, O	I, Om
Kostenkalkulation (Teilkostenkalk. Entwickler)	IH			IH, O	IH, O	I, Om	Im
Kundenauftrag	Im, IH						
Lastenheft	Im, O	Im, O		Im, O	Im, O	Im, O	Im
Leistungsprofil				IH			
Lessons Learned / Best Practices	IH, O	IH, O		IH, O	IH, O	IH, O	IH, O
Lieferanten-Datenblätter				IH		IH bzw. ImH	
Lösungsbeschreibungen / Prinzipielle Lösung				IH, Om	O, IH, I	Im	Im, Om

Abbildung E.1. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 1

FORFLOW - Prozessphase	Schritt 5: Systemgestaltung					Schritt 6
	Bauteil-gestaltung	Bauteil-absicherung	Teilsystem-gestaltung	Teilsystem-absicherung	Gesamtsystem-gestaltung	Gesamtsystem-absicherung
Artefakt						
Änderungsdokumente	O	O	O	O	O	O
Arbeitsanweisungen	ImH	ImH	ImH	ImH	ImH	ImH
Baugruppenbeschreibungen	Im, O	Im, O	Im, O, IH	Im, O, IH	Im, O	I, O
Berechnungsergebnisse	O, IH, Im	Om, Im, IH	O, IH, Im	Om, Im, IH	Om, Im, IH	
Betriebsdaten (geförderte, tatsächliche)	Im	Im, O	Im	Im, O	Im, O	Im
Bewertungsergebnisse	Im, IH	Im, IH, Om	Im, IH	Im, IH, Om	Im, IH, Om	O
CAD-Modell (v.a. 3D)	I, IH, Om	Im, IH	I, IH, Om	Im, IH	I, IH, Om	I
Checklisten	I, IH, O	I, IH, O	I, IH, O	I, IH, O	I, IH, O	I, IH, O
Einflussmatrix	I	I	I	I, O	I	I, O
Fachliteratur / -zeitschriften (+ Internetrecherchen)	IH	IH	IH	IH	IH	IH
Fertigungsprozessplan	IH, O	Im, Om	IH, Im, O	Im, Om	IH, Im, O	Im, O
Technische Zeichnung: Fertigungszeichnung	IH, Om	Im	I	Im	I	Im, O
Technische Zeichnung: Zusammenstellungszeichnung			IH, Om	Im	IH, Om	Im
FMEA-Analyse-Dokument		I, IH, O		I, IH, O		I, IH, O
Funktionsmodell / -struktur	I bzw. Im	I bzw. Im	I bzw. Im	I bzw. Im	I bzw. Im	
Gestaltstudie	I, O	I, O	I, O		I, O	
Ideenformular	I, O					
Konstruktions-/Funktionskataloge	IH					
Konzeptbeschreibung / -heft	Im		Im		Im	
Kostenkalkulation (Teilkostenkalk. Entwickler)	O, Im, IH	Im	O, Im, IH	Im	O, Im, IH	Im
Kundenauftrag						
Lastenheft						
Leistungsprofil		Om		Om		Om
Lessons Learned / Best Practices	IH, O	IH, O	IH, O	IH, O	IH, O	IH, O
Lieferanten-Datenblätter	IH	IH	IH	IH	IH	IH
Lösungsbeschreibungen / Prinzipielle Lösung	I, Om	I, Om	I, Om		I, Om	

Abbildung E.2. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 1

FORFLOW - Prozessphase		Schritt 1	Schritt 2	Schritt 3	Schritt 4: Gesamtkonzept gestalten		
Artefakt	Klären & Präzisieren der Aufgabenstellung	Ermitteln v. Funktionen & Strukturen	Suche Lösungsprinzipien & Strukturen	Modularisieren	Großgestaltung	Eigenschaftsabsicherung / -analyse	Konzeptauswahl
Machbarkeitsstudie	O, IH	O, IH		O, IH	IH	Om	Im
Marktanalyseergebnisse	IH, Om						I
Materialliste	IH			O	O	I	
Meilensteindokumentation (Meilensteinmatrix)	Om, IH	Om, Im	Om, Im	Om, Im	Om, Im	Om, Im	Om, Im
Messergebnisse	IH	IH	IH	IH	IH	IH	
Modulare Struktur	IH	IH	IH	IH	IH, Om	I, IH	Om
Montage-/Fertigungsplan			IH	IH	IH	IH	
Morphologischer Kasten			IH, O	I	I		I
Nachweis Konzepttauglichkeit			IH	IH	IH	Om	Im
Packaging			IH	IH, O	I, IH, Om	IH, Im	Im, IH
Patente	ImH	IH	IH	O	O	O	ImH, O
Pflichtenheft	IH	IH	IH, O	Im, O	O	O	Om
Präsentationen	O	O	O	O	O	O	O
Prinzipzeichnungen (Skizzen)	IH	IH, O	I, IH, O	I, IH, O	I, IH, O	I	I
Produktdokumentation (Bedienungs-, Wartungs-, Montageanleitung, ...)							
Produktlogik			IH, O	IH, O	Om	Im, O	Im
Produktstrukturplan			IH, O	IH, O	Om	Im, O	Im
Projektauftrag	IH, Om	I	I	I	I	I	I
Projektdokumentation	Om	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om
Projektplan	Om	Im	Im	Im	Im	Im	Im
Prototypen	IH	IH	IH	IH	IH	O	I
Prüfanweisungen	IH				O, IH	I	I
Prüfplan	IH				O, IH	I	I
Richtlinien/Normen/Gesetze/Vorschriften (Unternehmens-, Umwelt-, Sicherheits-, ...)	ImH	ImH	ImH	ImH	ImH	ImH	ImH

Abbildung E.3. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 2

FORFLOW - Prozessphase Artefakt	Schritt 5: Systemgestaltung					Schritt 6 Produktions- anlauf & -be- treuung
	Bauteil- gestaltung	Bauteil- absicherung	Teilsystem- gestaltung	Teilsystem- absicherung	Gesamtsystem- gestaltung	Gesamtsystem- absicherung
Machbarkeitsstudie						
Marktanalysergebnisse						
Materialliste	I, Om	Im	Im	Im	Im	Im
Meilensteindokumentation (Meilensteinmatrix)	Om, Im	Om, Im	Om, Im	Om, Im	Om, Im	Om, Im
Messergebnisse		Om	I, IH	Im, Om	I, IH	Im, Om
Modulare Struktur	Im	I	Im	I	Im	I
Montage-/Fertigungsplan			IH, Om	Im	IH, Im, Om	Im, O
Morphologischer Kasten	I		I		I	
Nachweis Konzepttauglichkeit						
Packaging	Im, O	Im	Im, Om	Im	Im, Om	Im
Patente	ImH, O	ImH	ImH, O	ImH	ImH, O	ImH, O
Pflichtenheft	Im	Im	Im	Im	Im	Im
Präsentationen	O	O	O	O	O	O
Prinzipzeichnungen (Skizzen)	I	I	I	I	I	
Produktdokumentation (Bedienungs-, Wartungs-, Montageanleitung, ...)	Om	Im	Om	Im	Om	Om
Produktlogik	I, O	Im	Im, O	Im	Im, O	Im
Produktstrukturplan	I, O	Im	Im, O	Im	Im, O	Im, O
Projektauftrag	I	I	I	I	I	I
Projektdokumentation	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om	I, IH, Om
Projektplan	Im	Im	Im	Im	Im	Im
Prototypen	I, IH	O	I, IH	O	I, IH	O
Prüfanweisungen	Om	Im	Om	Im	Om	Im, O
Prüfplan	I, Om	I	I, Om	I	I, Om	Im
Richtlinien/Normen/Gesetze/Vorschriften (Unternehmens-, Umwelt-, Sicherheits-, ...)	ImH	ImH	ImH	ImH	ImH	ImH

Abbildung E.4. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 2

FORFLOW - Prozessphase Artefakt	Schritt 1 Klären & Präzisieren der Aufgabenstellung	Schritt 2 Ermitteln v. Funktionen & Strukturen	Schritt 3 Suche Lösungsprin- zipien & Strukturen	Schritt 4: Gesamtkonzept gestalten			
				Modulari- sieren	Großgestal- tung	Eigenschafts- absicherung / -analyse	Konzeptplaus- wahl
Schaltpläne			IH	IH	IH, O	I	I, IH
Simulations- & Berechnungsergebnisse (Reports)	IH		IH	IH	IH	Om	Im
spezifische Anforderungsliste	ImH, Om	Im, O	Im, O	Im, O	Im, O	Im, O	Im, Om
Stücklisten			IH	IH	IH		
Teilekataloge			IH	IH	IH	IH	
Terminplan	Om	Im	Im	Im	Im	Im	Im
Vertriebsunterlagen (des eigenen Unternehmens)	I, O						
Design / Layout (= graphische Produktbeschreibung)	IH	IH	IH	IH	IH, O	IH, I	IH, I
Recyclingplan	IH		IH	IH	IH, Om		IH

Abbildung E.5. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 3

FORFLOW - Prozessphase Artefakt	Schritt 5: Systemgestaltung					Schritt 6
	Bauteil- gestaltung	Bauteil- absicherung	Teilsystem- gestaltung	Teilsystem- absicherung	Gesamtsystem- gestaltung	Gesamtsystem- absicherung
Schaltpläne	Om, I, IH	Im, IH	Om, I, IH	Im, IH	Om, I, IH	Im, IH
Simulations- & Berechnungsergebnisse (Reports)	I, IH, O	Om, I	I, IH, O	Om, I	I, IH, O	Om, I
spezifische Anforderungsliste	Im, O	Im, O	Im, O	Im, O	Im, O	Im, O
Stücklisten			Om	Im	Om	Im
Teilekataloge	IH	IH	IH	IH	IH	IH
Terminplan	Im	Im	Im	Im	Im	Im
Vertriebsunterlagen (des eigenen Unternehmens)						
Design / Layout (= graphische Produktbeschreibung)	IH, I, O	IH, I	IH, I, O	I, IH	IH, I, O	I, IH
Recyclingplan	Im, O, IH	Im	Im, O, IH	Im	Im, O, IH	Im

Abbildung E.6. — Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 3

Inhalt		Metadaten (Formate)						
Produkt- tun und best.- funktionen & Lösungen	Geome- treten struktur	Topologie Produkt- struktur	Ziehungs- (Maßstab, Toleranz- angaben,...)	gewünschte Produktan- forderun- gen	realisier- anforde- rungen	Technische Daten (Produkteinrichtung, -nummer, Material, Fertigungsform, Maßnahmen, Varianten, Lieferant, ...)	Metadaten Ersteller, Erstellungslam., verwendetes Softwaresystem, Zeichnungs- nummer Dateiname,....	Unterneh- men (Firmenname, Kunde)
Anderungsbilddatei			(x)				x	x
Anfertigungszeichnungen	x	x			(x)	x	x	x
Baugruppenelemente						x		
Bezugsdateien					x			x
Bewertungsergebnisse								
CAD-Modell 3D - JT	x	x				x	x	(x)
CAD-Modell - STEP	x	x	(x)			x	x	(x)
CAD-Modell 3D - VRML	x					(x)	(x)	(x)
Checklisten								
Einflussmatrix								
Fachliteratur / -zeitschriften								
Fertigungszeichnung / Technische Zeichnungen - DXF	x	x	x			x	x	x
Funktionskataloge	x				(x)	x	x	
Funktionsmodell	x					x	x	
Gesamtskizze		x	x				x	
Ideenformular	(x)							
Konstruktionskataloge					x			
Konzeptheit / -beschreibung	(x)	(x)			x	(x)	x	
Kostenkalkulation								
Kundenauftrag	x			x		(x)	x	x
Lastenheft				x	x			x
Lernungsprofil						x		
Lessons Learned und Best Practices				(x)	x	x		x
Lieferanten-Datenblätter							x	
Lösungsbeschreibungen / Prinzipielle Lösung				(x)	(x)		x	
Marktanalysedatensätze				x	(x)			(x)
Materialeigenschaften						x		x
Materialkostenkalkulation (Merkmalenmatrix)				x	x		x	
Modellstruktur						x		
Montage-/Fertigungsplan		x			x	x	x	
Morphologischer Kasten		x						
Nachweis Konstruktivfähigkeit	(x)				(x)	x	x	

Abbildung E.7. — Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informationsgehalts

Inhalt	Ergebnisse (Berechnungen, Simulationen, Bewertungen...)	(Ideen zu prinzipiellen Lösungsmög- lichkeiten)	Kosten Material-, Personal-, Fertigungs- kosten,	Projekt- ablauf	Termine	Aufgaben- träger	Mehrs- stene	Betriebs- sicherheit, Wartungs-, Montagean- weisungen	Erfah- rungen	Information über Design for X- Kriterien	System- verhalten	Umgebungs- bedingungen	Graphi- ken
Artefakte (Formate)													
Änderungsdokumente													
Arbeitsanweisungen								x					x
Baugruppeneinschreibungen	(x)		(x)					(x)		(x)	x	(x)	x
Berechnungsergebnisse	x									(x)	x		
Betriebsdaten													
Bewertungsergebnisse	x												
CAD-Modell 3D - JT						(x)							
CAD-Modell - STEP													
CAD-Modell 3D - VRML						(x)							
Checklisten													
Einflussmatrix													
Fachliteratur / Zeitschriften													
Fertigungszeichnung / Technische Zeichnungen - DXF					(x)	(x)							
FMEX-Analyse-Dokument	x									(x)			x
Funktionskataloge		x											(x)
Funktionsmodell													x
Gestaltstudie													
Identifikator													
Identifikationskataloge		x								(x)			x
Konzeptentwurf / -beschreibung		x											(x)
Kostenkalkulation			x										
Kundenanfrage			x		(x)					x	x	x	
Lastenheft					x					(x)	x	x	
Leistungsprofil													
Lessons Learned und Best Practices									x				(x)
Lieferanten-Datenblätter			(x)										x
Lösungsberechnungen / Prinzipielle Lösung	(x)	x									x		x
Machbarkeitsstudie	x	x							x				x
Marktanalyseergebnisse			(x)										
Materialliste													
Mehrschichtdokumentation (Mehrschichtmatrix)					x	x	x						
Messergebnisse	x											(x)	
Modulare Struktur		(x)											(x)
Montage- / Fertigungsplan													
Morphologischer Kasten						(x)							(x)
Nachweis Konzeptungstüchtigkeit	x	x											

Abbildung E.8. — Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informationsgehalts — Fortsetzung

Inhalt		Proble- funktion bzw. -ge- treuen funktionen & Lösungen	Geome- tredaten	Topologie / Produkt- struktur	Zeichnungs- daten (Maßstab, Toleranz- angaben, ...)	gewünschte Produktan- forderun- gen	realisier- bare Produkt- anforde- rungen	Technische Daten (Produktbezeichnung, Seriennummer, Material-, Fertigungsinfos, Maß- angaben, Varianten, Lieferant, ...)	Metadaten	
Artefakte (Formate)									Erschaffer, Erstellungsdatum, Änderungsdatum, verwendetes Software-System, Dateiname, ...	Zeichnungs- nummer
Packaging	Packaging	x	x	x						
Patente	Patente	x		x			(x)	x		(x)
Präsentationen	Präsentationen	x						(x)	x	x
Prinzipzeichnungen (Skizzen)	Prinzipzeichnungen (Skizzen)		x							
Produktdokumentation (Betriebs-, Wartungs-, Montageanleitung)	Produktdokumentation (Betriebs-, Wartungs-, Montageanleitung)							x	x	x
Produktlogik	Produktlogik	(x)		x				x		
Produktstrukturplan	Produktstrukturplan			x				x	x	x
Projektauftrag	Projektauftrag						x		x	x
Projektdokumentation	Projektdokumentation								x	x
Projektplan	Projektplan								x	
Prüfungsanweisungen	Prüfungsanweisungen									
Prüfplan	Prüfplan									
Richtlinien / Normen / Gesetze / Vorschriften (Umwelt-, Sicherheits-, Unternehmens- ...)	Richtlinien / Normen / Gesetze / Vorschriften (Umwelt-, Sicherheits-, Unternehmens- ...)								x	x
Schulpläne	Schulpläne	x						x	x	
Simulations- & Berechnungsergebnisse (Reports)	Simulations- & Berechnungsergebnisse (Reports)					x	x	x	x	
Spezifische Anforderungsliste	Spezifische Anforderungsliste	x		x			x	x		(x)
Stücklisten	Stücklisten						(x)	x		x
Teilkonstruktion	Teilkonstruktion									
Terminpläne	Terminpläne									
Verkehrsunterlagen	Verkehrsunterlagen	x		(x)			x	x		x
Recyclingplan	Recyclingplan								x	

Abbildung E.9. — Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informationsgehalts — Fortsetzung

Inhalt		Ergebnisse (Berechnungen, Simulationen, Bewertungen...)	(Ideen zu prinzipiellen Lösungsmög- lichkeiten)	Kosten Material-, Persönli- chkeits-, Fertigungs- kosten, ...	Projek- tablauf	Termine	Aufgaben- träger	Malen- stene	Betriebs- Sicherheits-, Wartungs-, Montagean- weisungen	Erfah- rungen	Information über Design for X- Kriterien	System- verhalten	Umgebungs- bedingungen	Graphi- ken
Aktivität (Vormate)														
Packaging														
Patente			x									(s)	(s)	x
Pflichtentwurf			(s)	x		x		x			x	x	x	
Präsentationen														
Prinzipzeichnungen (Skizzen)														
Produktdokumentation (Betriebs-, Wartungs-, Montageanleitung)			x						x					
Produktlogik														
Produktstrukturplan														
Projektauftrag						x	x	x	x					
Projektdokumentation				x		x	x	x	x					
Projektplan														
Profilanweisungen														
Prüfplan														
Rechnen / Normen / Gesetze / Vorschriften (Umwelt-, Sicherheits-, Unternehmens ...)										(s)	x			
Schulpläne												(s)		
Simulations- & Berechnungsergebnisse (Reports)		x											(s)	x
spezifische Anforderungsliste				(s)			x							
Stücklisten														
Teilekataloge									(s)					x
Terminpläne						x								
Vertriebsunterlagen									x			(s)	x	
Recyclingplan				(s)										x

Abbildung E.10. — Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informationsgehalts — Fortsetzung

F. IGES 5.3 – Spezifikationsauszug

Auszug aus der IGES 5.3-Spezifikation, die eine Auflistung der in der *Global Section* einer IGES-Datei übertragbaren Informationen enthält.

