BAMBERGER BEITRÄGE
ZUR WIRTSCHAFTSINFORMATIK UND ANGEWANDTEN INFORMATIK
ISSN 0937-3349

Nr. 84

A Process for Identifying Predictive Correlation Patterns in Service Management Systems

Werner Zirkel and Guido Wirtz

February 2010

FAKULTÄT WIRTSCHAFTSINFORMATIK UND ANGEWANDTE INFORMATIK
OTTO-FRIEDRICH-UNIVERSITÄT BAMBERG

Distributed Systems Group

Otto-Friedrich Universität Bamberg

Feldkirchenstr. 21, 96052 Bamberg, GERMANY

Prof. Dr. rer. nat. Guido Wirtz

`http://www.uni-bamberg.de/pi/`

Due to hardware developments, strong application needs and the overwhelming influence of the net in almost all areas, distributed systems have become one of the most important topics for nowadays software industry. Owing to their ever increasing importance for everyday business, distributed systems have high requirements with respect to dependability, robustness and performance. Unfortunately, distribution adds its share to the problems of developing complex software systems. Heterogeneity in both, hardware and software, permanent changes, concurrency, distribution of components and the need for inter-operability between different systems complicate matters. Moreover, new technical aspects like resource management, load balancing and guaranteeing consistent operation in the presence of partial failures and deadlocks put an additional burden onto the developer.
The long-term common goal of our research efforts is the development, implementation and evaluation of methods helpful for the realization of robust and easy-to-use software for complex systems in general while putting a focus on the problems and issues regarding distributed systems on all levels. Our current research activities are focussed on different aspects centered around that theme:

- *Reliable and inter-operable Service-oriented Architectures*: Development of design methods, languages, tools and middle-ware to ease the development of SOAs with an emphasis on provable correct systems that allow for early design-evaluation due to rigorous development methods. Additionally, we work on approaches and standards to provide truly inter-operable platforms for SOAs.
- *Implementation of Business Processes and Business-to-Business-Integration (B2Bi)*: Starting from requirements for successful B2Bi development processes, languages and systems, we investigate the practicability and inter-operability of different approaches and platforms for the design and implementation of business processes with a focus on combining processes from different business partners.
- *Quality-of-Service (QoS) Aspects for SOA and B2Bi*: QoS aspects, especially reliability and security, are indispensable when putting distributed systems into practical use. We work on methods that allow for a seamless observance of QoS aspects during the entire development process from high-level business processes down to implementation platforms.
- *Agent and Multi-Agent (MAS) Technology:* Development of new approaches to use Multi-Agent-Systems for designing, organizing and optimizing complex systems ranging from service management and SOA to electronic markets and virtual enterprises.
- *Visual Programming- and Design-Languages*: The goal of this long-term effort is the utilization of visual metaphors and languages as well as visualization techniques to make design- and programming languages more understandable and, hence, more easy-to-use.

More information about our work, i.e., projects, papers and software, is available at our homepage (see above). If you have any questions or suggestions regarding this report or our work in general, don't hesitate to contact me at `guido.wirtz@uni-bamberg.de`

Guido Wirtz

Bamberg, January 2010

A Process for Identifying Predictive Correlation Patterns in Service Management Systems

Werner Zirkel and Guido Wirtz

Abstract By using the remote functions of a modern IT service management system infrastructure, it is possible to analyze huge amounts of logfile data from complex technical equipment. This enables a service provider to predict failures of connected equipment before they happen. The problem most providers face in this context is finding 'a needle in a haystack' - the obtained amount of data turns out to be too large to be analyzed manually. This report describes a process to find suitable predictive patterns in log files for the detection of upcoming critical situations. The identification process may serve as a hands-on guide. It describes how to connect statistical means, data mining algorithms and expert domain knowledge in the domain of service management. The process was developed in a research project which is currently being carried out within the Siemens Healthcare service organization. The project deals with two main aspects: First, the identification of predictive patterns in existing service data and second, the architecture of an autonomous agent which is able to correlate such patterns. This paper summarizes the results of the first project challenge. The identification process was tested successfully in a proof of concept for several Siemens Healthcare products.

Keywords Service Management, Event Correlation, Pattern Identification

Contact:

Werner Zirkel
Siemens AG
Healthcare IT
Henkestr. 127
91052 Erlangen
Germany
werner.zirkel@siemens.com

Guido Wirtz
Distributed Systems Group
University of Bamberg
Feldkirchenstrasse 21
Bamberg 96045
Germany
guido.wirtz@uni-bamberg.de

Contents

1 Service Management Fundamentals

From a customer's point of view, a service management system offers a portfolio of services which are designed to help the customer in achieving her business goals. By doing so, a service management system creates added value for a customer. To understand state-of-the-art service management, it is helpful review the historic progress of service-related systems. Many of them are rooted in network management systems, which were used to control a set of technical equipment remotely. Due to the increasing complexity of computer networks, the task has shifted to the more abstract level of managing whole systems, which focus on complex objects, including hardware and software [5]. Modern systems may be designed to support entire business processes. Therefore they need an appropriate IT infrastructure as well as a process landscape which acts as an enabler to fulfill the customer's business goals.