<u>Index</u>	<u>Type</u>	<u>Description</u>
1	String	Parameter delimiter character.
2	String	Record delimiter character.
3	String	Product identification from sending system
4	String	File name
5	String	Native System ID
6	String	Preprocessor version
7	Integer	Number of binary bits for integer representation
8	Integer	Maximum power often representable in a single-precision floating point number on the sending system
9	Integer	Number of significant digits in a single-precision floating point number on the sending system
10	Integer	Maximum power of ten representable in a double-precision floating point number on the sending system
11	Integer	Number of significant digits in a double-precision floating point number on the sending system
12	String	Product identification for the receiving system
13	Real	Model space scale
14	Integer	Units flag
15	String	Units Name
16	Integer	Maximum number of line weight gradations. Refer to the Directory Entry Parameter 12.
17	Real	Width of maximum line weight in units. Refer to the Directory Entry Parameter 12 (see Section 2.2.4.4.12) for use of this parameter.
18	String	Date and time of exchange file generation 15HYYMMDD.HHNNSS or 13HYYMMDD.HHNNSS where: YYYY or YY is 4 or 2 digit year HH is hour (00-23) MM is month (01-12) NN is minute (00-59) DD is day (01-31) SS is second (00-59)
19	Real	Minimum user-intended resolution or granularity of the model in units specified by Parameter 14.
20	Real	Approximate maximum coordinate value occurring in the model in units specified by Parameter 14.
21	String	Name of author
22	String	Author's organization
23	Integer	Flag value corresponding to the version of the Specification to which this file complies.
24	Integer	Flag value corresponding to the drafting standard to which this file complies, if any.
25	String	Date and time the model was created or last modified, in same format as field 18.
26	String	Descriptor indicating application protocol, application subset, Mil-specification, or user-defined protocol or subset, if any.

Abbildung F.1. — Auflistung der in der *Global Section* einer IGES-Datei übertragbaren Informationen (entnommen aus [US 96, S. 18])

G. Recall-Precision-Diagramme der FV-basierten Konzepte

Die in der Evaluation für jedes FV-basierte Konzept in Abhängigkeit von der verwendeten Distanzfunktion ermittelten Recall- und Precisionwerte werden in einem Recall-Precision-Diagramm dargestellt und können somit miteinander verglichen werden.

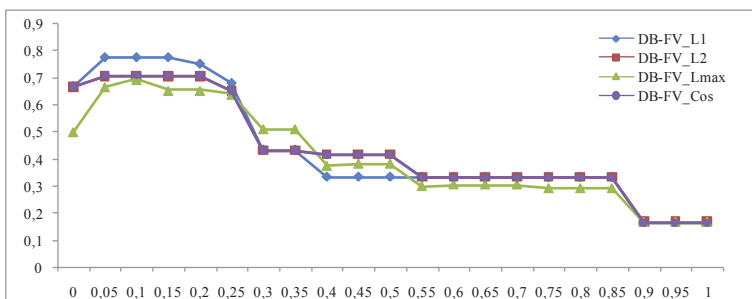


Abbildung G.1. — Recall-Precision-Diagramm für das FV-basierte Konzept DB-FV

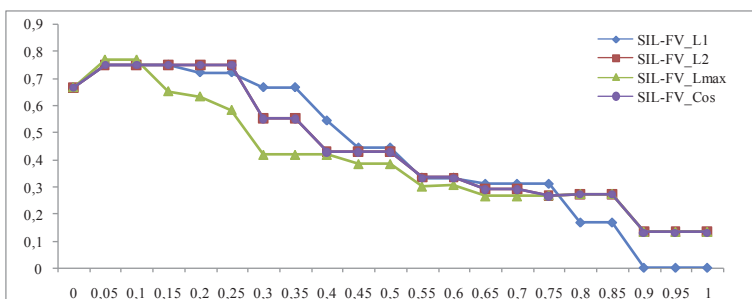


Abbildung G.2. — Recall-Precision-Diagramm für das FV-basierte Konzept SIL-FV

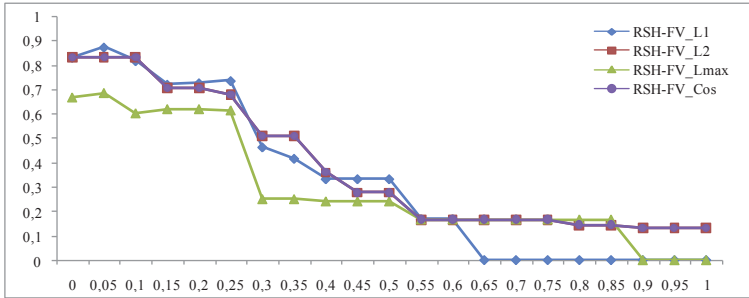


Abbildung G.3. — Recall-Precision-Diagramm für das FV-basierte Konzept RSH-FV

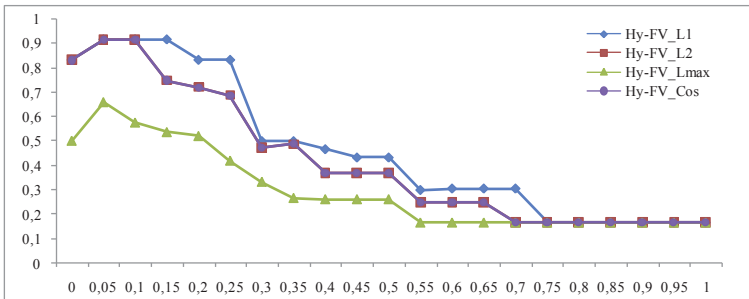


Abbildung G.4. — Recall-Precision-Diagramm für das FV-basierte Konzept Hy-FV

Abbildungsverzeichnis

1.1. Produzenten des <i>Digitalen Universums</i> gegliedert nach Industrien (nach [GCM ⁺ 08, S. 7])	2
1.2. Ursachen und Auswirkungen der Produktkomplexität (nach [Les01, S. 1])	5
1.3. Einordnung der Konstruktionsarten und Suchräume anhand der vier von Pahl et al. in [PBF07] definierten Prozessphasen	9
1.4. Beispiele für mögliche Fragestellungen bei der Entwicklung	14
1.5. Aufbau der Arbeit	24
2.1. Übersichtskatalog für Welle-Nabe-Verbindungen (entnommen aus [Rot01, S. 165])	27
2.2. Einordnung eines Fahrzeugs in eine a) Monohierarchie und eine b) Polyhierarchie	30
2.3. Screenshot der Kennwertsuche in der Werkstoffdatenbank M-Line 2 [SP07]	35
2.4. Screenshot des in DEPATISnet verfügbaren Suchmodus für Einsteiger	38
2.5. Beispiel einer in der Testversion von Geolus Search [Sie10a] durchgeführten Suche nach Schrauben, bei der die Suchanfrage mittels eines 3D-Modells formuliert ist und als Ergebnis ein Ranking ähnlicher Schrauben geliefert wird	43
2.6. Kommerzielle Anbieter von ES-Lösungen gemäß dem <i>Magic Quadrant for Information Access Technology</i> von Gartner, Inc. (nach [Whi09])	46
2.7. Bandbreite der von ES-Lösungen bereitgestellten Suchmöglichkeiten. Links: GUI der Google Search Appliance mit einer einfachen Schlagwortsuche (entnommen aus [Goo10a]); Rechts: GUI der Endeca Information Access Platform mit zahlreichen Möglichkeiten zur Anfrageverfeinerung (entnommen aus [End09])	47
3.1. a) Darstellung des 5D-Punktes $C = (c_1, c_2, c_3, c_4, c_5)$ in parallelen Koordinaten; b) Kartesische (oben) und parallele Koordinatendarstellung (unten) eines Quadrats (links), eines Würfels in 3D (mittig) und eines Würfels in 5D (rechts), wobei letzterer in kartesischen Koordinaten nicht mehr darstellbar ist (nach [Ins05, S. 153] und [Ins85, S. 90])	64

3.2. Konzept der multiplen Ebenen für die PE mit vier Artefakttypebenen und exemplarischen Artefaktbeziehungen sowohl innerhalb einer Ebene (durchgezogene Pfeile) als auch zwischen Ebenen (gestrichelte Pfeile)	66
3.3. Allgemeine Architektur des LFRP-Frameworks nach dem ADK-Strukturmodell	69
3.4. Graphische Benutzeroberfläche des LFRP-Frameworks	70
3.5. Relevanzklärung durch a) Gegenüberstellung oder b) Überlagerung von Sterndiagrammen, in welche zusätzlich c) die Gewichtung einzelner Facetten integriert werden kann	72
3.6. Erweiterung der Ergebnisliste mit Snippets (Ergebnis 1, 4 und 5) und Balkendiagrammen (Ergebnis 2 und 3) zur besseren Erklärung der Relevanz dieser Ergebnisse	74
3.7. Architektur der Suchkomponente des LFRP-Frameworks	75
3.8. Indexierungskomponente und Index des LFRP-Frameworks: Vereinfachte Darstellung des Indexierungsvorgangs	76
3.9. Artefaktbeschreibung als Polyrepräsentation	77
4.1. Für den Bereich der technischen PE zu differenzierende Artefakttypen, deren zugehörige Artefakte die drei Attribute <i>ArtefaktID</i> , <i>ArtefaktURI</i> und <i>Artefakttyp</i> (= Diskriminator) besitzen müssen	80
4.2. Exemplarische Artefakttyphierarchie für Produkte in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute und Fokus auf Spezialisierungsbeziehungen	82
4.3. Exemplarische Artefakttyphierarchie für Motoren als UML-Klassendiagramm mit relevanten Suchkriterien als Attribute und Fokus auf Spezialisierungs- und Aggregationsbeziehungen	86
4.4. Sammlung von Dokumenttypen, die in der PE von Bedeutung sind	88
4.5. Exemplarische Artefakttyphierarchie für Dokumente in Form eines UML-Klassendiagramms mit der <i>Dokumentkategorie</i> als Diskriminator und relevanten Suchkriterien als Attribute	90
4.6. Exemplarische Artefakttyphierarchie für Werkstoffe in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute	93
4.7. Überblick über wichtige Werkstoffeigenschaften, welche unterschiedliche, bei der Entwicklung zu berücksichtigende Aspekte adressieren (nach [MRB+08, S. 30])	95
4.8. Exemplarische Artefakttyphierarchie für Personen in Form eines UML-Klassendiagramms mit relevanten Suchkriterien als Attribute	96
4.9. Artefakttyp Projekt als UML-Klasse mit relevanten Suchkriterien als Attribute	98
4.10. Exemplarisches Beziehungsnetz bei Unterscheidung von fünf Artefakttypen (grün markiert): Produkt, Person, Dokument, Werkstoff und Projekt	99

4.11. Definition von Ausprägungen für die beiden Attribute <i>Area</i> (oben) und <i>Position</i> (unten) einer Kabeltülle (links) zur Verwendung in einem PDMS101	
4.12. Geforderte Struktur der Containerdatei für den Datenimport anhand eines Beispiels	104
5.1. Aufbau des Indexierers und vereinfachter Ablauf der darin stattfindenden Repräsentationserstellung unter Nutzung eines Facettenschemas .	110
5.2. Struktur der vom Interface IndexerFactory abgeleiteten Subklassen: Differenzierung nach dem Artefakttyp (erste Ebene) und Differenzierung nach dem Dokumenttyp (zweite Ebene)	120
5.3. Möglichkeiten zur Filterung anhand einer Ortsangabe: a) Anwendung einer einzigen hierarchischen Facette; b) Anwendung von drei eigenständigen Facetten	121
5.4. Hierarchie der zur Dokumentverarbeitung verfügbaren DocumentHandler	126
5.5. Vierstufiges Phasenmodell nach Pahl et al. [PBFG07] mit Beispielen für zugehörige Dokumenttypen	127
5.6. Auszüge aus unterschiedlichen Lastenheften eines Unternehmens . . .	131
5.7. Zwei Beispielseiten aus einem Pflichtenheft mit extrahierbaren dokument-spezifischen (grün umrandet), personenspezifischen (blau umrandet) und produktspezifischen (orange umrandet) Daten	135
5.8. Exemplarischer Auszug aus einem detaillierteren Lastenheft, welcher die gewünschten Produktfunktionen sowie die vom Auftraggeber geforderten Prüfungen beschreibt	138
5.9. Screenshot des GATE Developer	141
5.10. Elemente des IGES-Datenformates in der Version 5.1 (nach [And93, S. 185])	149
5.11. Approximation einer Zylindergeometrie durch Tessellierung (nach [AB ⁺ 07, S. Y 15])	152
5.12. Visualisierung eines Produktes im JT-Format mit PMI-Daten (links oben) und als <i>Properties</i> -Elemente gespeicherten Metadaten (links unten) sowie die zu einem PMI-Datum (rote Umrandung) extrahierbaren Informationen (rechts)	156
5.13. Beispiel einer STEP-Datei eines Deckels für einen Motorkanal, deren Metadaten nur folgende Standardinformationen enthalten: Pfad inkl. Dateiname (blau), Erstellungszeitpunkt (rot), Quell-CAD-System (grün), Verwendetes AP (orange), Produktname (lila), Produktgranularität (türkis)	161
5.14. Exemplarische Daten eines CAD-Modells zu dessen Produktstruktur (links) bzw. BRep-basierter Produktgestalt (rechts) im JT-Format . .	163
5.15. Vergleich von Produktstrukturen D_i mit der Struktur eines Anfrageproduktes Q_i (entnommen aus [TTP04, S. 1509])	164

5.16. Definition der triangulierten Geometriedaten im VRML-Format (links) und im JT-Format (rechts oben) sowie Darstellung der Zerlegung eines Polygons aus fünf Punkten in drei Dreiecke (rechts unten)	168
5.17. Mehrstufiger Prozess der Repräsentationserzeugung aus einem 3D-Modell (nach [BKS ⁺ 06, S. 40])	169
5.18. Gegenüberstellung der MAP@10- und MAP@250-Werte für die FV-basierten Konzepte: a) DB-FV, b) SIL-FV, c) RSH-FV und d) Hy-FV	175
5.19. Vergleich der MAP-Werte hinsichtlich der Parameterkombination bei Punktzahl $N = 50000$, Shellanzahl $s = 30$ und konstanter Ähnlichkeitsfunktion: QFD (oben) und L2 (unten)	177
5.20. Vergleich der MAP-Werte hinsichtlich der Ähnlichkeitsfunktion bei Punktzahl $N = 50000$, Shellanzahl $s = 30$ und konstanter Parameterkombination PDEm	178
5.21. Recall-Precision-Diagramm zum Vergleich der verschiedenen Parameter für das SD-Konzept, wobei für jede <i>shape function</i> das beste Evaluationsergebnis für den Vergleich verwendet wurde	179
5.22. Recall-Precision-Diagramm zum Vergleich der Konzepte mit den jeweils besten Parameterkombinationen	180
5.23. Beispiel eines automatisch generierten Berichts einer in Pro/E ausgeführten Modellanalyse, deren Ergebnis aus einer einfachen Auflistung von produktspezifischen Parameterwerten besteht (grün markiert)	183
5.24. Beispiel eines automatisch generierten Ergebnisberichts zu einer in Pro/MECHANICA durchgeführten Simulation bei einer Belastung durch Verschiebung	184
5.25. Beispiel eines im DOC/DOCX-Format gespeicherten Prüfberichts	186
5.26. Verschiedene Arten technischer Zeichnungen: a) Einzelteilzeichnung des Untergehäuses eines Lüfters; b) Gesamtzeichnung des Lüfters; c) Explosionszeichnung des Lüfters	189
5.27. Beispiel einer technischen Zeichnung aus der industriellen Praxis, deren Informationsgehalt in die drei Kategorien der organisatorischen, geometrischen und technischen Daten gruppiert wird	194
5.28. In einer technischen Zeichnung adressierte Artefakte mit einigen ihrer zugehörigen Daten	195
5.29. Beispiele für DXF-Objekte mit ihren jeweiligen Informationen, die aus der zu der in Abbildung 5.27 dargestellten technischen Zeichnung gehörenden DXF-Datei extrahiert werden können	196
5.30. Exemplarisches Schriftfeld einer technischen Zeichnung	197
5.31. Aufbau einer Extraktorkomponente für technische Zeichnungen im DXF-Format	199
5.32. a) Beschreibung eines komplexen Objektes anhand einer dominanten Geometrie (R1) und drei Hilfsgeometrien (R2, R3 und T1); b) Abbildung eines Objektes als Topologiegraph (nach [FFJ05, S. 9])	202

5.33. Beispiele für die zur Evaluation der Konzepte verwendeten einfachen technischen Zeichnungen	204
5.34. Recall-Precision-Diagramm für das topologiebasierte Konzept von Fonseca et al.	205
5.35. Recall-Precision-Diagramm zum Vergleich der Ähnlichkeitskonzepte für technische Zeichnungen unter Berücksichtigung der jeweils besten Einzelergebnisse	205
5.36. Verdeutlichung der drei grundsätzlichen Stücklistenformen anhand der Beschreibung des aus Einzelteilen (grau) und Baugruppen (grün) zusammengesetzten Produktes P (in Anlehnung an [PBF07, S. 554 - 564])	207
5.37. Beispiel einer Strukturstückliste aus der industriellen Praxis mit allgemeinen Daten und Bezugsdaten im Schriftfeld (oben) sowie Folgedaten im Stücklistenfeld (unten)	209
5.38. Überblick über die aus der in Abbildung 5.37 dargestellten Stückliste extrahierbaren Artefakte mit ihren Attribut- (nicht kursiv) und Beziehungsfacetten (kursiv): Dokument (1), Dokumentautor und Lieferanten als Personen (2), Hauptprodukt und Komponenten als Produkte (3) und Werkstoff (4)	211
6.1. Zuordnung der in einer Polyrepräsentation enthaltenen Repräsentationen auf mehrere Subindizes in Abhängigkeit von ihrer Facettenart . .	219
7.1. Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks .	227
7.2. Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks (Fortsetzung)	228
7.3. Leitfaden zur Realisierung des vorgestellten Indexierungsframeworks (Fortsetzung)	229
7.4. Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 1. bis 4. Suchschritt	231
7.5. Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 5. bis 8. Suchschritt	232
7.6. Beispielhaftes Suchvorgehen im LFRP-Framework für das Suchszenario D: 9. und 10. Suchschritt	233
8.1. Darstellung der aus drei Teilkomplexitäten bestehenden und durch das Indexierungs- und damit auch das LFRP-Framework zu beherrschenden Gesamtkomplexität	236
8.2. Kabelkanäle mit a) einem bestimmten Befestigungsclip (rot umrandet) und b) bzw. c) jeweils ohne diesen speziellen Befestigungsclip	238
E.1. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 1	269

E.2. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 1	270
E.3. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 2	271
E.4. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 2	272
E.5. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 1 bis 4 - Teil 3	273
E.6. Dokumenttypzuordnung zu den FORFLOW-Prozessschritten 5 und 6 - Teil 3	274
E.7. Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informa- tionsgehalts	275
E.8. Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informa- tionsgehalts — Fortsetzung	276
E.9. Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informa- tionsgehalts — Fortsetzung	277
E.10. Analyse der verschiedenen Dokumenttypen hinsichtlich ihres Informa- tionsgehalts — Fortsetzung	278
F.1. Auflistung der in der <i>Global Section</i> einer IGES-Datei übertragbaren Informationen (entnommen aus [US 96, S. 18])	279
G.1. Recall-Precision-Diagramm für das FV-basierte Konzept DB-FV . . .	281
G.2. Recall-Precision-Diagramm für das FV-basierte Konzept SIL-FV . . .	281
G.3. Recall-Precision-Diagramm für das FV-basierte Konzept RSH-FV . .	282
G.4. Recall-Precision-Diagramm für das FV-basierte Konzept Hy-FV . . .	282

Tabellenverzeichnis

1.1. Funktionsorientierte Einordnung von IT-Systemen der PE als erzeugende oder prozessbegleitende Systeme: + Hauptfunktionalität, o Teilfunktionalität und - keine Funktionalität im Aufgabenbereich der jeweiligen Systemgruppe	6
2.1. Übersicht und funktionelle Charakterisierung der kommerziell verfügbaren CAD-IR-Systeme	45
2.2. Übersicht der von den bereits existierenden Konzepten unterstützten Suchparadigmen mit Charakterisierung des Retrievalprozesses: S $\hat{=}$ Set Retrieval, R $\hat{=}$ Ranked Retrieval und SR $\hat{=}$ Set oder Ranked Retrieval (je nach Retrievalmodell)	55
3.1. Anforderungen an den Retrievalprozess hinsichtlich der zu unterstützenden Suchfunktionalitäten sowie der Ergebnispräsentation und die damit einhergehenden Anforderungen an die Indexierung	58
4.1. Zuordnung einer Auswahl von Dokumenttypen zu den drei Dokumentkategorien <i>Projektdokument</i> , <i>Produktmodell</i> und <i>Allgemeines Dokument</i>	92
5.1. Kategorisierung von Anforderungen nach Haberhauer [Hab08] (links) und nach Ullman [Ull03] (rechts) mit Beispielen	133
5.2. Einordnung der potentiell verfügbaren Metadaten in Kategorien	160
5.3. Übersicht über die prototypisch implementierten Konzepte	172
5.4. Übersicht über die prototypisch implementierten Konzepte — Fortsetzung	173
A.1. Vergleich von Vorgehensmodellen der technischen PE durch Gegenüberstellen der zugehörigen Prozessphasen	241
A.2. Vergleich von Vorgehensmodellen der technischen PE durch Gegenüberstellen der zugehörigen Prozessphasen - Fortsetzung	242
E.1. Bei der Zuordnung von Dokumenttypen zu Prozessphasen verwendete Kurzbezeichnungen und ihre Erklärung	268

Verzeichnis der Code-Listings

4.1. Exemplarische DTD der XML-Importdatei: artefaktimport.dtd	105
5.1. Inhalt eines <code><facet/></code> -Elementes gemäß facetSchema.xsd in Anhang C	111
5.2. Konkrete Definition von Facetten anhand von vier Beispielen	113
5.3. Inhalt eines <code><entity/></code> -Elementes gemäß facetSchema.xsd in Anhang C	115
5.4. Ausschnitt einer Artefakttyphierarchie mit exemplarischer Definition des Artefakttyps <i>Dokument</i> und seinen beiden Subklassen <i>Technische Zeichnung</i> und <i>CAD-Modell</i>	117
5.5. Exemplarische Definition einer Facette für Ortsangaben	122
5.6. Auszug aus einem Lastenheft für einen Kugelschreiber	130
5.7. Möglichkeit zur Speicherung von Metadaten in VRML-Dateien	154

Literaturverzeichnis

- [AB⁺07] ANDERL, R. ; BERGER, C. u. a. ; GROTE, K.-H. (Hrsg.) ; FELDHUSEN, J. (Hrsg.): *Dubbel - Taschenbuch für den Maschinenbau*. 22., neubearbeitete und erweiterte Auflage. Berlin, Heidelberg, New York : Springer-Verlag, 2007
- [ABB⁺04] ANDERMANN, H. ; BEKAVAC, B. ; BÖHLE, K. ; BOHNERT, R. u. a. ; KUHLEN, R. (Hrsg.) ; SEEGER, T. (Hrsg.) ; STRAUCH, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation. Band 2: Glossar*. 5., völlig neu gefasste Auflage. München : K. G. Saur Verlag GmbH, 2004
- [ABB⁺06] ALBER, B. ; BOCK, E. ; BÖHM, E. ; DENEFFLEH, R. ; DETERS, L. u. a. ; RIEG, F. (Hrsg.) ; KACZMAREK, M. (Hrsg.): *Taschenbuch der Maschinenelemente*. München, Wien : Carl Hanser Verlag, 2006
- [ABDP05] ANTINI, G. ; BERRETTI, S. ; DEL BIMBO, A. ; PALA, P.: Curvature Correlograms for Content Based Retrieval of 3D Objects. In: ROLI, F. (Hrsg.) ; VITULANO, S. (Hrsg.): *Image Analysis and Processing (ICIAP)* Bd. 3617. Berlin, Heidelberg : Springer-Verlag, 2005 (Lecture Notes in Computer Science), S. 859 – 866
- [Abl97] ABLAMEYKO, S. V.: *An Introduction to Interpretation of Graphic Images*. Bellingham, Washington : The Society of Photo-Optical Instrumentation Engineers (SPIE), 1997
- [ACI99] ACI - TECHNIK WUPPERTAL: *online-schr@uben.de*. Online-Shop. <http://www.online-schrauben.de>. Version: 1999. – letzter Zugriff: 08.06.2010
- [ADEK05] ARNOLD, V. ; DETTMERING, H. ; ENGEL, T. ; KARCHER, A.: *Product Lifecycle Management beherrschen: Ein Anwenderhandbuch für den Mittelstand*. Berlin, Heidelberg : Springer-Verlag, 2005
- [Ado08] ADOBE SYSTEMS INCORPORATED: *Document management - Portable document format - Part 1: PDF 1.7*. First Edition, Juli 2008. http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf. – letzter Zugriff: 30.07.2010
- [Ahm06a] AHMED, S.: An Approach to Assist Designers With Their Queries and Designs. In: *Proceedings of Design, Theory and Methodology Conference*

- (DTM). Pittsburgh, USA : American Society of Mechanical Engineers (ASME), September, 24 - 28 2006
- [Ahm06b] AHMED, S.: Prompting Designers to Design. In: *Proceedings of International Design Conference (DESIGN)*. Dubrovnik, Croatia, May, 15 - 18 2006, S. 1041 - 1048
- [AI99] APPELT, D. E. ; ISRAEL, D.: *Introduction to Information Extraction Technology*. <http://www.ai.sri.com/~appelt/ie-tutorial/>. Version: August 1999. - IJCAI-99 Tutorial, Stockholm, Schweden - letzter Zugriff: 20.02.2011
- [AKKP05] ASSFALG, J. ; KRIEGEL, H.-P. ; KRÖGER, P. ; PÖTKE, M.: Accurate and Efficient Similarity Search on 3D Objects Using Point Sampling, Redundancy, and Proportionality. In: BAUZER MEDEIROS, C. (Hrsg.) ; EGENHOFER, M. (Hrsg.) ; BERTINO, E. (Hrsg.): *Advances in Spatial and Temporal Databases* Bd. 3633. Berlin, Heidelberg : Springer-Verlag, 2005 (Lecture Notes in Computer Science), S. 200 - 217
- [AKKS99] ANKERST, M. ; KASTENMÜLLER, G. ; KRIEGEL, H.-P. ; SEIDL, T.: 3D Shape Histograms for Similarity Search and Classification in Spatial Databases. In: *Proceedings of the 6th International Symposium on Spatial Databases (SSD)* Bd. 1651. Hong Kong, China : Springer-Verlag, 1999 (Lecture Notes in Computer Science), 207 - 226
- [AKWK06] AKBAR, S. ; KÜNG, J. ; WAGNER, R. ; KEPLER, J.: Multi-Feature Integration on 3D Model Similarity Retrieval. In: *Proceedings of 1st International Conference on Digital Information Management*. Bangalore, 2006, S. 151 - 156
- [All83] ALLEN, J. F.: Maintaining Knowledge about Temporal Intervals. In: *Communications of the ACM* 26 (1983), November, Nr. 11, S. 832 - 843
- [AM09] ABRAMOVICI, M. ; MEIMANN, V.: Holistic Intellectual Property Protection of Virtual Product Models in Product Development Networks. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [And93] ANDERL, R. ; ABELN, O. (Hrsg.): *CAD-Schnittstellen: Methoden und Werkzeuge zur CA-Integration*. München, Wien : Carl Hanser Verlag, 1993
- [And05a] ANDREASEN, M. M.: Concurrent Engineering - effiziente Integration der Aufgaben im Entwicklungsprozess. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel IV 3, S. 293 - 315

- [And05b] ANDREASEN, M. M.: Vorgehensmodelle und Prozesse für die Entwicklung von Produkten und Dienstleistungen. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel IV 1, S. 247 – 263
- [AS94] AHLBERG, C. ; SHNEIDERMAN, B.: Visual Information Seeking Using the FilmFinder. In: *Conference Companion on Human Factors in Computing Systems (CHI)*. Boston, Massachusetts, USA : ACM Press, 1994, S. 433 – 434
- [Ash57] ASHBY, W. R.: *An Introduction To Cybernetics*. Second Impression. London : Chapman & Hall LTD, 1957
- [Ash07] ASHBY, M. F. ; WANNER, A. (Hrsg.) ; FLECK, C. (Hrsg.): *Materials Selection in Mechanical Design (Third Edition) - Das Original mit Übersetzungshilfen*. 1. Auflage. München : Elsevier GmbH, 2007
- [AT00] ANDERL, R. ; TRIPPNER, D.: *STEP Standard for the Exchange of Product Model Data: Eine Einführung in die Entwicklung, Implementierung und industrielle Nutzung der Normenreihe ISO 10303 (STEP)*. Stuttgart, Leipzig : B. G. Teubner, 2000
- [ATC+08] AU, O. K.-C. ; TAI, C.-L. ; CHU, H.-K. ; COHEN-OR, D. ; LEE, T.-Y.: Skeleton Extraction by Mesh Contraction. In: *ACM Transactions on Graphics* 27 (2008), Nr. 3, Article 44
- [ATO05] ABUSALAH, M. ; TAIT, J. ; OAKES, M.: Literature Review of Cross Language Information Retrieval. In: *World Academy of Science, Engineering and Technologie* 4 (2005), S. 175 – 177
- [Aut08] AUTODESK, INC.: *Autodesk MAGAZIN*. Website. <http://www.autodeskmagazin.de>. Version: 2008. – letzter Zugriff: 30.07.2010
- [Aut09] AUTOCAD MAGAZIN: *Planen und Bauen in 5 Dimensionen*. Online-Artikel vom 05.10.2009. http://www.autocad-magazin.de/index.php?level=1&CatID=19&inhalt_id=19&presse=6843&do=showDetail. Version: Oktober 2009. – letzter Zugriff: 30.07.2010
- [Aut10a] AUTODESK, INC.: *AutoCAD 2011 - DXF Reference*. v.u.25.1.01, February 2010. <http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454>. – letzter Zugriff: 15.08.2010
- [Aut10b] AUTODESK, INC.: *Autodesk*. Website. <http://www.autodesk.de>. Version: 2010. – letzter Zugriff: 30.07.2010

- [Aut10c] AUTONOMY CORP.: *Autonomy KeyView IDOL*. Produktbeschreibung. <http://www.autonomy.com/content/Products/idol-modules-connectors/index.en.html>. Version: 2010. – letzter Zugriff: 14.04.2010
- [Aut10d] AUTONOMY CORP.: *Enterprise Search*. Website. <http://www.autonomy.com/content/Products/enterprise-search/index.en.html>. Version: 2010. – letzter Zugriff: 14.04.2010
- [AW03] AHMED, S. ; WALLACE, K.: Indexing Design Knowledge Based upon Descriptions of Design Processes. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 14th International Conference on Engineering Design (ICED)*. Stockholm, Schweden, August, 19 - 21 2003
- [AWF85] AWF (Hrsg.): *Integrierter EDV-Einsatz in der Produktion*. Eschborn : Ausschuss für wirtschaftliche Fertigung, 1985
- [Bar09] BARTOLINI, I.: A Multi-faceted Browsing Interface for Digital Photo Collections. In: *Proceedings of the Seventh International Workshop on Content-Based Multimedia Indexing (CBMI)*, IEEE Computer Society, 2009, S. 237 – 242
- [Bat89] BATES, M. J.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. In: *Online Review* 13 (1989), Nr. 5, S. 407 – 424
- [Bat90] BATES, M. J.: Where Should the Person Stop and the Information Search Interface Start? In: *Information Processing and Management* 26 (1990), Nr. 5, S. 575 – 591
- [BB98] BURGE, J. E. ; BROWN, D. C.: Design Rationale Types and Tools / Computer Science Department of Worcester Polytechnic Institute, Artificial Intelligence in Design Group. Version: 1998. <http://web.cs.wpi.edu/Research/aidg/DR-Rpt98.html>. 1998. – Forschungsbericht. – letzter Zugriff: 23.04.2010
- [BBH⁺02] BECKER, M. ; BÜCHEL, A. ; HABERFELLNER, R. ; MASSOW, H. von ; NAGEL, P. ; DAENZER, W. F. (Hrsg.) ; HUBER, F. (Hrsg.): *Systems Engineering: Methodik und Praxis*. 11., durchgesehene Auflage. Zürich : Verlag Industrielle Organisation, 2002
- [BEE⁺97] BEITZ, W. ; EHRENSPIEL, K. ; ERB, J. ; GAUSEMEIER, J. u. a. ; GRABOWSKI, H. (Hrsg.) ; GEIGER, K. (Hrsg.): *Neue Wege zur Produktentwicklung*. Stuttgart : Dr. Josef Raabe Verlags-GmbH, 1997
- [Ber94] BERNDT, O.: *Dokumenten-Management-Systeme*. Neuwied : Hermann Luchterhand Verlag GmbH, 1994

- [Ber97] BERCHTOLD, S.: *Geometrie-basierte Suche ähnlicher Bauteile*, Fakultät für Mathematik der Ludwig-Maximilians-Universität München, Diss., Juni 1997
- [BGB⁺92] BAYA, V. ; GEVINS, J. ; BAUDIN, C. ; MABOGUNJE, A. ; TOYE, G. ; LEIFER, L.: An Experimental Study of Design Information Reuse. In: *Proceedings of the 4th International Conference on Design Theory and Methodology*, ASME, 1992, S. 141 – 147
- [BJ95] BYSTRÖM, K. ; JÄRVELIN, K.: Task Complexity Affects Information Seeking and Use. In: *Information Processing and Management* 31 (1995), Nr. 2, S. 191 – 213
- [BJ02] BANACH, D. T. ; JONES, T.: *Autodesk InventorTM: From the Top*. Canada : Autodesk Press and Thomson Learning, 2002
- [BK09] BODENDORF, F. ; KAISER, C.: *Case Based Reasoning*. Enzyklopädie der Wirtschaftsinformatik - Online Lexikon. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/technologien-methoden/KI-und-Softcomputing/Case-Based-Reasoning>. Version: März 2009. – letzter Zugriff: 26.04.2010
- [BKK⁺04] BRECHEISEN, S ; KRIEGEL, H.-P. ; KUNATH, P. ; PFEIFLE, M. ; RENZ, M.: Der virtuelle Prototyp: Datenbankunterstützung für CAD-Anwendungen. In: *Datenbank-Spektrum* 10 (2004), S. 22 – 29
- [BKS⁺04] BUSTOS, B. ; KEIM, K. ; SAUPE, D. ; SCHRECK, T. ; VRANIC, D. V.: Automatic Selection and Combination of Descriptors for Effective 3D Similarity Search. In: *Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE)*. Washington, DC, USA : IEEE Computer Society, 2004, S. 514 – 521
- [BKS⁺05] BUSTOS, B. ; KEIM, D. A. ; SAUPE, D. ; SCHRECK, T. ; VRANIC, D. V.: Feature-Based Similarity Search in 3D Object Databases. In: *ACM Computer Survey* 37 (2005), December, Nr. 4, S. 345 – 387
- [BKS⁺06] BUSTOS, B. ; KEIM, D. A. ; SAUPE, D. ; SCHRECK, T. ; VRANIC, D.V.: An Experimental Effectiveness Comparison of Methods for 3D Similarity Search. In: *International Journal on Digital Libraries* 6 (2006), Nr. 1, S. 39 – 54
- [BMM⁺06] BRANDT, S. C. ; MORBACH, J. ; MIATIDIS, M. ; THEISSEN, M. ; JARKE, M. ; MARQUARDT, W.: Ontology-Based Information Management in Design Processes. In: *Computer Aided Chemical Engineering* 21 (2006), S. 2021 – 2026

- [BR⁺99] BAEZA-YATES, R. ; RIBEIRO-NETO, B. u. a. ; BAEZA-YATES, R. (Hrsg.) ; RIBEIRO-NETO, B. (Hrsg.): *Modern Information Retrieval*. Harlow, England : ACM press, Addison Wesley Longman Limited, 1999
- [Bro80] BROOKES, B. C.: The foundations of information science: Part I. Philosophical aspects. In: *Journal of Information Science* 2 (1980), S. 125 – 133
- [BRS03] BESPALOV, D. ; REGLI, W. C. ; SHOKOUFANDEH, A.: Reeb Graph Based Shape Retrieval for CAD. In: *Proceedings of DETC: ASME Design Engineering Technical Conferences*. Chicago, Illinois USA, September 2003
- [BS07] BASSHUYSSEN, R. van (Hrsg.) ; SCHÄFER, F. (Hrsg.): *Handbuch Verbrennungsmotoren - Grundlagen, Komponenten, Systeme, Perspektiven*. 4., aktualisierte und erweiterte Auflage. Wiesbaden : Friedr. Vieweg & Sohn Verlag / GWV Fachverlage GmbH, 2007
- [BS10] BASSHUYSSEN, R. van ; SCHÄFER, F.: *Motorlexikon.de*. Website. <http://www.motorlexikon.de/>. Version: 2010. – letzter Zugriff: 09.06.2010
- [BW97] BLOTT, S. ; WEBER, R.: A Simple Vector-Approximation File for Similarity Search in High-Dimensional Vector Spaces / Bundesamt für Bildung und Wissenschaft (BBW) Schweiz: European ESPRIT project HERMES (project no. 9141). 1997 (19). – Technical Report
- [Car97] CARSTENSEN, P. H.: Towards Information Exploration Support for Engineering Designers. In: *Advances in Concurrent Engineering* (1997), S. 26 – 33
- [CDK⁺07] CONRAD, J. ; DEUBEL, T. ; KÖHLER, C. ; WANKE, S. ; WEBER, C.: Comparison of Knowledge Representation in PDM and by Semantic Networks. In: *Proceedings of the 16th International Conference on Engineering Design (ICED)*. Paris, France, August, 28 - 31 2007
- [CGDK06] CARDONE, A. ; GUPTA, S. K. ; DESHMUKH, A. ; KARNIK, M.: Machining Feature-Based Similarity Assessment Algorithms for Prismatic Machined Parts. In: *Computer-Aided Design* 38 (2006), S. 954 – 972
- [CGK03] CARDONE, A. ; GUPTA, S. K. ; KARNIK, M.: A Survey of Shape Similarity Assessment Algorithms for Product Design and Manufacturing Applications. In: *Journal of Computing and Information Science in Engineering* 3 (2003), June, S. 109 – 118
- [CH06] CHU, C. ; HSU, Y.: Similarity Assessment of 3D Mechanical Components for Design Reuse. In: *Robotics and Computer-Integrated Manufacturing* 22 (2006), Nr. 4, S. 332 – 341