Many companies commonly face two important challenges - increasing the customer's service organizations uptime and reducing internal service costs. Thus, the service providers may implement an event management process to help to achieve these goals. The event management activities have changed over time with an increase in the abstraction level of service management systems, from polling simple status surveillance information to analyzing complex workflow data [3]: This is done by using intelligent agents to retrieve complex information about the status and the workload, thereby enabling service providers to ensure the continual uptime of a service object. Because of this, the service management often uses a distributed software architecture [10] where software agents are used to read the service data and correlate the information to provide a service operator with meaningful information.

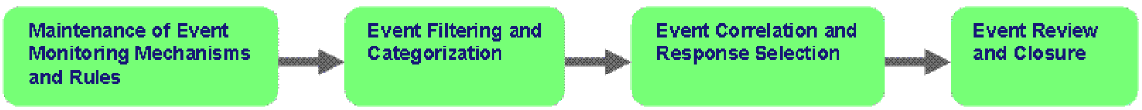

Figure 1: ITIL v3 Event Management Process

The Information Technology Infrastructure Library (ITIL) was developed by the Office of Government Commerce (OGC), and for many service-related organizations, it serves as a reference framework for the implementation of such service processes. As a best practice approach, ITIL also recommends an event management reference process (c.f. Fig.1). This process definition can be subdivided into the four basic activities: *maintenance of event monitoring mechanisms and rules, event filtering and categorization, event correlation and response selection and event review and closure* [16].

ITIL defines an event in this context as 'an alert or notification created by any IT service, Configuration Item or monitoring tool. Events typically require IT operations personnel to take actions, and often lead to Incidents being logged' [16]. ITIL proposes three categories for event classification:

- *Information events* - appear commonly and might become important in a certain context
- *Warning events* - indicate a precedent critical situation
- *Exception events* - indicate that a certain threshold has been exceed

Considering this definition, it is possible to separate two similar research topics which are both covered by the ITIL event management definition. While *event management* in a closer sense works only with exceptions, *property monitoring*, which is sometimes referred to as condition-based monitoring, works also with continuous low-level information events (e.g. machine values). This differentiation is essential regarding the service data characteristics. Sensor values such as temperature, voltage or speed values are generally collected continually, so many property monitoring solutions are able to use regression methods to predict failure situations. Event management systems have to correlate events which might appear sporadically, not continually. Therefore, these systems have to fall back on more robust methods for prediction.

By identifying a set of mechanisms and rules, events may be classified and filtered. In this paper, we assume that the idea of the activity *event correlation* is rooted in the experience that an event may not be uniquely interpretable for a service operator and should be regarded in the context of its preceding and successive events. A correlation consists therefore of a set of events which has a certain structure. This reproducible sequence forms a *pattern*. With these patterns, failure situations may be identified[1]. By analyzing the workflow data of service objects in this way, failures should be found as soon as possible and appropriate actions can be planned in time. A better planning strategy may help minimize working time costs as well as material logistics costs on the service provider side.

ITIL proposes a final event review to keep the event management process transparent and controllable. From an organizational point of view, it is possible to differentiate the event management quality in consideration of the *reaction time*.

In the worst case, an alarm message is generated after the customer has perceived the problem. In this case the customer would call the service provider first. Thus, this behavior can described as *reactive* service. If the service provider is alarmed directly before or at the same time as the failure is noticed by the customer, it is possible to start emergency actions, e.g. to pre-clarify the situation or to contact the customer first. This reaction can be considered as *proactive* behavior, because it allows the service organization to start emergency counter-measures such as pre-clarification and/or emergency orders before the customer acts. One major disadvantage of proactive service is the lack of time to handle the failure situation as a standard operative action, which implies that alternative emergency mechanisms have to be used.

From an economic standpoint, correlations that are able to predict future system failures are most valuable, because the customers' uptime is maximized and previously unplanned actions may be planned in future. In this sense, it is reasonable to separate proactive behavior from *predictive* behavior. A pattern can be labeled *predictive*, if it enables its service organization to pre-plan upcoming actions and/or to set up appropriate counter-measures. Figure 2 shows a time-line illustrating the sequence of predictive, proactive and reactive behavior. Note that the reaction time depends on the point of failure perception as well as the event management process implementation. The point of failure perception can be influenced by the quality of

[1] This paper focuses on failure situations which occur repeatedly and are therefore predictable by analyzing historical data. It does not consider situations which appear only once.

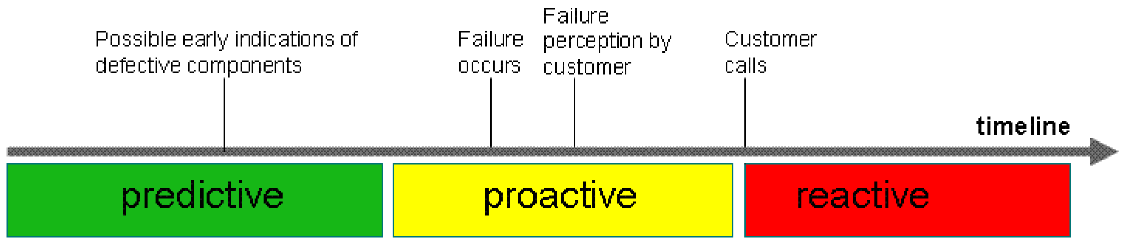

Figure 2: Behavior of the Service Organization

patterns; the process refers to the service organization's ability to act immediately. Therefore, it is not possible to define exact time intervals which are generally valid.

2 Statistical Methods for Event Correlation and Failure Prediction

Based on different research disciplines, there exists a huge variety of methods to implement event correlation mechanisms. Some examples: Hanemann and Marcu differentiate hypotheses-based approaches such as the codebook approach, case-based reasoning and active probing. [3] Hanemann also proposes a combination of rule-based mechanisms and case-based reasoning for correlation. Besides these approaches there exist data-driven methods. For example, Wietgrefe et al. use a neural net to implement a correlation system [15]. The definition of event correlation does not basically consider time distances which may occur from the correlation trigger to the actual occurrence of the site problem. Many correlation systems work with patterns and notify the service operator after the problem occurred. From the customer's point of view, this results in reactive behavior, because a necessary action does not start before the failure occurs and usually the customer informs the service provider about it. The main purpose of a reactive system may therefore be seen as support for field service engineers.

In this sense, event correlation systems have to be enhanced with mechanisms for probability forecasting. A lot of scientific approaches deal with this problem. Flatin et al. summarize different approaches in 'Event Correlation in Integrated Management-Lessons Learned and Outlook' [2]. As well as classic statistical methods [9], regression- and classification algorithms are commonly used [7, 14]. In addition, there are already a variety of more specialized algorithms available. For example, Gary Weiss proposes the *Timeweaver* algorithm, which may be suitable for identifying predictive patterns [13]. Salfner describes in his doctoral thesis 'Event-based Failure Prediction - An Extended Hidden Markov Model Approach' how Hidden semi-Markov Models (HSMMs) can be applied to predict system failures. He compares the application with the efficiency of standard classification algorithms like Support Vector Machines (SVM) und Naïve Bayes [17]. Li et al. use the Cox Regression algorithm to identify variables which influence an upcoming hard down situation at most [6]. Taking these factors into consideration, it is also possible to build a pattern which predicts system failures.

3 Developing a Process for Pattern Identification

The goal of this paper is not to find a new algorithm which finds patterns better, but to *define a process to identify proactive and predictive patterns*. This process uses statistical means and data mining algorithms, as well as expert knowledge, which is commonly available in cost-intensive service organizations, and searches for semantically connected data structures. These structures are used as indicators for upcoming equipment problems. The process consists of a set of activities. The diagram in Fig.3 shows how these steps are linked in a sequence. Based on economic considerations and technical feasibility, a use case has to be identified which describes a certain failure situation and its economic consequences.[2] Starting with the analysis of frequency distributions, significance tests and the application of the *Generalized Sequential Pattern (GSP) algorithm*, a basic set of events may be identified. This set may then be structured by domain experts into semantically dependent sequences. The process ends with a final significance test to confirm the pattern's efficiency. The process is basically procedural; though practical application showed that in some cases it might become necessary to enhance the raw/detail pattern by redefining the use case and/or retrieving additional data from the domain experts.

In the following subsections, this process will be explained in detail, starting with some basic information on the economical perspective. Afterwards, we describe how this process was applied in a case study.

3.1 Business Impact and Technical Feasibility

To identify patterns, it is important to consider the connection of cause and effect. Our research experiences proved that single events can not generally be interpreted without context information. Because the actual cause of a system failure is a defect of one or more sub-components, every use case in the process refers directly to a certain material defect and not to a certain system condition. We assume, that the impact of a certain defect always manifests itself in a specific constellation of events. An inverse deduction, namely the conclusion from one or more events to a certain system condition is usually only possible in a few cases, because a specific condition is often caused by different defects.[5] Figure 4 visualizes the difference between the two approaches. Low level approaches (e.g. property monitoring) start usually with the knowledge of which properties may cause a system to fail. This approach may not be generally adapted for event correlation systems, because an event by definition represents aggregated sub-component information. Therefore, an a priori deduction of the connection event set/system error is often not possible.

From an economic view, there are also different ways of approaching the search for patterns. One possibility is a top-down analysis which focuses more on the generation of economically useful patterns rather than on a maximum amount of patterns. Considering spare part defects as the original problem and its consequence, a certain system error, the top-down approach

[2] In the service domain, a use case might be based on a problem which is rooted in a common defect of a certain spare part.