- [Cha05] CHAKRABORTY, T.: Shape-Based Clustering Of Enterprise CAD Databases. In: *Computer-Aided Design & Applications* 2 (2005), Nr. 1 - 4, S. 145 – 154
- [CKS08] CHA, H.-S. ; KIM, K. ; SHIN, D.-W.: Development of 5D-CAD based Work Process Model for Aged-Housing Revamp Projects. In: *International Conference on Information Technology in Construction*. Santiago, Chile, 2008
- [CLT06] CAO, L. ; LIU, J. ; TANG, X.: 3D Object Retrieval Using 2D Line Drawing and Graph Based Relevance Feedback. In: *MULTIMEDIA '06: Proceedings of the 14th Annual ACM International Conference on Multimedia*. New York, NY, USA : ACM Press, 2006, S. 105 – 108
- [CM08] CAPRA, R. G. ; MARCHIONINI, G.: The Relation Browser Tool for Faced Exploratory Search. In: *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*. Pittsburgh, Pennsylvania, USA : ACM Press, June, 16 - 20 2008, S. 420 – 420
- [CMBT02] CUNNINGHAM, H. ; MAYNARD, D. ; BONTCHEVA, K. ; TABLAN, V.: GATE: An Architecture for Development of Robust HLT Applications. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*. Morristown, NJ, USA : Association for Computational Linguistics, 2002, S. 168 – 175
- [CMO+07] CAPRA, R. ; MARCHIONINI, G. ; OH, J. S. ; STUTZMAN, F. ; ZHANG, Y.: Effects of Structure and Interaction Style on Distinct Search Tasks. In: *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*. Vancouver, British Columbia, Canada : ACM Press, June, 18 - 23 2007, S. 442 – 451
- [Con05] CONRAD, K.-J.: *Grundlagen der Konstruktionslehre: Methoden und Beispiele für den Maschinenbau*. 3. aktualisierte und erweiterte Auflage. München, Wien : Carl Hanser Verlag, 2005
- [Con08a] CONRAD, K.-J.: Gestaltungsrichtlinien. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 23, S. 416 – 465
- [Con08b] CONRAD, K.-J.: Konstruktionsablauf. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 15, S. 249 – 305
- [Con08c] CONRAD, K.-J.: Konstruktionsberechnung. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München

- : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 8, S. 154 – 180
- [Con08d] CONRAD, K.-J.: Rechnerunterstützung der Konstruktion. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 26, S. 560 – 579
- [Con08e] CONRAD, K.-J.: Technische Zeichnungen. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 2, S. 28 – 68
- [Cou97] COURT, A. W.: The Relationship Between Information and Personal Knowledge in New Product Development. In: *International Journal of Information Management* 17 (1997), Nr. 2, S. 123 – 138
- [CPZ97] CIACCIA, P. ; PATELLA, M. ; ZEZULA, P.: M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In: JARKE, M. (Hrsg.) ; CAREY, M. J. (Hrsg.) u. a.: *Proceedings of the 23rd International Conference on Very Large Data Bases (VLDB)*. San Francisco : Morgan Kaufmann, 1997, S. 426 – 435
- [CR01] CICIRELLO, V. ; REGLI, W. C.: Machining Feature-Based Comparisons of Mechanical Parts. In: *Proceedings of the International Conference on Shape Modeling and Applications (SMA)*. Washington DC, USA : IEEE Computer Society, May 2001, S. 176 – 187
- [CRC+02] CORNEY, J. ; REA, H. ; CLARK, D. ; PRITCHARD, J. ; BREAKS, M. ; MACLEOD, R.: Coarse Filters for Shape Matching. In: *IEEE Computer Graphics and Applications: Graphics in Advanced Computer-Aided Design* 22 (2002), Nr. 3, S. 65 – 74
- [CSY87] CHANG, S. K. ; SHI, Q. Y. ; YAM, C. W.: Iconic Indexing of 2D Strings. In: *IEEE Transactions of Pattern Analysis and Machinable Intelligence* 9 (1987), Nr. 3, S. 413 – 428
- [CTSO03] CHEN, D.-Y. ; TIAN, X.-P. ; SHEN, Y.-T. ; OUHYOUNG, M.: On Visual Similarity Based 3D Model Retrieval. In: *Eurographics*. Granada, Spain, 2003, S. 223 – 232
- [Dan99] DANDL, J.: Dokumenten-Management-Systeme - Eine Einführung / Johannes-Gutenberg-Universität Mainz; Lehrstuhl für Allgemeine BWL und Wirtschaftsinformatik. 1999 (9). – Arbeitspapiere WI
- [Das10a] DASSAULT SYSTÈMES: *Dassault Systèmes*. Website. <http://www.3ds.com/de>. Version: 2010. – letzter Zugriff: 30.07.2010

- [Das10b] DASSAULT SYSTÈMES SOLIDWORKS CORP.: *SolidWorks*. Website. <http://www.solidworks.com>. Version: 2010. – letzter Zugriff: 30.07.2010
- [Das10c] DASSAULT SYSTÈMES SOLIDWORKS CORP.: *Unternehmensgeschichte*. Website. http://www.solidworks.de/sw/656_DEU_HTML.htm. Version: 2010. – letzter Zugriff: 30.07.2010
- [Den03] DENTON, W.: *How to Make a Faceted Classification and Put it on the Web*. Website. <http://www.miskatonic.org/library/facet-web-howto.html>. Version: November 2003. – letzter Zugriff: 09.05.2010
- [Deu62] DEUTSCHES INSTITUT FÜR NORMUNG: *DIN 199: Technische Zeichnungen: Benennungen*. September 1962
- [Deu77] DEUTSCHES INSTITUT FÜR NORMUNG: *DIN 199 Teil 2: Begriffe im Zeichnungs- und Stücklistenwesen: Stücklisten*. Dezember 1977
- [Deu09] DEUTSCHES INSTITUT FÜR NORMUNG: *DIN 69901-5: Projektmanagement - Projektmanagementsysteme - Teil 5: Begriffe*. 2009
- [Deu10] DEUTSCHES PATENT- UND MARKENAMT: *Die internationale Patentklassifikation*. Homepage des DPMA. <http://www.dpma.de/service/klassifikationen/ipc/index.html>. Version: Februar 2010. – letzter Zugriff: 26.02.2010
- [DK07] DECKER, K.-H. ; KABUS, K.: *Maschinenelemente - Funktion, Gestaltung und Berechnung*. 16., neu bearbeitete Auflage. München : Carl Hanser Verlag, 2007
- [DLW05] DATTA, R. ; LI, J. ; WANG, J. Z.: Content-based Image Retrieval - Approaches and Trends of the New Age. In: *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*. New York, NY, USA : ACM Press, 2005, S. 253 – 262
- [Dör87] DÖRNER, D. ; HERRMANN, T. W. (Hrsg.) ; TACK, W. H. (Hrsg.) ; WEINERT, F. E. (Hrsg.): *Problemlösen als Informationsverarbeitung*. 3. Auflage. Stuttgart, Berlin : W. Kohlhammer GmbH, 1987
- [Dru93] DRUCKER, P. F.: *Post-Capitalist Society*. New York : HaperCollins Publishers, Inc. (HaperBusiness), 1993
- [eCl07] eCL@SS E. V.: *eCl@ss-Leitfaden für kleine und mittlere Unternehmen*, 2007. http://www.eclass.de/user/documents/broschuere_eiclass_leitfaden.pdf. – letzter Zugriff: 01.03.2010
- [eCl09] eCL@SS E. V.: *eCl@ss 7.0 - Produktdatenmanagement vom Hersteller bis zum CAD: Bericht der eCl@ss-Arbeitsgruppe CAX-Anwendungen*, 2009.

- http://www.eclass.de/user/documents/broschuere_cax.pdf. – letzter Zugriff: 02.03.2010
- [eCl10] eCl@ss E. V.: *eCl@ss - Internationaler Standard zur Klassifizierung und Beschreibung von Produkten und Dienstleistungen*. eCl@ss-Service Portal. <http://www.eclass.de>. Version: 2010. – letzter Zugriff: 01.03.2010
- [EFH⁺09] ECKSTEIN, R. ; FAERBER, M. ; HENRICH, A. ; JABLONSKI, S. ; JOCHAUD, F. u. a. ; HENRICH, A. (Hrsg.) ; JABLONSKI, S. (Hrsg.) ; KRCMAR, H. (Hrsg.) ; LINDEMANN, U. (Hrsg.) ; MEERKAMM, H. (Hrsg.) ; RIEG, F. (Hrsg.): *Flexible Prozessunterstützung in der Produktentwicklung: Prozesse - Daten - Navigation*. Aachen : Shaker Verlag GmbH, 2009
- [Ehr07] EHRENSPIEL, K.: *Integrierte Produktentwicklung - Denkabläufe, Methodeneinsatz, Zusammenarbeit*. München, Wien : Carl Hanser Verlag, 2007 (3., aktualisierte Auflage)
- [EHW09] ECKSTEIN, R. ; HENRICH, A. ; WEBER, N.: Das LFRP-Framework - ein Ansatz zur interaktiven Suche für unternehmensspezifische Informationen. In: *Datenbank-Spektrum* 9 (2009), Oktober, Nr. 30, S. 20 – 29
- [EHW10] ECKSTEIN, R. ; HENRICH, A. ; WEBER, N.: LFRP-Search: Multi-Layer Ranked Visual Faceted Search - An Approach to Cope with Complex Search Situations. In: *Proceedings of the 25th ACM Symposium on Applied Computing (SAC)*. Sierre, Schweiz : ACM, March, 22 - 26 2010, S. 1713 – 1717
- [EM91] EIGNER, M. ; MAIER, H.: *Einstieg in CAD: Lehrbuch für CAD-Anwender*. 8. Nachdruck. München, Wien : Carl Hanser Verlag, 1991
- [EN09] ELMASRI, R. A. ; NAVATHE, S. B.: *Grundlagen von Datenbanksystemen*. 3., aktualisierte Auflage. München : Pearson Studium (Pearson Education Deutschland GmbH), 2009. – Bachelorausgabe
- [End09] ENDECA TECHNOLOGIES, INC.: *Search Applications on the Endeca Information Access Platform*. Technical White Paper. <http://www.endeca.com/resource-center-white-papers.htm>. Version: 2009. – letzter Zugriff: 14.04.2010
- [End10] ENDECA TECHNOLOGIES, INC.: *Enterprise Search Tools & Solutions*. Website. <http://www.endeca.com/solutions-enterprise-search.htm>. Version: 2010. – letzter Zugriff: 14.04.2010
- [Ern08] ERNST, H.: *Grundkurs Informatik: Grundlagen und Konzepte für die erfolgreiche IT-Praxis - Eine umfassende, praxisorientierte Einführung*. 4., vollständig überarbeitete Auflage. Wiesbaden : Vieweg + Teubner Verlag, 2008

- [ES01] EIGNER, M. ; STELZER, R.: *Produktdatenmanagement-Systeme: Ein Leit-faden für Product Development und Life Cycle Management*. Berlin, Hei-delberg : Springer-Verlag, 2001
- [Exa08] EXALEAD S. A.: *Enterprise Search*. Website. <http://www.exalead.com/software/solutions/enterprise-search/>. Version: 2008. – letzter Zu-griff: 13.04.2010
- [Fer03] FERBER, Reginald: *Information Retrieval - Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. Heidelberg : dpunkt.verlag, 2003
- [Fer08] FERDINAND GROSS GMBH & CO. KG: *Fachwörterverzeichnis*. Unterneh-menshomepage. <http://www.schrauben-gross.de/index.php?id=104>. Version: 2008. – letzter Zugriff: 08.06.2010
- [FFJ05] FONSECA, M. J. ; FERREIRA, A. ; JORGE, J. A.: Content-Based Retrieval of Technical Drawings. In: *International Journal of Computer Applications in Technology* 23 (2005), March, Nr. 2-3, S. 86 – 100
- [FGS08] FRICK, D. ; GADATSCH, A. ; SCHÄFFER-KÜLZ, U. G.: *Grundkurs SAP ERP: Geschäftsprozess-orientierte Einführung mit durchgehendem Fallbei-spiel*. 1. Auflage. Wiesbaden : Friedr. Vieweg & Sohn Verlag / GWV Fachverlage GmbH, 2008
- [Fin03] FINK, G. A.: *Mustererkennung mit Markov-Modellen: Theorie - Praxis - Anwendungsgebiete*. 1. Auflage. Wiesbaden : B. G. Teubner Verlag / GWV Fachverlage GmbH, 2003
- [Fin05] FINGER, D.: *Handlungsempfehlungen für die Implementierung eines ERP-System-basierten Wissensmanagements*. Lohmar, Köln : JOSEF EUL Ver-lag GmbH, 2005. – Dissertation, Rechts- und Wirtschaftswissenschaftliche Fakultät der Universität des Saarlandes
- [FL97] FRICKE, G. ; LOHSE, G.: *Entwicklungsmanagement: Mit methodischer Pro-duktentwicklung zum Unternehmenserfolg*. Berlin, Heidelberg : Springer-Verlag, 1997
- [FMK⁺03] FUNKHOUSER, T. ; MIN, P. ; KAZHDAN, M. ; CHEN, J. ; HALDERMAN, A. ; DOBKIN, D.: A Search Engine for 3D Models. In: *ACM Transactions on Graphics* (2003), January, S. 83 – 105
- [FMKK00] FRÄNTI, P. ; MEDNONOGOV, A. ; KYRKI, V. ; KÄLVÄINEN, H.: Content-based Matching of Line-Drawing Images Using the Hough Transform. In: *International Journal on Document Analysis and Recognition* 3 (2000), Nr. 3, S. 117 – 124

- [Foc09] FOCUS ONLINE: *Vor 20 Jahren: Die Geburtsstunde des World Wide Web*. Online-Artikel (Quelle: claf/dpa). http://www.focus.de/digital/internet/tid-13637/vor-20-jahren-die-geburtsstunde-des-world-wide-web_aid-379711.html. Version: 13. März 2009. – letzter Zugriff: 02.10.2009
- [Fon04] FONSECA, M. J.: *Sketch-Based Retrieval in Large Sets of Drawings*. Lisbon, Portugal, Instituto Superior Técnico, Diss., Juli 2004
- [FS08] FERSTL, O. K. ; SINZ, E. J.: *Grundlagen der Wirtschaftsinformatik*. 6., überarb. und erw. Auflage. München : Oldenbourg Wissenschaftsverlag GmbH, 2008
- [FSV01] FINK, A. ; SCHNEIDERREIT, G. ; VOSS, S.: *Grundlagen der Wirtschaftsinformatik*. Heidelberg : Physica-Verlag, 2001
- [Fuh96] FUHR, N.: *Ziele und Aufgaben der Fachgruppe Information Retrieval*. Homepage der Fachgruppe Information Retrieval, Gesellschaft für Informatik. http://www.uni-hildesheim.de/fgir/index.php?option=com_content&task=view&id=14&Itemid=41. Version: Januar 1996. – letzter Zugriff: 26.01.2010
- [Gab10] GABLER VERLAG (HRSG.): *Gabler Wirtschaftslexikon; Stichwort: Konsistenzpostulat*. Online-Lexikon. <http://wirtschaftslexikon.gabler.de/Archiv/1792/konsistenzpostulat-v6.html>. Version: 2010. – letzter Zugriff: 12.09.2010
- [Gab11a] GABLER VERLAG (HRSG.): *Gabler Wirtschaftslexikon; Stichwort: kritische Erfolgsfaktoren*. Online-Lexikon. <http://wirtschaftslexikon.gabler.de/Definition/kritische-erfolgsfaktoren.html>. Version: 2011. – letzter Zugriff: 30.01.2011
- [Gab11b] GABLER VERLAG (HRSG.): *Gabler Wirtschaftslexikon; Stichwort: Nutzen*. Online-Lexikon. <http://wirtschaftslexikon.gabler.de/Definition/nutzen.html>. Version: 2011. – letzter Zugriff: 21.01.2011
- [Gar97] GARFIELD, E.: A Tribute To Calvin N. Mooers, A Pioneer Of Information Retrieval. In: *The Scientist* 11 (1997), Nr. 6, S. 9
- [Gau05] GAUS, W.: *Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval*. Fünfte, überarbeitete Auflage. Berlin, Heidelberg : Springer-Verlag, 2005
- [GC06] GAL, R. ; COHEN-OR, D.: Salient Geometric Features for Partial Shape Matching and Similarity. In: *ACM Transactions on Graphics* 25 (2006), January, Nr. 1, S. 130 – 150

- [GCM⁺08] GANTZ, J. F. ; CHUTE, C. ; MANFREDIZ, A. ; MINTON, S. ; REINSEL, D. ; SCHLICHTING, W. ; TONCHEVA, A.: *The Diverse and Exploding Digital Universe - An Updated Forecast of Worldwide Information Growth Through 2011*. IDC White Paper - sponsored by EMC Corporation. <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>. Version: März 2008. – letzter Zugriff: 02.10.2009
- [GEG05] GAUSEMEIER, J. ; EBBESMEYER, P. ; GRAFE, M.: Prototyping und Produktmodellierung. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel VI 6, S. 603 – 623
- [GMPS00] GREENE, S. ; MARCHIONINI, G. ; PLAISANT, C. ; SHNEIDERMAN, B.: Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information Seeking. In: *Journal of the American Society for Information Science* 51 (2000), Nr. 4, S. 380 – 393
- [Goo09] GOOGLE INC.: *Google Search Appliance 6.0: Universelle Suche für Unternehmen*. Produktdatenblatt. http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.de/de/de/enterprise/pdf/gsa6_ds_DE.pdf. Version: 2009. – letzter Zugriff: 14.04.2010
- [Goo10a] GOOGLE INC.: *Funktionen für Endnutzer: Beispiel-Suchanfrage*. On-Demand-Produkttour: Video. http://www.google.de/enterprise/gsa/555_google_user_interface.html. Version: 2010. – letzter Zugriff: 14.04.2010
- [Goo10b] GOOGLE INC.: *Google Search Appliance*. Website. http://www.google.de/enterprise/gsa/index.html#utm_campaign=de&utm_source=de-ha-emea-de-bk&utm_medium=ha&utm_term=google%20search%20appliance. Version: 2010. – letzter Zugriff: 14.04.2010
- [GR05] GAUSEMEIER, J. ; REDENIUS, A.: Entwicklung mechatronischer Systeme. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel VI 4, S. 547 – 575
- [GR09] GANTZ, J. F. ; REINSEL, D.: *As the Economy Contracts, the Digital Universe Expands*. IDC White Paper - sponsored by EMC Corporation. http://idcdocserv.com/EMC_MMWP_Digital_Universe. Version: Mai 2009. – letzter Zugriff: 02.10.09

- [GR10] GANTZ, J. F. ; REINSEL, D.: *The Digital Universe Decade - Are you Ready?* IDC White Paper - sponsored by EMC Corporation. <http://idcdocserv.com/925>. Version: Mai 2010. – letzter Zugriff: 10.10.2010
- [GRC⁺07] GANTZ, J. F. ; REINSEL, D. ; CHUTE, C. ; SCHLICHTING, W. ; MCARTHUR, J. ; MINTON, S. ; XHENETI, I. ; TONCHEVA, A. ; MANFREDIZ, A.: *The Expanding Digital Universe - A Forecast of Worldwide Information Growth Through 2010*. IDC White Paper - sponsored by EMC Corporation. <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>. Version: März 2007. – letzter Zugriff: 02.10.2009
- [Gro04] GRONAU, N.: *Enterprise Resource Planning und Supply Chain Management: Architektur und Funktionen*. München : Oldenbourg Wissenschaftsverlag GmbH, 2004
- [Gro08] GRONAU, N.: *ERP-System, Anbieter*. Artikel in Enzyklopädie der Wirtschaftsinformatik - Online-Lexikon. <http://www.encyklopaedie-der-wirtschaftsinformatik.de/wi-encyklopaedie/lexikon/informationssysteme/Sektorspezifische-Anwendungssysteme/Enterprise-Resource-Planning-System/ERP-System--Anbieter>. Version: September 2008. – letzter Zugriff: 18.03.2010
- [Gün05] GÜNZLER, A. R.: *Integrationskonzepte in der modellbasierten Produktentwicklung*, Lehrstuhl für Informatik an der Technischen Universität München, Diss., 2005
- [Gus08] GUSIG, L.-O.: Kosten in der Konstruktion. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 20, S. 358 – 384
- [Hab08] HABERHAUER, H.: Methodisches Konstruieren. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Carl Hanser Verlag, 2008, Kapitel 11, S. 196 – 216
- [Ham05] HAMMERSCHALL, U.: *Verteilte Systeme und Anwendungen: Architekturkonzepte, Standards und Middleware-Technologien*. München : Pearson Studium, 2005
- [Hat07] HATCHER, E.: *The Mail Archive: Does solr handle hierarchical facets?* Website. <http://www.mail-archive.com/solr-user@lucene.apache.org/msg07826.html>. Version: 2007. – letzter Zugriff: 08.07.2010
- [Haw04] HAWKING, D.: Challenges in Enterprise Search. In: *Proceedings of the 15th Australasian Database Conference (ADC)* Bd. 27. Dunedin, New Zealand :

- Australian Computer Society, Inc., 2004, S. 15 – 24
- [HB88] HEINRICH, L. J. ; BURGHOLZER, P.: *Informationsmanagement: Planung, Überwachung und Steuerung der Informations-Infrastruktur*. 2., überarbeitete und ergänzte Auflage. München, Wien : R. Oldenbourg Verlag GmbH, 1988
- [Hea09] HEARST, M. A.: *Search User Interfaces*. New York, NY, USA : Cambridge University Press, 2009
- [HEE+02] HEARST, M. ; ELLIOTT, A. ; ENGLISH, J. ; SINHA, R. ; SWEARINGEN, K. ; YEE, K.-P.: Finding the Flow in Web Site Search. In: *Communications of the ACM* 45 (2002), Nr. 9, S. 42 – 49
- [Hei05] HEISIG, P.: *Integration von Wissensmanagement in Geschäftsprozesse*, Technische Universität Berlin, Diss., 2005
- [Her00] HERBST, D.: *Erfolgsfaktor Wissensmanagement*. 1. Auflage. Berlin : Cornelsen Verlag, 2000
- [HHMS06] HINDEL, B. ; HÖRMANN, K. ; MÜLLER, M. ; SCHMIED, J.: *Basiswissen Software-Projektmanagement: Aus- und Weiterbildung zum Certified Project Management nach iSQI-Standard*. 2., überarbeitete und erweiterte Auflage. Heidelberg : dpunkt.verlag GmbH, 2006
- [HK02] HAIK, Y. ; KILANI, M.: *Essentials of Pro/ENGINEER*. Canada : Wadsworth Group, Brooks/Cole, Thomson Learning, 2002
- [HM04] HEESOM, D. ; MAHDJOUBI, L.: Trends of 4D CAD Applications for Construction Planning. In: *Construction Management and Economics* 22 (2004), Februar, S. 171 – 182
- [Hod00] HODGE, G.: *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, DC : The Digital Library Federation, 2000
- [Hos08] HOSTETTER, C.: *Hierarchical Faceting*. Online-Beitrag im Lucene-Forum. <http://lucene.472066.n3.nabble.com/Hierarchical-Faceting-td490420.html#a490420>. Version: 2008. – letzter Zugriff: 08.07.2010
- [HP00] HERTZUM, M. ; PEJTERSEN, A. M.: The Information-Seeking Practices of Engineers: Searching for Documents as well as for People. In: *Information Processing and Management* 36 (2000), S. 761 – 778
- [HP02] HAPERIN, D. ; PECKER, E.: Iterated Snap Rounding. In: *Computational Geometry: Theory and Applications* 23 (2002), Nr. 2, S. 209 – 225

- [HPR00] HAN, J. ; PRATT, M. ; REGLI, W. C.: Manufacturing Feature Recognition from Solid Models: A Status Report. In: *IEEE Transactions on Robotics and Automation* 16 (2000), December, Nr. 6, S. 782 – 796
- [HS09] HEINRICH, L. J. ; STELZER, D.: *Informationsmanagement: Grundlagen, Aufgaben, Methoden*. 9., vollst. überarb. Auflage. München : Oldenbourg Wissenschaftsverlag GmbH, 2009
- [HSKK01] HILAGA, M. ; SHINAGAWA, Y. ; KOHMURA, T. ; KUNII, T. L.: Topology Matching for Fully Automatic Similarity Estimation of 3D Shapes. In: *SIGGRAPH: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. New York, USA : ACM Press, 2001, S. 203 – 212
- [HSW89] HENRICH, A. ; SIX, H.-W. ; WIDMAYER, P.: The LSD Tree: Spatial Access to Multidimensional Point and Nonpoint Objects. In: APERS, P. M. G. (Hrsg.) ; WIEDERHOLD, G. (Hrsg.): *Proceedings of the Fifteenth International Conference on Very Large Data Bases (VLDB)*. San Francisco : Morgan Kaufmann, 1989, S. 45 – 53
- [Hub84] HUBKA, V.: *Theorie Technischer Systeme: Grundlagen einer wissenschaftlichen Konstruktionslehre*. 2., völlig neu bearbeitete und erweiterte Auflage. Berlin, Heidelberg : Springer-Verlag, 1984
- [HZ00] HÜNEKE, K. ; ZIMMERMANN, B.: Skill-Datenbanken. In: *Computer Fachwissen* 8 - 9 (2000), S. 51 – 55
- [IJ05] INGWERSEN, P. ; JÄRVELIN, K.: *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, The Netherlands : Springer-Verlag, 2005
- [IJL⁺05] IYER, N. ; JAYANTI, S. ; LOU, K. ; KALYANARAMAN, Y. ; RAMANI, K.: Three-Dimensional Shape Searching: State-of-the-Art Review and Future Trends. In: *Computer-Aided Design* 37 (2005), S. 509 – 530
- [IMT99] IGARASHI, T. ; MATSUOKA, S. ; TANAKA, H.: Teddy: A Sketching Interface for 3D Freeform Design. In: *Proceedings of ACM SIGGRAPH*. Los Angeles, USA, 1999, S. 409 – 416
- [Ins85] INSELBERG, A.: The Plane with Parallel Coordinates. In: *The Visual Computer* 1 (1985), Nr. 4, S. 69 – 91
- [Ins05] INSELBERG, A.: Visualization of Concept Formation and Learning. In: *Kybernetes* 34 (2005), Nr. 1/2, S. 151 – 166
- [Int10a] INTERNATIONAL BUSINESS MACHINES (IBM) CORPORATION: *OmniFind*. Website. <http://www-01.ibm.com/software/data/enterprise->

- search/. Version: 2010. – letzter Zugriff: 13.04.2010
- [Int10b] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *SC4-NG - TC184/SC4 Industrial Data*. Website. <http://ng.tci84-sc4.org>. Version: 2010. – letzter Zugriff: 01.08.2010
- [Int10c] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *Schematic representation of ISO deliverables*. Website. http://www.iso.org/iso/standards_development/processes_and_procedures/deliverables/deliverables_schema-2.htm. Version: 2010. – letzter Zugriff: 02.08.2010
- [Jan10] JANSEN, R.: *CouchDB - ein angesagter Vertreter der NoSQL-Datenbanken: Freiheit für die Dokumente*. Online-Artikel. <http://www.heise.de/developer/artikel/CouchDB-angesagter-Vertreter-der-NoSQL-Datenbanken-929079.html>. Version: Februar 2010. – letzter Zugriff: 12.02.2010
- [KAE⁺05] KARNIK, M. V. ; ANAND, D. K. ; EICK, E. ; GUPTA, S. K. ; KAVETSKY, R.: Integrated Visual and Geometric Search Tools for Locating Desired Parts in a Part Database. In: *Computer Aided Design and Applications 2* (2005), Nr. 6, S. 727 – 736
- [KBK⁺06] KRIEDEL, H. P. ; BRECHSEIN, S. ; KRÖGER, P. ; PFEIFLE, M. ; SCHUBERT, M. ; ZIMEK, A.: Density-Based Data Analysis and Similarity Search. In: PETRUSHIN, V. A. (Hrsg.) ; KHAN, L. (Hrsg.): *Multimedia Data Mining and Knowledge Discovery*. London : Springer-Verlag, 2006, S. 94 – 115
- [KCJ94] KING, D.W. ; CASTO, J. ; JONES, H.: Communication by Engineers: A Literature Review of Engineers' Information Needs, Seeking Processes, and Use / University of Tennessee; Center for Information Studies. Washington, D.C., USA, 1994. – Forschungsbericht
- [Kei99] KEIM, D. A.: Efficient Geometry-Based Similarity Search of 3D Spatial Databases. In: *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*. New York, NY, USA : ACM Press, 1999, S. 419 – 430
- [KEL⁺09] KREHMER, H. ; ECKSTEIN, R. ; LAUER, W. ; ROELOFSEN, J. ; STÖBER, C. ; TROLL, A. ; WEBER, N. ; ZAPF, J.: Coping with Multidisciplinary Product Development - A Process Model Approach. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [Kem01] KEMMERER, S. J.: Initial Graphics Exchange Specifications. Version: 2001. http://ts.nist.gov/Standards/IGES/upload/246_249.pdf. In: LIDE, D. L. (Hrsg.): *A Century of Excellence in Measurements, Standards, and*

- Technology - A Chronicle of Selected NBS/NIST Publications, 1901 - 2000*. National Technical Information Service (NTIS), 2001 (NIST Special Publication 958), 246 - 249
- [KF98] KOO, B. ; FISCHER, M.: Feasibility Study of 4D CAD in Commercial Construction / Stanford University, Center for Integrated Facility Engineering (CIFE). Stanford, Palo Alto, August 1998 (118). – Forschungsbericht
- [KFR03] KAZHDAN, M. ; FUNKHOUSER, T. ; RUSINKIEWICZ, S.: Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. In: *Eurographics Symposium on Geometry Processing* (2003), S. 167 – 175
- [KGA+05] KARNIK, M. V. ; GUPTA, S. K. ; ANAND, D. K. ; VALENTA, F. J. ; WEXLER, I. A.: Design Navigator System: A Case Study in Improving Product Development through Improved Information Management. In: *ASME Computers and Information in Engineering Conference*. Long Beach, CA, September 2005
- [KGM05] KARNIK, M. ; GUPTA, S. K. ; MAGRAB, E. B.: Geometric Algorithms for Containment Analysis of Rotational Parts. In: *Computer Aided Design* 37 (2005), Nr. 2, S. 213 – 230
- [KGS09] KIM, A. ; GWUN, O. ; SONG, J.: Extraction of 3D Feature Descriptor Using the Distribution of Normal Vectors. In: ANDREASEN, T. (Hrsg.) ; YAGER, R. R. (Hrsg.) ; BULSKOV, H. (Hrsg.) ; CHRISTIANSEN, H. (Hrsg.) ; LARSEN, H. L. (Hrsg.): *Flexible Query Answering Systems (FQAS)* Bd. 5822. Berlin, Heidelberg : Springer-Verlag, 2009 (Lecture Notes in Artificial Intelligence), S. 191 – 200
- [KKM+03] KRIEDEL, H.-P. ; KRÖGER, P. ; MASHAEL, Z. ; PFEIFLE, M. ; PÖTKE, M. ; SEIDL, T.: Effective Similarity Search on Voxelized CAD Objects. In: *Proceedings of the 8th International Conference on Database Systems for Advanced Applications (DASFAA)*. Kyoto, Japan, 2003, S. 27 – 36
- [KMW09] KREHMER, H. ; MEERKAMM, H. ; WARTZACK, S.: The Product's Degree of Maturity as a Measurement for the Efficiency of Design Iterations. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [Kol98] KOLLER, R.: *Konstruktionslehre für den Maschinenbau: Grundlagen zur Neu- und Weiterentwicklung technischer Produkte mit Beispielen*. 4., neu bearb. und erw. Auflage. Berlin, Heidelberg : Springer-Verlag, 1998
- [Kor76] KOREIMANN, D. S.: *Methoden der Informationsbedarfsanalyse*. Berlin, New York : Walter de Gruyter GmbH & Co. KG, 1976

- [KPNK03] KÖRTGEN, M. ; PARK, G.-J. ; NOVOTNI, M. ; KLEIN, R.: 3D Shape Matching with 3D Shape Contexts. In: *The 7th Central European Seminar on Computer Graphics*. Budmerice, Slovakia, April 2003
- [Krc05] KRCMAR, H.: *Informationsmanagement*. 4. Auflage. Berlin, Heidelberg : Springer-Verlag, 2005
- [KSS97] KRIEGEL, H.-P. ; SCHMIDT, T. ; SEIDL, T.: 3D Similarity Search by Shape Approximation. In: *Proceedings of the Fifth International Symposium on Large Spatial Databases (SSD)*. Berlin, Germany, 1997, S. 11 – 28
- [Kuh04] KUHLEN, R.: Information. In: KUHLEN, R. (Hrsg.) ; SEEGER, T. (Hrsg.) ; STRAUCH, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 5., völlig neu gefasste Ausgabe. München : K. G. Saur Verlag GmbH, 2004, Kapitel A1, S. 3 – 20
- [LAR05] LI, Z. ; ANDERSON, D. C. ; RAMANI, K.: Ontology-Based Design Knowledge Modeling for Product Retrieval. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the International Conference on Engineering Design (ICED)*. Melbourne, August, 15 - 18 2005
- [Lau09] LAUER, W.: Gezielte Integration von Produktmodellen in den Entwicklungsprozess. In: HENRICH, A. (Hrsg.) ; JABLONSKI, S. (Hrsg.) ; KRCMAR, H. (Hrsg.) ; LINDEMANN, U. (Hrsg.) ; MEERKAMM, H. (Hrsg.) ; RIEG, F. (Hrsg.): *Flexible Prozessunterstützung in der Produktentwicklung: Prozesse - Daten - Navigation*. Aachen : Shaker Verlag GmbH, 2009, Kapitel 4.2, S. 63 – 80
- [LB04] LOVE, D. ; BARTON, J.: Aspects of Design Retrieval Performance Using Automatic GT Coding of 2D Engineering Drawings. In: *4th International Conference on Integrated Design and Manufacture in Mechanical Engineering*. Bath, April 2004
- [LBM04] LIU, R. ; BABA, T. ; MASUMOTO, D.: Attributed Graph Matching Based Engineering Drawings Retrieval. In: MARINAI, S. (Hrsg.) ; DENGEL, A. (Hrsg.): *Document Analysis Systems VI - Proceedings of the 6th International Workshop (DAS)*. Florence, Italy : Springer Verlag, Berlin, Heidelberg, September 2004, S. 378 – 388
- [Leh09] LEHNER, F.: *Wissensmanagement: Grundlagen, Methoden und technische Unterstützung*. 3., aktualisierte und erweiterte Auflage. München, Wien : Carl Hanser Verlag, 2009
- [Les01] LESZINSKI, C. M.: *Ein Visualisierungs- und Navigationsassistent für Produktstrukturen in der Produktentwicklung*. Bochum : Shaker Verlag GmbH,

2001

- [LFB96] LIN, J. ; FOX, M. S. ; BILGIC, T.: A Requirement Ontology for Engineering Design. In: *Proceedings of 3rd International Conference on Concurrent Engineering* Technomic Publishing Company, Inc., 1996, S. 343 – 351
- [Lin05] LINDEMANN, U.: *Methodische Entwicklung technischer Produkte: Methoden flexibel und situationsgerecht anwenden*. Berlin, Heidelberg : Springer-Verlag, 2005
- [LJ10] LI, B. ; JOHAN, H.: View Context: A 3D Model Feature for Retrieval. In: BOLL, S. (Hrsg.) ; TIAN, Q. (Hrsg.) ; ZHANG, L. (Hrsg.) ; ZHANG, Z. (Hrsg.) ; CHEN, Y.-P. P. (Hrsg.): *Advances in Multimedia Modeling* Bd. 5916. Berlin, Heidelberg : Springer-Verlag, 2010 (Lecture Notes in Computer Science), S. 185 – 195
- [LJI+03] LOU, K. ; JAYANTI, S. ; IYER, N. ; KALYANARAMAN, Y. ; PRABHAKAR, S. ; RAMANI, K.: A Reconfigurable 3D Engineering Shape Search System Part II: Database Indexing, Retrieval and Clustering. In: *Proceedings of ASME DETC' 03 23rd Computers and Information in Engineering (CIE) Conference*. Chicago, Illinois, September, 2 - 6 2003
- [LMD+06] LIU, S. ; MCMAHON, C. A. ; DARLINGTON, M. J. ; CULLEY, S. J. ; WILD, P. J.: A Computational Framework for Retrieval of Document Fragments Based on Decomposition Schemes in Engineering Information Management. In: *Advanced Engineering Informatics* 20 (2006), S. 401 – 413
- [LMD+07] LIU, S. ; MCMAHON, C. ; DARLINGTON, M. ; CULLEY, S. ; WILD, P.: EDCMS: A Content Management System for Engineering Documents. In: *International Journal of Automation and Computing* 4 (2007), January, Nr. 1, S. 56 – 70
- [LPR04] LOU, K. ; PRABHAKAR, S. ; RAMANI, K.: Content-Based Three-Dimensional Engineering Shape Search. In: *Proceedings of the 20th International Conference on Data Engineering (ICDE)*. Boston, USA, March 30 - April 2 2004, 754 - 765
- [LRR07] LI, Z. ; RASKIN, V. ; RAMANI, K.: A Methodology of Engineering Ontology Development for Information Retrieval. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 16th International Conference on Engineering Design (ICED)*. Paris, France, August, 28 - 31 2007
- [LS96] LIPSON, H. ; SHPITALNI, M.: Optimization-Based Reconstruction of a 3D Object from a Single Freehand Line Drawing. In: *Computer-Aided Design* 28 (1996), Nr. 8, S. 651 – 663