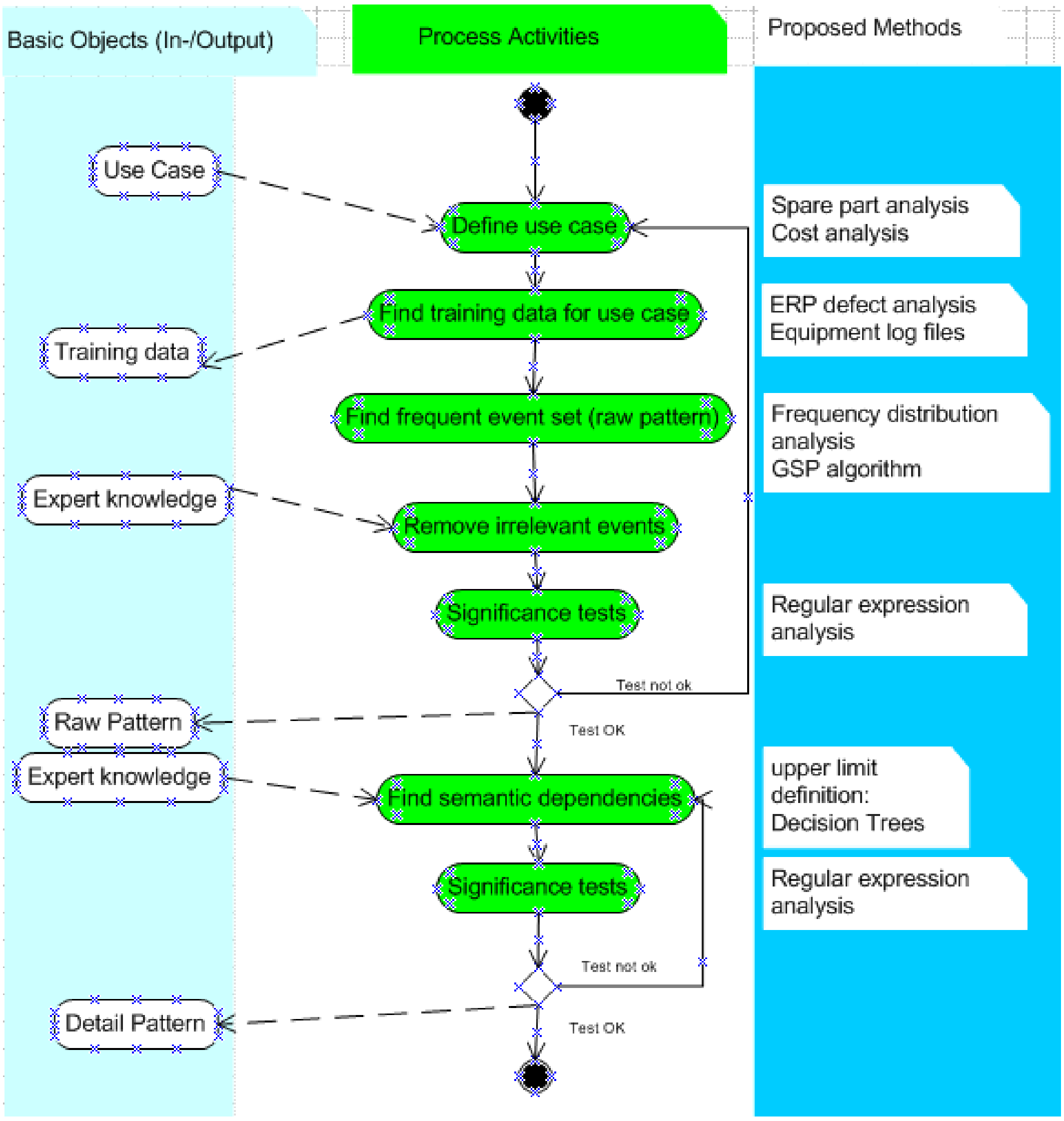

Figure 3: Process Overview

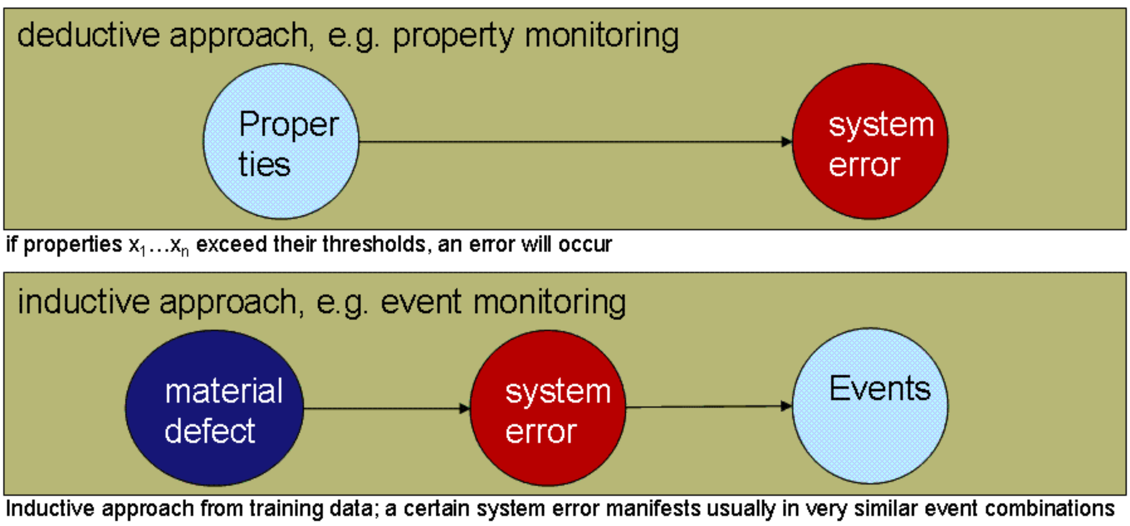

Figure 4: Consideration of Cause and Effect

starts by defining a use case. Therefore, the most frequent root causes of all defective spare parts (defining a *use case*) should be analyzed in order to continue with an economic investigation. Two aspects of service management are often emphasized: to ensure high uptime for customers and to work as cost efficiently as possible. Therefore, it is necessary to analyze the potential uptime increase and cost savings in order to guarantee efficient patterns. The uptime analysis considers the information of how frequently a defect appeared within a certain period of time and which spare parts would cause the most down time:

*Additional uptime potential = Number of defects per period * Mean Time To Restore Service*

A cost analysis depends mainly on the structure of the service organization. Therefore, it is necessary to clarify what are the main cost drivers in the service escalation process. If it is possible to determine the mean service costs of a single service intervention, the potential cost saving may be calculated as:

Potential cost savings[3] *= Number of defects per period * Mean Service Costs per Service Intervention*

These calculations have to be adapted depending on the characteristics of the service organization and its business environment. As a result, a calculation for each pattern shows the cost-saving potential and the possible uptime increase that can be reached for a pattern. This approach alleviates the implementation of a control mechanism according to the ITIL event management process activity *event review and closure* and supports continual service improvement. As soon as a use case is identified as economically valuable, domain experts have to be consulted to ensure that the pattern development is technically feasible. The main aspects required are:

- There has to be reproducible underlying defect, which is not considered as an '*Act of God*'[4].
- There has to be a proper amount of training data which can be used to build training sets and which can be used for significance tests/statistical validation. This also implies the need of a certain level of homogeneity of subcomponents.

3.2 Developing a Raw Pattern

The process is based on the *assumption that certain events are very frequent just before the failure occurs* and that they are measurable in a weaker instance as predictive indicators. To build a pattern, training data has to be acquired which can be examined a posteriori. Thus, in a first step, the frequency distributions of two training sets are compared (c.f. Fig.5). To create those sets it is necessary to define a *window width* w which limits the time span of each training case. The size of w depends on the service data and the use case itself and has to be adapted appropriately.

Set A is a set of log files from different sites where a spare part was replaced. Set B contains log files of sites where no failure occurred. The result of the comparison reveals two important

[3]Considering prevented on-site repair activities

[4]It is not possible to predict a technical problem which is caused by suddenly appearing incidents (e.g. lightning stroke, defective fuse, etc).

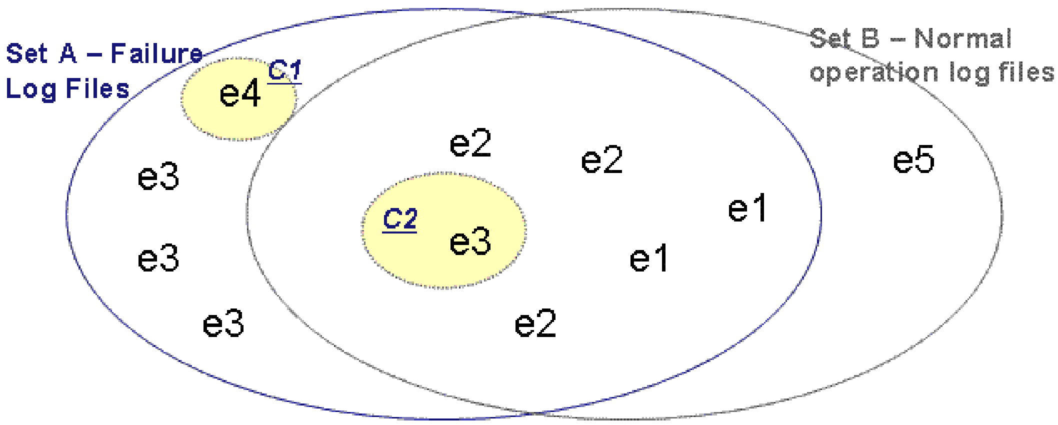

Figure 5: Building a Raw Pattern

subsets: C1, which is a set of events which occur only in set A. Because these events do not appear in B, it is reasonable to assume that these events are in a certain relationship to the failure. Furthermore, the result will contain a set of events C2 which appear very frequent in A compared to B. These events are promising pattern candidates and could be used to develop the detail pattern later on. We call C = C1 + C2 a *raw pattern*. A raw pattern is an unstructured set of events, which seem to be connected to the failure situation.

3.3 Significance Tests

The evaluation of frequency distributions may be limited to a small sample due to software capacity restrictions. To get more details, it is possible to use significance tests for each event of the raw pattern where larger subsets A' and B' can be analyzed. From a physical point of view, events are commonly stored as text strings in a log files. Therefore, the amount of data which is used in set A' and B' for these tests may be larger than in A and B because the data may be analyzed by a standard regular expression search tool.

To get a detail pattern from the raw pattern, it may be helpful to use an additional sequence analysis. This can be done by using the Generalized Sequential Pattern Algorithm (GSP).[11] By applying this algorithm to training set A, possible sequence candidates can be identified quickly. The input set may be focused on raw pattern events. To identify the correct window width of interesting sequences, set A should be analyzed in adjustment with field experts.

3.4 Developing a Detail Pattern

The raw pattern is a set of unstructured events which can be used as a basis for the development of a detail pattern. Here, the events are structured into *semantic blocks*. A semantic block is a semantically connected, structured sequence of events. The dependencies between events from the raw pattern are identified by domain experts and are weighted for pattern sensitization if necessary.