- [LSLL06] LEE, T. ; SHIM, J. ; LEE, H. ; LEE, S.: A Pragmatic Approach to Model and Exploit the Semantics of Product Information. In: *Journal on Data Semantics (LNCS 4244)* (2006), S. 242 – 266
- [Mac80] MACHLUP, F.: *Knowledge: Its Creation, Distribution, And Economic Significance*. Bd. 1: *Knowledge and Knowledge Production*. Princeton : Princeton University Press, 1980
- [Man04] MANECKE, H.-J.: Klassifikation, Klassieren. In: KUHLEN, R. (Hrsg.) ; SEEGER, T. (Hrsg.) ; STRAUCH, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 5., völlig neu gefasste Ausgabe. München : K. G. Saur Verlag GmbH, 2004, Kapitel B1, S. 127 – 140
- [Mar06] MARCHIONINI, G.: Exploratory Search: From Finding to Understanding. In: *Communications of the ACM* 49 (2006), April, Nr. 4, S. 41 – 46
- [MBA07] MOUGENOT, C. ; BOUCHARD, C. ; AOUSSAT, A.: A Study of Designers' Cognitive Activity in Design Informational Phase. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 16th International Conference on Engineering Design (ICED)*. Paris, France, August, 28 - 31 2007. – CD-ROM Proceedings, Paper 224
- [MCF02] MIN, P. ; CHEN, J. ; FUNKHOUSER, F.: A 2D Sketch Interface for a 3D Model Search Engine. In: *SIGGRAPH 2002*. San Antonio, Texas, July 2002
- [Mew73] MEWES, D.: *Der Informationsbedarf im konstruktiven Maschinenbau*. Düsseldorf : VDI-Verlag GmbH, 1973
- [MHC09] MCALPINE, H. ; HICKS, B. ; CULLEY, S.: Comparing the Information Content of Formal and Informal Design Documents: Lessons for More Complete Design Records. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [MHKF03] MIN, P. ; HALDERMAN, J. A. ; KAZHDAN, M. ; FUNKHOUSER, T. A.: Early Experiences with a 3D Model Search Engine. In: *Proceedings of Web3D Symposium*. St. Malo, France, March 2003, S. 7 – 18
- [MK97] MCCRICKARD, D. S. ; KEHOE, C. M.: Visualizing Search Results using SQWID. In: *Proceedings of the Sixth International World Wide Web Conference*. Santa Clara, California, April 1997
- [ML05] MASRY, M. ; LIPSON, H.: A Sketch-Based Interface for Iterative Design and Analysis of 3D Objects. In: *2nd Eurographics Workshop on Sketch-Based Interfaces and Modeling* (2005)

- [MLC04] McMAHON, C. ; LOWE, A. ; CULLEY, S.: Waypoint: An Integrated Search and Retrieval System for Engineering Documents. In: *Journal of Computing and Information Science in Engineering* 4 (2004), December, Nr. 4, S. 329 – 338
- [MM04] MUKHERJEE, R. ; MAO, J.: Enterprise Search: Tough Stuff. In: *ACM Queue* 2 (2004), Nr. 2, S. 36 – 46
- [MNPT05] MANOLOPOULOS, Y. ; NANOPOULOS, A. ; PAPADOPOULOS, A. N. ; THEODORIDIS, Y.: *R-Trees: Theory and Applications*. Berlin, Heidelberg : Springer-Verlag, 2005
- [Moe09] MOEHRINGER, S.: Mechatronic Design - Historical Background and Challenges for the Future. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [MP97] MAHER, M. L. ; PU, P.: *Issues and Applications of Case-Based Reasoning in Design*. Mahwah, NJ : Lawrence Erlbaum Associates, 1997
- [MPSR01] McWHERTER, D. ; PEABODY, M. ; SHOKOUFANDEH, A. ; REGLI, W. C.: Database Techniques for Archival of Solid Models. In: *Proceedings of the sixth ACM Symposium on Solid Modeling and Applications (SMA)*. New York, NY, USA : ACM Press, 2001, S. 78 – 87
- [MRB⁺08] MOELLER, E. ; REUTER, M. ; BLECK, W. u. a. ; MOELLER, E. (Hrsg.): *Handbuch Konstruktionswerkstoffe - Auswahl, Eigenschaften, Anwendung*. München : Carl Hanser Verlag, 2008
- [MRS08] MANNING, C. D. ; RAGHAVAN, P. ; SCHÜTZE, H.: *Introduction to Information Retrieval*. Cambridge, New York : Cambridge University Press, 2008
- [MT03] MCKINSEY & COMPANY ; TU DARMSTADT, INSTITUT PTW: *HAWK 2015 - Wissensbasierte Veränderung der automobilen Wertschöpfungskette*. Heinrich Druck + Medien GmbH, Frankfurt a. M., 2003. – Verband der Automobilindustrie e. V. (Hrsg.); Materialien zur Automobilindustrie Nr. 30
- [MT08] MERKEL, M. ; THOMAS, K.-H.: *Taschenbuch der Werkstoffe*. 7. Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008
- [Mül90] MÜLLER, J.: *Arbeitsmethoden der Technikwissenschaften: Systematik, Heuristik, Kreativität*. Berlin, Heidelberg : Springer-Verlag, 1990
- [Mül09] MÜLLER, P.: *Mathematisches Seminar: Spektrale Graphentheorie*. Website. <http://www.mathematik.uni-muenchen.de/~mueller/lehre/09/spek-graphen.php>. Version: 2009. – letzter Zugriff: 18.08.2010

- [MWJV07] MUHS, D. ; WITTEL, H. ; JANNASCH, D. ; VOSSIEK, J.: *Roloff / Matek Maschinenelemente: Normung, Berechnung, Gestaltung*. 18., vollständig überarbeitete Auflage. Wiesbaden : Friedrich Vieweg & Sohn Verlag / GWV Fachverlage GmbH, 2007
- [Nat08] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST): *The Initial Graphics Exchange Specification (IGES)*. Website. <http://ts.nist.gov/standards/iges/>. Version: 2008. – letzter Zugriff: 01.08.2010
- [Neu01] NEUMANN, G.: Informationsextraktion. In: CARSTENSEN, K. U. (Hrsg.) ; EBERT, Ch. (Hrsg.) ; ENDRISS, C. (Hrsg.) ; JEKAT, S. (Hrsg.) ; KLABUNDE, R. (Hrsg.) ; LANGER, H. (Hrsg.): *Computerlinguistik und Sprachtechnologie - Eine Einführung*. Heidelberg : Spektrum Akademischer Verlag, 2001, S. 448 – 516
- [NHS84] NIEVERGELT, J. ; HINTERBERGER, H. ; SEVCIK, K. C.: The Grid File: An Adaptable, Symmetric Multikey File Structure. In: *ACM Transactions on Database Systems* 9 (1984), Nr. 1, S. 38 – 71
- [NK04] NOVOTNI, M. ; KLEIN, R.: Shape Retrieval Using 3D Zernike Descriptors. In: *Computer-Aided Design* 36 (2004), Nr. 11, S. 1047 – 1062
- [NNS96] NABIL, M. ; NGU, A. H. H. ; SHEPHERD, J.: Picture Similarity Retrieval Using the 2D Projection Interval Representation. In: *IEEE Transactions on Knowledge and Data Engineering* 8 (1996), August, Nr. 4, S. 533 – 539
- [NS06] NIEDERMEIER, S. ; SCHOLZ, M.: *Java und XML: Grundlagen, Einsatz, Referenz*. 1. Auflage. Bonn : Galileo Press, 2006
- [NT97] NONAKA, I. ; TAKEUCHI, H.: *Die Organisation des Wissens: Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*. Frankfurt am Main : Campus Verlag GmbH, 1997
- [Nud10] NUDELMAN, G.: *More Like This: A Design Pattern*. Online-Artikel. <http://www.uxmatters.com/mt/archives/2010/01/more-like-this-a-design-pattern.php>. Version: January 2010. – letzter Zugriff: 09.05.2010
- [OFCD01] OSADA, R. ; FUNKHOUSER, T. ; CHAZELLE, B. ; DOBKIN, D.: Matching 3D Models with Shape Distributions. In: *Shape Modeling International* (2001), May, S. 154 – 166
- [OJ93] O'DAY, V. L. ; JEFFRIES, R.: Orienteering in an Information Landscape: How Information Seekers Get from Here to There. In: ASHLUND, S. (Hrsg.) ; HENDERSON, A. (Hrsg.) ; HOLLNAGEL, E. (Hrsg.) ; MULLET, K. (Hrsg.) ; WHITE, T. (Hrsg.): *Proceedings of the INTERCHI' 93 conference on human factors in computing systems*. Amsterdam : IOS Press Amsterdam, The

- Netherlands, April 1993, S. 438 – 445
- [OMT03] OHBUCHI, R. ; MINAMITANI, T. ; TAKEI, T.: Shape-Similarity Search of 3D Models by Using Enhanced Shape Functions. In: *Proceedings of the Theory and Practice of Computer Graphics (TPCG)*. Washington DC, USA : IEEE Computer Society, 2003, S. 97 – 106
- [Opi66] OPITZ, H.: *Werkstückbeschreibendes Klassifizierungssystem*. Essen : Verlag W. Girardet, 1966
- [Par10] PARAMETRIC TECHNOLOGY CORPORATION: *PTC*. Website. <http://www.ptc.com>. Version: 2010. – letzter Zugriff: 30.07.2010
- [PBFG07] PAHL, G. ; BEITZ, W. ; FELDHUSEN, J. ; GROTE, K.-H.: *Konstruktionslehre: Grundlagen erfolgreicher Produktentwicklung - Methoden und Anwendung*. 7. Auflage. Berlin, Heidelberg : Springer-Verlag, 2007
- [Pie00] PIERER, H. von: *Siemens - The E-Driven Company*. Pressekonferenz, München. <http://www.tse.de/papiere/internet%20und%20netze/ecommerce/pierer.html>. Version: Oktober 2000. – letzter Zugriff: 10.01.2010
- [PMHR09] PAVKOVIC, N. ; MARJANOVIC, D. ; HOIC, M. ; ROHDE, D.: An Explorer Style Interface for Design Knowledge Indexing Based on Taxonomies. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August 24 - 27 2009
- [Pol66] POLANYI, M.: *The Tacit Dimension*. Garden City, New York : Doubleday & Company, Inc., 1966
- [Pon07] PONZI, L. J.: Knowledge Management: Birth of a Discipline. In: KOENIG, M. E. D. (Hrsg.) ; SRIKANTIAH, T. K. (Hrsg.): *Knowledge Management Lessons Learned: What Works and What Doesn't*. Second printing. Medford, New Jersey : Information Today, Inc., 2007, S. 9 – 26
- [PR05] PU, J. ; RAMANI, K.: A 2D Sketch-Based User Interface for 3D CAD Model Retrieval. In: *Computer Aided Design and Application* 2 (2005), Nr. 6, S. 717 – 727
- [PR06] PU, J. ; RAMANI, K.: ShapeLab: A Unified Framework for 2D & 3D Shape Retrieval. In: *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, S. 1072 – 1079
- [Pra01] PRATT, M. J.: Introduction to ISO 10303 - the STEP Standard for Product Data Exchange. In: *Journal of Computing and Information Science in Engineering* 1 (2001), March, Nr. 1, S. 102 – 103

- [Pro09] PROSTEP iViP VEREIN: *JT ist ISO Standard*. Pressemitteilung. http://www.prostep.org/uploads/media/PI_JT_de.pdf. Version: September 2009. – letzter Zugriff: 02.08.2010
- [PRR06] PROBST, G. ; RAUB, S. ; ROMHARDT, K.: *Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen*. 5., überarbeitete Auflage. Wiesbaden : Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, 2006
- [Prz08a] PRZYWARA, R.: Informationsmanagement in der Konstruktion. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Carl Hanser Verlag, 2008, Kapitel 10, S. 188 – 194
- [Prz08b] PRZYWARA, R.: Wissensmanagement. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Carl Hanser Verlag, 2008, Kapitel 9, S. 182 – 187
- [PU99] PARK, J. H. ; UM, B. S.: A New Approach to Similarity Retrieval of 2-D Graphic Objects Based on Dominant Shapes. In: *Pattern Recognition Letters* 20 (1999), S. 591 – 616
- [Pul04] PULM, Udo: *Eine systemtheoretische Betrachtung der Produktentwicklung*, Fakultät für Maschinenwesen an der Technischen Universität München, Diss., Juni 2004
- [QR03] QIN, X. ; REGLI, W. C.: A Study in Applying Case-Based Reasoning to Engineering Design: Mechanical Bearing Design. In: *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 17 (2003), Nr. 3, S. 235 – 252
- [Rat85] RAT DER EUROPÄISCHEN GEMEINSCHAFTEN: *Richtlinie 85/374/EWG zur Angleichung der Rechts- und Verwaltungsvorschriften der Mitgliedstaaten über die Haftung für fehlerhafte Produkte*. Website. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31985L0374:de:HTML>. Version: Juli 1985. – letzter Zugriff: 02.12.2009
- [Reu08] REUTER, M.: Produktentstehung. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Fachbuchverlag Leipzig im Carl Hanser Verlag, 2008, Kapitel 25, S. 498 – 557
- [RG02] *Kapitel 6: Query-By-Example (QBE)*. In: RAMAKRISHNAN, R. ; GEHRKE, J.: *Database Management Systems*. 3 edition. McGraw-Hill, Inc., 2002, S. 177 – 192
- [RHAS00] REGLI, W. C. ; HU, X. ; ATWOOD, M. ; SUN, W.: A Survey of Design Rationale Systems: Approaches, Representation, Capture and Retrieval. In: *Engineering with Computers* 16 (2000), S. 209 – 235

- [RHC97] RUI, Y. ; HUANG, T. S. ; CHANG, S.-F.: Image Retrieval: Past, Present, and Future. In: *International Symposium on Multimedia Information Processing*, 1997
- [Rij79] RIJSBERGEN, C. J.: *Information Retrieval*. Second Edition. London : Butterworths, 1979
- [Rot00] ROTH, K.: *Konstruieren mit Konstruktionskatalogen - Band I: Konstruktionslehre*. 3. Auflage. Berlin, Heidelberg : Springer-Verlag, 2000
- [Rot01] ROTH, K.: *Konstruieren mit Konstruktionskatalogen - Band II: Konstruktionskataloge*. 3. Auflage. Berlin, Heidelberg : Springer-Verlag, 2001
- [RSW00] RUDOLPH, D. ; STÜRNICKEL, T. ; WEISSENBERGER, L.: *DXF intern: Das Zeichnungsaustauschformat DXF vollständig dokumentiert mit DXF-Werkzeugkasten auf Diskette*. Essen : CR/LF GmbH, 2000
- [RTG00] RUBNER, Y. ; TOMASI, C. ; GUIBAS, L. J.: The Earth Mover's Distance as a Metric for Image Retrieval. In: *International Journal of Computer Vision* 40 (2000), November, Nr. 2, S. 99 – 121
- [Rud00] RUDOLPH, D.: *AutoCAD-Objekte*. Düsseldorf : SYBEX-Verlag GmbH, 2000
- [Rus08] RUST, W.: Finite-Elemente-Methode. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Carl Hanser Verlag, 2008, Kapitel 27, S. 580 – 596
- [RYD01] RAMESH, M. ; YIP-HOI, D. ; DUTTA, D.: Feature Based Shape Similarity Measurement for Retrieval of Mechanical Parts. In: *Journal of Computing and Information Science in Engineering* 1 (2001), September, Nr. 3, S. 245 – 256
- [Sam06] SAMET, H.: *Foundations of Multidimensional and Metric Data Structures*. San Francisco : Morgan Kaufmann, 2006
- [SAP09] SAP AG: *SAP Glossary*. Online-Glossar. http://help.sap.com/saphelp_glossary/de/index.htm. Version: Oktober 2009. – letzter Zugriff: 18.03.2010
- [SAP10] SAP AG: *SAP-Bibliothek - Dokumentenverwaltung*. Online - SAP Dokumentation. http://help.sap.com/erp2005_ehp_04/helpdata/DE/c1/1c28a543c711d1893e0000e8323c4f/frameset.htm. Version: 2010. – letzter Zugriff: 25.03.2010
- [Sar06] SARACEVIC, T.: Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. In: *Advances in Librarianship* 30 (2006), S. 3 – 71

- [SC07] SARKAR, P. ; CHAKRABARTI, A.: Understanding Search in Design. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 16th International Conference on Engineering Design (ICED)*. Paris, France, August, 28 - 31 2007
- [Sch87] SCHWAIGER, L.: *CAD-Begriffe - Ein Lexikon*. Berlin, Heidelberg : Springer-Verlag, 1987
- [Sch98] SCHLÜTER, O.: *VRML: Sprachmerkmale, Anwendungen, Perspektiven*. 1. Auflage. Köln : O'Reilly Verlag, 1998
- [Sch99] SCHÖTTNER, J.: *Produktdatenmanagement in der Fertigungsindustrie: Prinzip - Konzepte - Strategien*. München, Wien : Carl Hanser Verlag, 1999
- [Sch02a] SCHICHTEL, M.: *Produktdatenmodellierung in der Praxis*. München, Wien : Carl Hanser Verlag, 2002
- [Sch02b] SCHWANKL, L.: *Analyse und Dokumentation in den frühen Phasen der Produktentwicklung*, Lehrstuhl für Produktentwicklung der Technischen Universität München, Diss., 2002
- [Sch04] *Kapitel Dokumentenmanagement*. In: SCHÜTZ, T.: *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 5., völlig neu gefasste Ausgabe. München : K. G. Saur Verlag GmbH, 2004, S. 339 – 349
- [Sch05a] SCHABACKER, M. ; VAJNA, S. (Hrsg.): *Solid Edge - kurz und bündig: Grundlagen für Einsteiger*. 1. Auflage. Wiesbaden : Friedr. Vieweg & Sohn Verlag, GWV Fachverlage GmbH, 2005
- [Sch05b] SCHÄPPI, B.: Integrierte Produktentwicklung - Entwicklungsprozesse zielorientiert und effizient gestalten. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel I 1, S. 3 – 27
- [Sch06] SCHMITT, I.: *Ähnlichkeitssuche in Multimedia-Datenbanken: Retrieval, Suchalgorithmen und Anfragebehandlung*. München, Wien : Oldenbourg Verlag, 2006
- [Sch08] SCHUSTER, M.: *Ermittlung von Kontextdaten aus ERP-Systemen zur Unterstützung von Suchanfragen im Konstruktionsprozess*, Universität Bamberg, Fakultät WIAI, Lehrstuhl für Medieninformatik, Diplomarbeit, April 2008
- [SCR06] SCRA: *STEP Application Handbook: ISO 10303: Version 3*. North Charleston, SC, USA, June 2006. <http://www.tc184->