The example in Fig.6 shows the connection of single events to a semantic block: Block 1 consists of all the events e1Ĕe4 and the error events E1 and E3 in the time order e3, e1, e2, E3, e1, e4, E1.

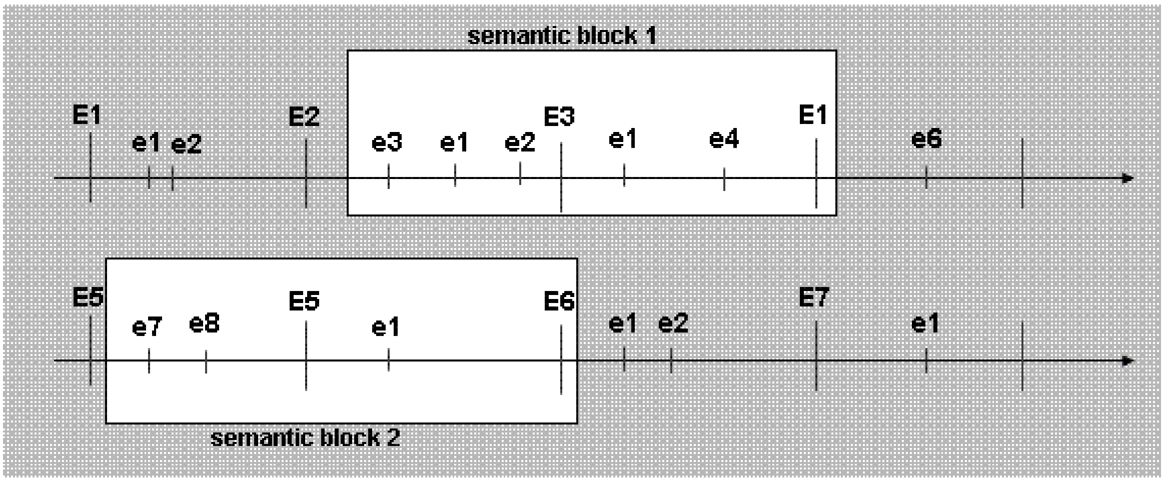

Figure 6: Semantic Blocks – Abstract Example

Semantic blocks occur within a certain time frame in the log file. The semantic blocks may now be counted depending on the chosen window width w. The detail pattern has to be verified by significance tests. A critical site status may be classified by using *Shewhart maps*. Practical experience showed that it is necessary to base the classification on site-specific values, because the semantic block count may not be null in call cases, but fluctuates around a certain *normal operation value*. The reasons are easy to comprehend from a practical standpoint: on the one side, a complex technical site has a variety of subcomponents where construction details may differ to some degree. On the other side, the degree of utilization or the workflow for each service object may vary significantly. Thus, it is necessary to calculate a normal operation value for each pattern and/or semantic block to set an actual count value in relation to the normal operation value. In some cases, it may be necessary to set up independent control charts for each semantic block. Therefore, it might become difficult to define the upper limits in the charts manually - a decision tree model might be useful for the pattern designer to reduce the complexity.[5] For the final significance tests, a regular expression tool may provide the necessary functions to identify chronologically connected semantic blocks.

3.5 Using the Pattern for Event Correlation

The developers of service data log files do usually not focus on machine readable processability. Therefore, it is often not possible to retrieve continuous metrics which can for example be used as a basis for a regression algorithm. For this reason service data mining experts have to find another appropriate metaphor. Regression uses the idea of a 'steadily abrasion': based on historic values, the most probable future values will be predicted. The idea we propose is based on the metaphor of a '*juddering*' system: a single or multiple misbehaviors of a site regarding

[5] For example, this might be the C4.5- or CHAID-Algorithm; the count values of the semantic blocks might be used as input variables.

historic normal operation values, which indicates a future failure situation. These misbehaviors can be statistically interpreted as outliers on a c-chart, where the crossing of a defined upper control limit triggers an action (e.g. sending a warning message to a service Operator).