- sc4.org/SC4_Open/SC4_Standards_Developers_Info/Files/STEP_application_handbook_63006.pdf. – letzter Zugriff: 20.02.2011
- [SD01] SMITH, J. S. ; DUFFY, A. H. B.: Re-Using Knowledge: Why, What and Where. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the International Conference on Engineering Design (ICED)*. Glasgow, Scotland, August, 21 - 23 2001, S. 227 – 234
- [SGR97] STICKEL, E. (Hrsg.) ; GROFFMANN, H.-D. (Hrsg.) ; RAU, K.-H. (Hrsg.): *Gabler Wirtschaftsinformatik-Lexikon*. Wiesbaden : Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, 1997
- [SHI95] SHUM, H. ; HEBERT, M. ; IKEUCHI, K.: On 3D Shape Similarity / Carnegie Mellon University. Pittsburgh, Pennsylvania, November 1995 (CMU-CS-95-212). – Forschungsbericht
- [Shn94] SHNEIDERMAN, B.: Dynamic Queries for Visual Information Seeking. In: *IEEE Software* 11 (1994), Nr. 6, S. 70 – 77
- [Sie09] SIEMENS PLM SOFTWARE INC.: *JT File Format Reference Version 8.1 Rev-D*. first edition, 2009
- [Sie10a] SIEMENS INDUSTRY SOFTWARE GMBH & CO. KG: *Geolus Search Application*. Website für Zugang zu Testversion. http://www.plm.automation.siemens.com/de_de/products/open/geolus/index.shtml. Version: 2010. – letzter Zugriff: 03.04.2010
- [Sie10b] SIEMENS INDUSTRY SOFTWARE GMBH & CO. KG: *NX*. Website. http://www.plm.automation.siemens.com/de_de/products/nx/index.shtml. Version: 2010. – letzter Zugriff: 30.07.2010
- [Sie10c] SIEMENS INDUSTRY SOFTWARE GMBH & CO. KG: *Solid Edge*. Website. http://www.plm.automation.siemens.com/de_de/products/velocity/solidedge/index.shtml. Version: 2010. – letzter Zugriff: 30.07.2010
- [Sie10d] SIEMENS INDUSTRY SOFTWARE GMBH & CO. KG: *Solid Edge with Synchronous Technology*. Online-FAQ-Artikel. http://www.plm.automation.siemens.com/de_de/Images/solid_edge_ST_FAQ_tcm73-62427.pdf. Version: 2010. – letzter Zugriff: 30.07.2010
- [Sie10e] SIEMENS PLM SOFTWARE INC.: *Siemens PLM Components - JT Open*. Website. http://www.plm.automation.siemens.com/en_us/products/open/jtopen/index.shtml. Version: 2010. – letzter Zugriff: 02.08.2010
- [SM87] SALTON, G. ; MCGILL, M. J.: *Information Retrieval - Grundlegendes für Informationswissenschaftler*. Hamburg, New York : McGraw-Hill, Inc., 1987

- [Soe94] SOERGEL, D.: Indexing and retrieval performance: The logical evidence. In: *Journal of the American Society for Information Science* 45 (1994), Nr. 8, S. 589 – 599
- [SP05] SHNEIDERMAN, B. ; PLAISANT, C.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Fourth Edition. San Francisco : Pearson Addison Wesley, 2005
- [SP07] SCHMITT-THOMAS, Kh. G. ; PETROPOULOS, A.: *werkstoffe.de - Bayerisches Online-Materialinformationssystem*. Website. <http://www.werkstoffe.de>. Version: 2007. – letzter Zugriff: 19.06.2010
- [Spe01] SPENCE, Robert: *Information Visualization*. Harlow, London, New York : ACM Press (Addison-Wesley), 2001
- [SRSB00] SZYKMAN, S. ; RACZ, J. ; SRIRAM, R. D. ; BOCHENEK, C.: Web-Based System for Design Artifact Modeling. In: *Journal of Design Studies* 21 (2000), S. 145 – 165
- [SSB+00] SZYKMAN, S. ; SRIRAM, R. D. ; BOCHENEK, C. ; RACZ, J. W. ; SENFAUTE, J.: Design Repositories: Next-Generation Engineering Design Databases / Nationale Institute of Standards and Technology (NIST). Version: July 2000. http://www.mel.nist.gov/publications/view_pub.cgi?pub_id=821539. 2000. – Forschungsbericht
- [SSBR98] SZYKMAN, S. ; SRIRAM, R. D. ; BOCHENEK, C. ; RACZ, J.: The NIST Design Repository Project. In: *Advances in Soft Computing - Engineering Design and Manufacturing*. London : Springer-Verlag, 1998
- [SSGD03] SUNDAR, H. ; SILVER, D. ; GAGVANI, N. ; DICKINSON, S.: Skeleton Based Shape Matching and Retrieval. In: *SMI* (2003), S. 130 – 139
- [SSR01] SZYKMAN, S. ; SRIRAM, R. ; REGLI, W.: The Role of Knowledge in Next-Generation Product Development Systems. In: *Journal of Computation and Information Science in Engineering* 1 (2001), Nr. 1, S. 3 – 11
- [Sta09] STANDARDIZATION ADMINISTRATION OF PR. CHINA & DEUTSCHES INSTITUT FÜR NORMUNG E. V.: *Sachmerkmale*. Homepage des Deutsch-Chinesischen Normeninformationsportals. http://www.standards-portal.de/web_de/topics_de/product_property_de. Version: 2009. – letzter Zugriff: 02.03.2010
- [Ste01] STELZER, D.: Informationsbedarf. In: MERTENS, P. (Hrsg.): *Lexikon der Wirtschaftsinformatik*. 4., völlig neu bearb. und erw. Auflage. Berlin : Springer-Verlag, 2001, S. 238 – 239

- [Ste05] STEINSCHADEN, J.: Entwicklung mechanischer Systeme. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel VI 1, S. 481 – 505
- [Sto07] STOCK, Wolfgang G.: *Information Retrieval: Informationen suchen und finden*. München, Wien : Oldenbourg Wissenschaftsverlag GmbH, 2007
- [Stö09] STÖBER, C.: Einfluss der Aspekte des Design for X auf den Produktentwicklungsprozess. In: HENRICH, A. (Hrsg.) ; JABLONSKI, S. (Hrsg.) ; KRCMAR, H. (Hrsg.) ; LINDEMANN, U. (Hrsg.) ; MEERKAMM, H. (Hrsg.) ; RIEG, F. (Hrsg.): *Flexible Prozessunterstützung in der Produktentwicklung: Prozesse - Daten - Navigation*. Aachen : Shaker Verlag GmbH, 2009, Kapitel 5.3, S. 141 – 160
- [Str08] STRASSBURG, K.: Schutzrechte in der Konstruktion. In: CONRAD, K.-J. (Hrsg.): *Taschenbuch der Konstruktionstechnik*. 2., aktualisierte Auflage. München : Carl Hanser Verlag, 2008, Kapitel 30, S. 628 – 635
- [Sut96] SUTTON, M. J. D.: *Document Management for the Enterprise: Principles, Techniques, and Applications*. New York, Brisbane : John Wiley & Sons, Inc., 1996
- [SV01a] SAUPE, D. ; VRANIC, D. J.: 3D Shape Descriptor Based on 3D Fourier Transform. In: FAZEKAS, K. (Hrsg.): *Proceedings of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS)*. Budapest, Hungary, September 2001, S. 271 – 274
- [SV01b] SAUPE, D. ; VRANIC, D. V.: 3D Model Retrieval with Spherical Harmonics and Moments. In: RADIG, B. (Hrsg.) ; FLORCZYK, S. (Hrsg.): *Proceedings of the DAGM*. München, September 2001, S. 392 – 397
- [SW08] SENDLER, U. ; WAVER, V.: *CAD and PDM - Optimizing Processes by Integrating Them*. München, Wien : Carl Hanser Verlag, 2008
- [SWY75] SALTON, G. ; WONG, A. ; YANG, C. S.: A Vector Space Model for Automatic Indexing. In: *Communications of the ACM* 18 (1975), November, Nr. 11, S. 613 – 620
- [Tja78] TJALVE, E.: *Systematische Formgebung für Industrieprodukte*. Düsseldorf : VDI-Verlag, 1978. – Institute for Product Development, Technical University of Denmark
- [TS06] TÜRKER, C. ; SAAKE, G.: *Objektrelationale Datenbanken: Ein Lehrbuch*. 1. Auflage. Heidelberg : dpunkt.verlag GmbH, 2006

- [TTP04] TSAI, C.-Y. ; TIEN, F.-C. ; PAN, T.-Y.: Development of an XML-based Structural Product Retrieval System for Virtual Enterprises. In: *International Journal of Production Research* 42 (2004), Nr. 8, S. 1505 – 1524
- [Tun09] TUNKELANG, D. ; MARCHIONINI, G. (Hrsg.): *Faceted Search*. San Rafael, CA : Morgan & Claypool Publishers, 2009 (Synthesis Lectures on Information Concepts, Retrieval, and Services 5)
- [TV04] TANGELDER, J. W. H. ; VELTKAMP, R. C.: A Survey of Content Based 3D Shape Retrieval Methods. In: *Proceedings of the International Conference on Shape Modeling and Applications*. Genova, Italy : IEEE Computer Society, Los Alamitos, June 2004, S. 145 – 156
- [Tv08] TANENBAUM, A. S. ; VAN MAARTEN, S.: *Verteilte Systeme: Prinzipien und Paradigmen*. 2., aktualisierte Auflage. München : Pearson Studium, 2008
- [TVD09] TIERNY, J. ; VANDEBORRE, J.-P. ; DAOUDI, M.: Partial 3D Shape Retrieval by Reeb Pattern Unfolding. In: *Computer Graphics forum* 28 (2009), Nr. 1, S. 41 – 55
- [TWT03] TABBONE, S. ; WENDLING, L. ; TOMBRE, K.: Matching of Graphical Symbols in Line-Drawing Images Using Angular Signature Information. In: *International Journal on Document Analysis and Recognition* 6 (2003), S. 115 – 125
- [Ull03] ULLMAN, D. G.: *The Mechanical Design Process*. Third Edition. New York : The McGraw-Hill Companies, Inc., 2003
- [US 96] US PRO (U.S. PRODUCT DATA ASSOCIATION) ; IGES/PDES ORGANIZATION (Hrsg.): *Initial Graphics Exchange Specification IGES 5.3*. Charleston: IGES/PDES Organization, 1996. http://www.uspro.org/documents/IGES5-3_forDownload.pdf. – letzter Zugriff: 20.02.2011
- [Vaj05] VAJNA, S.: Informationsmanagement - Management der produkt- und prozessbezogenen Information in der integrierten Produktentwicklung. In: SCHÄPPI, B. (Hrsg.) ; ANDREASEN, M. M. (Hrsg.) ; KIRCHGEORG, M. (Hrsg.) ; RADERMACHER, F.-J. (Hrsg.): *Handbuch Produktentwicklung*. München, Wien : Carl Hanser Verlag, 2005, Kapitel V 3, S. 419 – 447
- [Vel01] VELTKAMP, R. C.: Shape Matching: Similarity Measures and Algorithms / Utrecht University, The Netherlands. 2001 (UU-CS-2001-03). – Forschungsbericht
- [Ver77] VEREIN DEUTSCHER INGENIEURE: *VDI 2222 Blatt 2: Erstellung und Anwendung von Konstruktionskatalogen*. VDI-Handbuch Konstruktion, Juli 1977. – Beuth Verlag GmbH, Berlin

- [Ver79] VEREIN DEUTSCHER INGENIEURE: *VDI 2217: Datenverarbeitung in der Konstruktion - Begriffserläuterungen*. VDI-Handbuch Konstruktion, März 1979. – Beuth Verlag GmbH, Berlin
- [Ver86] VEREIN DEUTSCHER INGENIEURE: *VDI 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte*. VDI-Handbuch Konstruktion, November 1986. – Beuth Verlag GmbH, Berlin
- [Ver97] VEREIN DEUTSCHER INGENIEURE: *VDI 2222 Blatt 1: Konstruktionsmethodik: Methodisches Entwickeln von Lösungsprinzipien*. VDI-Handbuch Konstruktion, Juni 1997. – Beuth Verlag GmbH, Berlin
- [Ver99] VEREIN DEUTSCHER INGENIEURE: *VDI 2223: Methodisches Entwerfen technischer Produkte*. VDI-Handbuch Konstruktion, März 1999. – Beuth Verlag GmbH, Berlin
- [Ver02] VEREIN DEUTSCHER INGENIEURE: *VDI 2219: Informationsverarbeitung in der Produktentwicklung - Einführung und Wirtschaftlichkeit von EDM/PDM-Systemen*. VDI-Handbuch Konstruktion, November 2002. – Beuth Verlag GmbH, Berlin
- [Ver04] VEREIN DEUTSCHER INGENIEURE: *VDI 2206: Entwicklungsmethodik für mechatronische Systeme*. VDI-Handbuch Konstruktion, Juni 2004. – Beuth Verlag GmbH, Berlin
- [Viv10] VIVISIMO, INC.: *Velocity Platform*. Website. <http://vivisimo.com/technology/velocity-platform.html>. Version: 2010. – letzter Zugriff: 13.04.2010
- [Vog05] VOGEL, F.: *Beschreibende und schließende Statistik: Formeln, Definitionen, Erläuterungen, Stichwörter und Tabellen*. 13., korrigierte und erweiterte Auflage. München : Oldenbourg Wissenschaftsverlag GmbH, 2005
- [Vog06] VOGEL, H.: *Konstruieren mit SolidWorks*. 2. überarbeitete Auflage. München, Wien : Carl Hanser Verlag, 2006
- [Vra04] VRANIC, D. V.: *3D Model Retrieval*, University of Leipzig, Diss., 2004
- [VWSS94] VAJNA, S. ; WEBER, C. ; SCHLINGENSIEPEN, J. ; SCHLOTTMANN, D.: *CAD/CAM für Ingenieure: Hardware - Software - Strategien*. Braunschweig / Wiesbaden : Vieweg & Sohn Verlagsgesellschaft mbH, 1994
- [WA03] WALLACE, K. M. ; AHMED, S.: How Engineering Designers Obtain Information. In: LINDEMANN, U. (Hrsg.): *Human Behaviour in Design: Individuals, Teams, Tools*. München, Berlin : Springer-Verlag, 2003, S. 184 – 194

- [Wal86] WALL, R. A.: *Finding and Using Product Information: From Trade Catalogues to Computer Systems*. Aldershot, Hants : Gower Publishing Company Limited, 1986
- [Web10] WEB 3D CONSORTIUM: *X3D and Related Specifications*. Website. <http://www.web3d.org/x3d/specifications>. Version: 2010. – letzter Zugriff: 01.08.2010
- [Whi09] WHIT, A.: *Magic Quadrant for Information Access Technology*. Gartner RAS Core Research Note G00169927 (Gartner, Inc.). <http://www.gartner.com/technology/media-products/reprints/microsoft/vol17/article2/article2.html>. Version: September 2009. – letzter Zugriff: 12.04.2010
- [Wie08] WIENDAHL, H.-P.: *Betriebsorganisation für Ingenieure*. 6., aktualisierte Auflage. München, Wien : Carl Hanser Verlag, 2008
- [Wik10a] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *Ähnlichkeit*. Online-Artikel. http://de.wikipedia.org/wiki/%C3%84hnlichkeit_%28Philosophie%29. Version: 2010. – letzter Zugriff: 12.05.2010
- [Wik10b] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *Aluminium*. Online-Artikel. <http://de.wikipedia.org/wiki/Aluminium>. Version: 2010. – letzter Zugriff: 10.03.2010
- [Wik10c] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *Gewinde*. Online-Artikel. <http://de.wikipedia.org/wiki/Gewinde>. Version: 2010. – letzter Zugriff: 08.06.2010
- [Wik10d] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *JT (visualization format)*. Online-Artikel. http://en.wikipedia.org/wiki/JT_%28visualization_format%29. Version: 2010. – letzter Zugriff: 02.08.2010
- [Wik10e] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *Werkstoffnummer*. Online-Artikel. <http://de.wikipedia.org/wiki/Werkstoffnummer>. Version: 2010. – letzter Zugriff: 18.06.2010
- [Wik10f] WIKIPEDIA: DIE FREIE ENZYKLOPÄDIE: *ZIP (Dateiformat)*. Online-Artikel. http://de.wikipedia.org/wiki/ZIP_%28Dateiformat%29. Version: 2010. – letzter Zugriff: 20.06.2010
- [Wit10] WITHERTON, P. G.: *Produktspezifikation*. Online-Wirtschaftslexikon. <http://www.wirtschaftslexikon24.net/d/produktspezifikation/produktspezifikation.htm>. Version: 2009-2010. – letzter Zugriff: 19.07.2010

- [WJB09] WANG, H. ; JOHNSON, A. L. ; BRACEWELL, R. H.: Supporting Design Rationale Retrieval for Design Knowledge Re-Use. In: THE DESIGN SOCIETY (Hrsg.): *Proceedings of the 17th International Conference on Engineering Design (ICED)*. Palo Alto, CA, USA, August, 24 - 27 2009
- [Wol10] WOLFF, E.: *Spring 3 - Framework für die Java-Entwicklung*. 3., überarbeitete und erweiterte Auflage. Heidelberg : dpunkt.verlag, 2010
- [Wor10] WORLD INTELLECTUAL PROPERTY ORGANIZATION: *International Patent Classification (IPC)*. WIPO-Homepage. <http://www.wipo.int/classifications/ipc/en/index.html>. Version: 2010. – letzter Zugriff: 26.02.2010
- [WPS04] WEBER, C. ; POHL, M. ; STEINBACH, M.: New Ideas for Knowledge Management in Product Development Projects. In: MARJANOVIC, D. (Hrsg.): *Proceedings of the 8th International Design Conference*. Dubrovnik, Croatia, May, 18 - 21 2004, S. 77 – 82
- [WR09] WHITE, R. W. ; ROTH, R. A. ; MARCHIONINI, G. (Hrsg.): *Exploratory Search: Beyond the Query-Response Paradigm*. San Rafael, CA : Morgan & Claypool Publishers, 2009 (Synthesis Lectures on Information Concepts, Retrieval, and Services 3)
- [WSB98] WEBER, R. ; SCHEK, H.-J. ; BLOTT, S.: A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In: GUPTA, A. (Hrsg.) ; SHMUELI, O. (Hrsg.) ; WIDOM, J. (Hrsg.): *Proceedings of the 24th International Conference on Very Large Data Bases (VLDB)*. San Francisco : Morgan Kaufmann, 1998, S. 194 – 205
- [YSLH03] YEE, K.-P. ; SWEARINGEN, K. ; LI, K. ; HEARST, M.: Faceted Metadata for Image Search and Browsing. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. Ft. Lauderdale, Florida, USA : ACM Press, April, 5 - 10 2003, S. 401 – 408
- [Yu02] YU, C.: *High-Dimensional Indexing: Transformational Approaches to High-Dimensional Range and Similarity Searches*. Berlin, Heidelberg : Springer-Verlag, 2002 (LNCS 2341)
- [ZADB06] ZEZULA, P. ; AMATO, G. ; DOHNAL, V. ; BATKO, M.: *Advances in Database Systems*. Bd. 32: *Similarity Search: The Metric Space Approach*. New York : Springer Science + Business Media, 2006
- [Zlo76] ZLOOF, M. M.: Query-by-Example: Operations on Hierarchical Data Bases. In: *AFIPS National Computer Conference*, 1976, S. 845 – 853

Abkürzungsverzeichnis

2D	zweidimensional
2D-PIR	2D Projection Interval Relationship
3D	dreidimensional
ANSI	American National Standards Institute
AP	Anwendungsprotokoll
ASCII	American Standard Code for Information Interchange
AT	Aufgabenträger
AwS	Anwendungssystem
B	Browsing ($\hat{=}$ Monohierarchisches Navigieren)
BRep	Boundary Representation
CAD	Computer Aided Design
CAE	Computer Aided Engineering
CAM	Computer Aided Manufacturing
CAP	Computer Aided Planning
CAQ	Computer Aided Quality
Cx	Computer Aided x
CBR	Case-Based Reasoning
CIM	Computer Integrated Manufacturing
CMS	Content Management System
CRM	Customer Relationship Management
Cos	Kosinusmaß
CSG	Constructive Solid Geometry