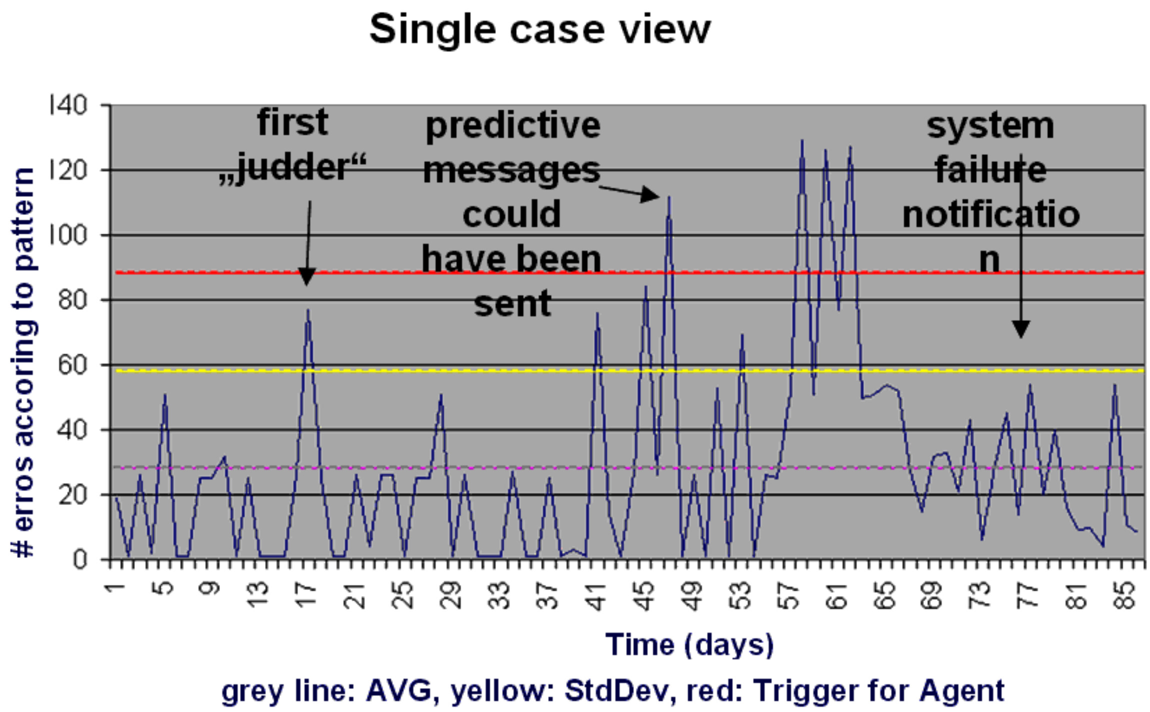

Figure 7: Example: Pattern detection on a single site

The c-chart in Fig.7 shows a site specific example how a predictive pattern classification would look like. The figure shows a pattern that indicates vertical movement problems of the patient table of a computer tomography. The customer called the service provider on day 76 (x-axis). The graph shows a significant spike on day 17 (y-axis). This could be interpreted as a first '*judder*' of a defect motor controller in the table. The failure would have been predictable on day 48 (spike), 28 days before the customer's call. In this example, we used one c-chart for all semantic blocks, whereas the upper threshold for escalation was defined at 2*STDDEV of the last 30 days[6].

4 Case Study

The service object we chose for our example analysis is a computer tomograph. The use case manifests itself, from the customers point of view, as an abort of a spiral scan. The defect itself happens in the data acquisition phase, where information is being transferred from the gantry to the computer. Defective hard disks and/or communication problems are the reasons why the scan raw data cannot be written to the hard disks of the data acquisition computer. As

[6]The shown upper limit (red line) and the warning limit (yellow line) are simplified as straight line here which refers to day 76. Note, that in some cases, more complex models (e.g. ARIMA) might increase the accuracy. See also Miloucheva/Anzaloni/Müller ([9]) for additional information.

a consequence, the scan may be aborted. Unfortunately, we were not able to obtain enough data to build separate patterns for both the disk problem and the communication device, so we had to build one collective pattern for both defects. This combination is possible because both device defects manifest themselves in a very similar way in the log files and are finally causing the scan abort.