CSV	Comma Separated Values
D	Durchmesser
DB	Datenbank
DBS	Datenbanksystem
DBMS	Datenbankmanagementsystem
DB-FV	Depth Buffer Feature Vector
DIN	Deutsches Institut für Normung e.V.
DFT	Diskrete Fouriertransformation
DfX	Design for X
DMS	Dokumentenmanagementsystem
DMU	Digital Mock-Up
DOC/DOCX	Document-Format von MS Word (DOCX ab Version 2007)
DR	Design Rationales
DTD	Document Type Definition
DXF	Drawing Interchange Format
EDM	Engineering Data Management
ERP	Enterprise Resource Planning
ES	Enterprise Search
ESS	Enterprise Search-System
FEM	Finite Elemente Methode
FN	Facettierte Navigation ($\hat{=}$ Polyhierarchisches Navigieren)
FS	Faktensuche
FV	Featurevektor
GATE	General Architecture for Text Engineering
GT	Group Technology
GUI	Graphical User Interface
HTML	Hypertext Markup Language

Hy-FV	Hybrid Feature Vector
ID	Identifikationsnummer
IGES	Initial Graphics Exchange Specification
IKT	Informations- und Kommunikationstechnologien
IM	Informationsmanagement
IPC	International Patent Classification
IR	Information Retrieval
IRS	Information Retrieval-System
IS	Informationssystem
ISO	International Organization for Standardization
IT	Informationstechnik
IWM	Informations- und Wissensmanagement
JDK	Java Development Kit
JNI	Java Native Interface
JT	Jupiter Tessellation
KMU	kleine und mittlere Unternehmen
KOS	Knowledge Organization System
L	Länge
L1	Manhattan-Distanz
L2	Euklidische Distanz
LFRP	Multi- L ayer F aceted Search with R anking using P arallel Coordinates
Lmax	Max-Distanzfunktion
LoD	Level of Detail
MAP	Mean Average Precision
MPS	Metadaten- bzw. Parametersuche
NC	Numerical Control
NLP	Natural Language Processing

OCR	Optical Character Recognition
PDF	Portable Document Format
PDM	Produktdatenmanagement
PDMS	Produktdatenmanagementsystem
PE	Produktentwicklung
PEP	Produktentwicklungsprozess
PLIB	Parts Library (ISO-Standard 13584)
PMS	Projektmanagementsystem
POJO	Plain Old Java Object
QbE	Query-by-Example
QbT	Query-by-Text
QFD	Quadratic Form Distance
RSH-FV	Ray-based Feature Vector with Spherical Harmonics
RTF	Rich Text Format
SD	Shape Distributions
SH	Shape Histogram
SIL-FV	Silhouette-based Feature Vector
SML	Sachmerkmaliste
SQL	Structured Query Language
SRM	Supplier Relationship Management
STEP	Standard for the Exchange of Product Model Data
SVG	Scalable Vector Graphics
UGS	Unigraphics Solutions Inc.
UML	Unified Modeling Language
URI	Uniform Resource Identifier
UUID	Universally Unique Identifier
VDI	Verein Deutscher Ingenieure

VM	Vorgehensmodell
VR	Virtual Reality
VRML	Virtual Reality Modeling Language
WM	Wissensmanagement
WfMS	Workflowmanagementsystem
WWW	World Wide Web
XLS/XLSX	Extensible Spreadsheet-Format von MS Excel (XLSX ab Version 2007)
XML	Extensible Markup Language

Stichwortverzeichnis

Ähnlichkeit	66	Artefakt	12, 65, 75, 79,
partielle	237	99, 106, 109, 118, 123, 159, 162,	
Ähnlichkeitsfacette	71, 78,	188, 194, 199, 209, 210, 213, 215,	
99, 111, 138, 162, 166, 170, 188,		218, 219, 221, 222, 231, 235	
195, 203, 209, 210, 217, 239		Artefaktbeziehungen ..	65, 99–100, 102,
Ähnlichkeitsfunktion/-maß ...	67, 165,	107	
175, 177, 178, 202, 204		Artefaktquellen	100–103
Ähnlichkeitskonzepte		Artefakttyp	12, 55, 63, 65, 77, 79,
graphbasierte	170	87, 97, 106, 118, 137, 216, 224,	
histogrammbasierte	170	225, 235	
Rastergraphik-basierte	200	Artefakttypenebene	65, 70, 100, 116
vektorbasierte	169, 174	Artefakttyphierarchie	79, 110, 114, 215
Vektorgraphik-basierte	201	Attributfacette	71, 78, 80,
Ähnlichkeitssuche ..	36, 38, 43, 54, 59,	99, 111, 119, 124, 135, 160, 162,	
63, 66–68, 71		166, 182, 186, 195, 209, 210, 214,	
fertigungstechnische	49	217	
geometrische	48, 225	Aufgabenklärung	125
2D-Modell	170, 203	Aufgabenträger	
3D-Modell	42, 142, 167, 170, 178	maschinelle	18, 69
3D-Ähnlichkeitssuche	167	personelle	18, 20, 69
ADK-Strukturmodell	68	Ausarbeitungsphase	128, 188
Aktualisierungsszenario	221, 222	Austauschformate	147–152, 191
Allgemeines Dokument	91	AutoCAD	144
Anforderungen ..	57–60, 132, 137, 216,	Autodesk Inventor	144
231		Balkendiagramm	73
funktionale	57	Baugruppe	81, 162, 189, 206
nicht-funktionale	59	Baukastenstückliste	206
Anforderungsliste	131	Berechnung	181
Anfrage	13, 19, 20, 71, 171, 238	Berechnungsergebnis	180
Anfrageformulierung	59, 60, 62	Informationsgehalt	181–185
Anfragevorschau	61, 71, 217	Berechnungsvorschrift	85, 112
Anpassungskonstruktion	8	Bewegungsdaten	40
Architektur	68–70		

Beziehungsfacette .. 78, 100, 116, 119, 124, 136, 160, 162, 186, 194, 209, 210, 217	Design for X-Kriterien 11
Beziehungsnetz 100, 116, 210	Design for X-Strategie 98
Bezugsdaten 208	Design Rationales 11, 50
Boundary Representation 143, 166	Digital Mock-Up 143, 181
Browsing .. 28, 29, 39, 42, 47, 51, 53, 54	Digitales Universum 1
CAD 141	Diskriminator 80, 81, 96, 116
CAD-IR-Systeme 42–44	Distanzfunktion/-maß 67, 175, 200
CAD-Modell 42, 85, 128, 141, 166, 181, 238	Dokument 37, 85, 87– 93, 101, 123, 124, 135, 159, 166, 186, 194, 209, 210, 214, 225, 233, 235
Informationsgehalt 151–157	semi-strukturiert 88
CAD-Systeme 141–147, 189, 206	strukturiert 88, 130, 182, 206
der 1. Generation 142	unstrukturiert 88, 182
der 2. Generation 142	Dokumentenmanagementsysteme .. 37– 39
der 3. Generation 143	Dokumentinformationen 101
CAE 180	Dokumentinhalt 51
CATIA 145	Dokumentkategorie 90
Constructive Solid Geometry 143	Dokumenttyp .. 88, 123–125, 194, 206, 226, 238
Containerdatei .. 77, 103–107, 109, 123, 124	DXF-Format 191–192, 199, 203
Dateiformat 125	dynamische Facettenbereitstellung 115, 216
Daten 16	Ebenenkonzept 65, 100, 235
Aktualität 219	eClass-Standard 32–33
organisatorische 106, 119, 129, 135, 157, 185, 192, 208	Eigenteil 83
semi-strukturierte 46, 60	Einzelteil 81, 188
strukturierte 46, 60, 103	Einzelteilzeichnung 188
tesselierte 166	Engineering Data Management 41
triangulierte 153, 166	Enterprise Search-Systeme 44–47
unstrukturierte ... 46, 60, 103, 196	Entwicklungsinformationen 10
Datenbank 33, 42, 101	Entwurfphase 128, 141, 181, 188
objektorientierte 34, 217	Ergebnispräsentation .. 71–74, 165, 171
objektrelationale 34	ERP-Systeme 39–40, 101, 102, 206
relationale 34, 38, 40, 53	Evaluation 171, 176, 178, 200, 203
Datenbankmanagementsystem 33	Explosionszeichnung 189
Datenbankschema 34	Extraktorkomponente .. 134, 166, 187, 196, 199, 210, 234, 235
Datenbanksysteme 33–36	Facette 29, 60, 71, 80, 107, 109, 110, 124, 134, 137,
Datenimport 75, 79, 215	
Datenmodell 34, 150	
Depth Buffer FV 175	
Design for X-Aspekte 84, 98	

- 165, 188, 199, 213, 215, 216, 221, 222, 224, 235
- einwertige 62
- fakultative 116, 214, 215
- flache 62
- hierarchische 62
- mehrwertige 62
- Merkmale 61
- metrische 61
- nominale 61
- obligatorische 116, 214, 215
- ordinale 61
- Wertigkeit 62, 84, 111, 216
- Facettenausprägung 61, 107, 213, 218, 221, 222
- Facettenklassifikation 29
- Facettenmodell 153
- Facettenschema 78, 110–119, 137, 214, 216, 218, 224
- Facettentyp 61
- Facettierte Navigation . 29, 47, 54, 60, 63
- Faktensuche 36, 40, 54
- Feature-Bäume 161
- Features 67, 168
 - bildbasierte 169
 - geometriebasierte 169
- Featurevektor 169, 200, 202
- Fertigungsähnlichkeit 49
- Finite Elemente Methode 181
- Flächenmodell 142
- Folgedaten 208
- Form-Features 49, 143, 161
- Fremdteil 84
- Geometrieinformationen 153
- Geometriesuchmaschine 43
- Gesamtzeichnung 188
- Gestaltungsrichtlinien 11
- Hauptkomponentenanalyse 168
- Histogramm 170
- Hybrid FV 175
- IGES 147, 155, 159
- Index . 70, 136, 213, 214, 216, 219, 222
 - invertierter 218
- Indexer 118
- Indexierung 19, 87, 115, 213, 224
- Indexierungsframework ... 3, 103, 205, 211, 224, 225, 231, 233, 235, 236
- Indexierungskomponente .. 69, 75, 171
- Indexierungsmethode
 - feine 211
 - grobe 211
- Indexierungsmodus 76
- Indexierungsprozess ... 57, 75–78, 100, 235
- Indexierungstiefe 60
- Indexstrukturen ... 217–218, 233, 239
- Indexverwaltung 78, 216
- industrielle Anwendungssysteme ... 33
- Information 16–17
 - geometrische 192
 - technologische 192
- Information Retrieval 18–20
- Information Retrieval-System 3, 19, 68
- Informationsüberflutung 63
- Informationsbedürfnis 12, 71, 162
- Informationsbedarf 48, 54
 - geäußerter 13
 - objektiver 8–12, 79, 235
 - subjektiver 12–13, 235
- Informationsexplosion 1
- Informationsextraktion 139
- Informationsgehalt 123, 188
- Informationskomplex 12
- Informationsmanagement 16–18
- Informationsverlust 225
- Informationsvielfalt
 - Klassifizierung 11
- interaktives Retrievalmodell 3, 57, 235
- JT 150, 156, 158
- Kantenmodell 142, 147
- Klassifikation
 - analytisch-synthetische 29
 - analytische 29

Klassifikationssysteme	28–215	Nutzen	230
im engeren Sinne	26	NX	146
im weiteren Sinne	26	Ontologie	53
Komplexität	13, 87, 224, 225, 236	Opitz-Klassifikationssystem	31
Komponentenauflösung	209	Ordnungsschemata	26–28
Komponentenverwendung	209	Parallele Koordinaten ..	64–65, 71, 235
Konsistenz	213, 219	Patentklassifikation	30
Konsistenzprüfung ..	78, 136, 210, 213, 214, 219, 224	Pattern Matching ..	137, 139, 154, 182, 188, 197
Konstruktionsarten	8, 98	PDM-Systeme ...	41–42, 100, 102, 206
Konstruktionskataloge	26	Person	95–97, 135, 186, 195
Konzipierungsphase	127, 141	Personendaten	102
kritische Erfolgsfaktoren	230, 239	Pflichtenheft	128, 129, 185
Kunde	96	Informationsgehalt	131
Lastenheft	127, 128	Physical Mock-Up	143
Informationsgehalt	129–131	Polygonerkennung	198, 204
Lastfälle	184	Polyhierarchie	29, 60
Layertechnik	192, 198	Polyrepräsentation ..	78, 109, 118, 124, 165, 194, 199, 209, 210, 213, 214, 216, 218
Leitfaden	226	Präferenzfunktion	66, 71
Level of Detail	167	Precision	19, 174
LFRP-Framework ..	68, 70, 74, 89, 103, 157, 213, 217, 235	Pro/ENGINEER	145
Lieferant	96	Problemlösungsprozess	7
Management	16	Produkt	2, 48, 81–87, 127, 136, 141, 162, 166, 180, 186, 188, 195, 203, 206, 207, 209
Mean Average Precision	174	Produktdaten	100
Mechatronik	4	Produktdatenmanagement	41
Mengenstückliste	206	Produktdatenmodell	41
Metadaten	28, 38, 41, 42, 63, 75, 89, 101, 118, 123, 124, 153, 155, 159, 199	Produktentstehungsprozess	5
Metadaten- bzw. Parametersuche ..	28, 35, 38, 40, 42, 47, 54	Produktentwickler	10, 235
Metadatenstandard	89	Produktentwicklung	2, 235
Middleware	77	Herausforderungen	4–7
Modell	141	kundenanonyme	127
Monohierarchie	29, 31, 81	kundenindividuelle	127
Name/Wert-Paare ..	103, 107, 118, 154, 157, 160, 218	Suchsituationen	13–15
Natural Language Processing ..	139, 188	Produktentwicklungsprozess	5
Neukonstruktion	8	Produktfunktion	84, 93, 112
Normalisierung	167, 170	Produktgeometrie	166
Nullwerte	213, 215	Produktgestalt	153, 167, 193
		Produktgranularität	81

- Produktkomplexität 4
- Produktmodell . 91, 100, 102, 141, 214
- Produktstruktur 83, 153, 162, 206, 208
- Projekt 97–98
- Projektdokument 91
- Projektinformationen 102
- proprietäre Datenformate 152, 190
- Prozesskomplexität 5
- Pullmodus 76
- Pushmodus 77
- Query-by-Example 36, 55, 67
- Query-by-Text 38, 55, 67
- Ranked Retrieval 19, 66
- Ranking 19, 43, 66–68, 235
- Rastergraphik 200
- Ray-based FV with Spherical Harmonics 175
- Recall 19, 174
- Rechteprüfung 74
- Reflexion
 - ergebnisorientierte 230–234
 - kritische 225
 - prozessorientierte 225–226
- Relevanz 19
- Repräsentation 19, 43, 48, 67, 78, 118, 157, 159, 161, 162, 166, 178, 181, 185, 194, 209, 210, 224
- Repräsentationserstellung .73, 77, 109, 169, 171, 210, 213, 226
- Retrieval 19
- Retrievalprozess .. 57, 70–74, 109, 159, 216
- Sachmerkmale 26
- Sachmerkmalleisten 26
- Schlagwortsuche 38, 63, 138
- Schriftfeld 192, 197, 207
- Set Retrieval 19, 61
- Shape Distributions 178, 203
- Shape Histogram 176
- Silhouette-based FV 175
- Simulation 181
- Simultaneous Engineering 4
- Softwarebibliothek 157, 187, 195
- Solid Edge 146
- SolidWorks 145
- spektrale Graphentheorie 202
- SQL 36
- Stückliste 128, 206
 - Informationsgehalt 207–208
- Stücklistenfeld 207, 209
- Stammdaten 40
- STEP 148, 155, 158
- Sterndiagramm 72
- Strukturstückliste 206, 208, 210
- Suche
 - explorative 28, 29, 54, 60
 - facettierte ... 60–64, 215, 218, 235
 - inhaltsbasierte 200, 201
 - multikriterielle 55, 59
 - zielgerichtete 54, 60
- Suchkomponente 69, 74, 219
- Suchkriterien 59, 60, 79, 83, 84, 89, 94, 109, 188, 234, 235
- Suchparadigmen 20, 25, 47, 54–56
- Synset 112, 197
- System
 - erzeugendes 5
 - prozessbegleitendes 5
 - technisches 2, 81
- Taxonomie 52
- Technische Zeichnung ... 85, 128, 188, 189, 199, 203, 238
 - Informationsgehalt 192–194
- Tesselierung 153
- Testbericht 180
 - Informationsgehalt 185–188
- Thumbnail 73
- Thumbnailfacette ... 78, 165, 166, 171, 195, 210, 217
- Triangulation 153
- Unternehmensmitarbeiter 97
- Validierung 216, 219, 224
- Variantenkonstruktion 8
- Varietät 236

Verarbeitungspipeline	139
Verifikation	219, 224
Volltextsuche	38, 42, 47, 51, 60
Volumenmodell	143, 147
Vorgehensmodell	9, 125, 226
VRML	151, 153, 158
Werkstoff	85, 93–95, 157, 195, 210
Werkstoffdaten	101
Werkstoffeigenschaften	94
Wiederholteile	31
Wissen	16–17
explizites	17
implizites	17
Wissenskreislauf	18, 20
Wissensmanagement	16–18
Zeichnungsfläche	193



Durch den zunehmenden Einsatz von Informations- und Kommunikationstechnologien sowie den schnellen Technologiefortschritt steht die Entwicklung technischer Produkte vor immer neuen Herausforderungen. Dabei ist die Aufgabe der Produktentwicklung selbst als Problemlösungsprozess zu betrachten, in dem Lösungen mittels intensiver Informationsverarbeitung gefunden werden. Somit werden täglich unterschiedlichste Arten von Informationen erstellt, benötigt und verarbeitet, die primär in digitaler Form vorliegen. Diese werden in heterogenen Anwendungssystemen verwaltet, was eine Wiederverwendung bereits existierender Informationen erschwert. Damit beansprucht die Suche nach Informationen noch immer einen erheblichen Anteil der Entwicklungszeit.

Zur Verbesserung der Informationsversorgung im Bereich der technischen Produktentwicklung wird ein interaktives Information Retrieval-System – das LFRP-Framework – vorgestellt. Dieses kombiniert die vier Basiskonzepte der multiplen Ebenen, der facettierten Suche, des Rankings und der parallelen Koordinaten, um hochkomplexe Informationsbedürfnisse zu befriedigen. Seine Realisierung erfordert neben einer geeigneten Suchoberfläche die Entwicklung einer Indexierungskomponente, welche die vorhandenen Informationen in eine für das LFRP-Framework rechnerverarbeitbare Form transformiert. Dieser als Indexierung bezeichnete Prozess stellt die Grundvoraussetzung für die Funktionsfähigkeit eines Suchsystems dar und liegt daher im Fokus der Betrachtung. Es wird ein Lösungsansatz vorgestellt, welcher eine Indexierung in Form facettenbasierter Suchkriterien ermöglicht und dabei nicht nur Informationen aus heterogenen Anwendungssystemen, sondern insbesondere aus entwicklungsspezifischen Dokumenten, wie CAD-Modellen, technischen Zeichnungen oder Stücklisten, berücksichtigt.

ISBN 978-3-86309-058-6

ISSN 1867-7401

28,00 Euro