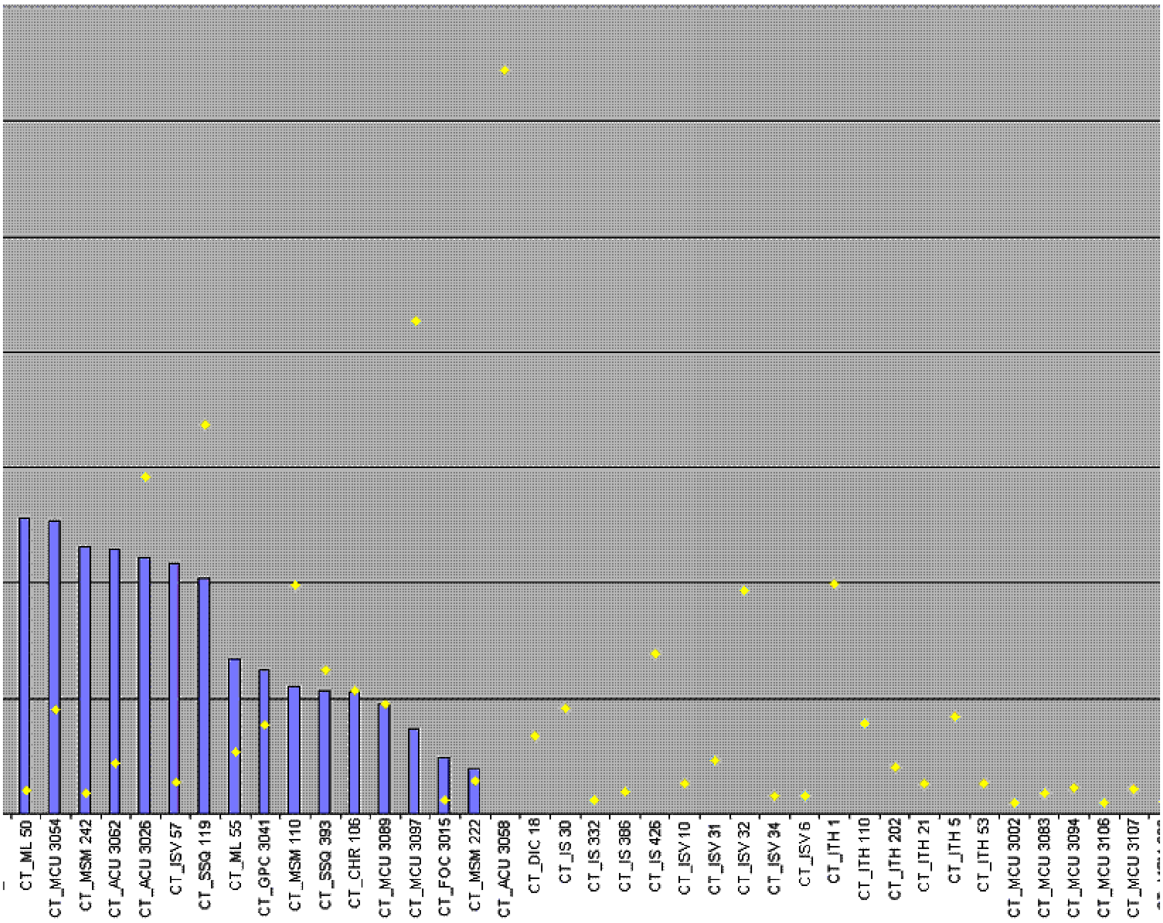

Figure 8: Comparison of Frequency Distributions

The analysis of the error frequency was based on the business data which was obtained from the ERP system. The uptime analysis results proved an average uptime increase of approximately two hours per defect. The cost analysis showed that these defects caused a significant amount of service cost due to multiple on-site repairs, overtime fees and emergency spare part transports. So we calculated a potential profit which motivated the development of an associated pattern.

For research purposes, we collected 52 training cases for set A with continuing log files of 30 days. Domain experts could initially pre-clarify single failure events and assumed uncommon failure distributions in a majority of the training set. To build the raw pattern, the frequency distributions of 13 (=25%) cases in set A have been analyzed. The raw pattern set held 28 single events, which occurred in $\geq$ 12 cases (92 %) very frequently or only in the failure situation. Due to the significance tests of each event, some events could be identified as not relevant and had been rejected. The detail pattern was built with the help of CT experts. The following chart shows an example for a frequency distribution analysis: The events CT_ACU 3058 to CT

_STC69 appeared only in set A and CT_ISV 37 to CT_MSM222 appeared significantly more frequently in the training set A than those messages would appear in a normal operation mode (set B). The chart (blue columns) shows the relative frequencies of each single event of set A divided by set B. For example, CT_ISV 37 appeared in A about 3 times as much often as in B. Events with no column exist only in set A and could not be found in B. The point markers which are dimensioned on the second y-axis additionally show how frequent the set A events occurred in relation to all set A events. CT_ITH 202 is an example for a so-called *’rare event’* [7] – it is obvious that this event seems to be very rare in comparison to all events, but it is frequent regarding set proportion of set A and B.

Support	Transaction 0	Transaction 1	
0.667	CT_ISV 32, CT_ITH 1		1
0.611	CT_ISV 32	CT_DIC 18	2
0.611	CT_DIC 18	CT_DIC 18	2
0.611	CT_ITH 1	CT_DIC 18	2
0.611	CT_ML 50	CT_DIC 18	2
0.611	CT_ISV 32, CT_ITH 1	CT_DIC 18	2
0.611	CT_ISV 32	CT_ISV 32	2
0.611	CT_ITH 1	CT_ISV 32	2

Figure 9: GSP Algorithm Application Example

Figure 9[8] shows the result of a GSP application. 23 training cases were analyzed for frequent item sets using a support factor of 0.6. The GSP window width parameter was set to w = 1 minute, because we were only interested in items that come in a close chronological sequence. In this case it seemed to be reasonable to use only a subset of all events as GSP input due to reduce the calculation time of GSP. The highlighted data set contains different events. The events CT_ISV32 and CT_ITH1 are a part of the raw pattern. We investigated the sequence CT_ISV32/CT_ITH1 more closely, because these messages appear together in the given distance of s = 1[9]. The GSP results had been discussed with a domain expert who confirmed in this case that CT_ISV32 is an error that comes with a nearby ITH1 explanation. Based on the raw pattern events and the GSP results, 20 semantic blocks were identified, which are directly connected to hard disk problems and/or the communication system. The ISV32 event was identified as serious problem of the control computer and was associated with a reset event (CT_SSQ211). Fig.10 shows an excerpt from the original log file where the combination CT_ISV32/SSQ211/ITH1 appears: The significance tests consisted of 39 failure cases (set A, True Positive) and 5974 placebo files (set B, True Negative). In set A, 27 out of 39 cases were identified correctly, where 5966 out of 5974 log files were classified correctly in set B. By rating the true positive result it has to be considered that it was not possible to calculate an exact

[7]*’Rare Events’* for further details see definition from G. Weiss in[8], S.765 and *’Small disjuncts’* in [14], S.1.

[8]X-axis shows event types, Y-axis shows the relative frequencies of each single event of set A divided by set B.

[9]s was set to 1 minute. In this case it seemed to be reasonable to use only a subset of all events as input variables for GSP to minimize the calculation effort

2008-10-28	11:21:09	CT_ITH	1	IRS error information: NET: Connection closed: IRSAQP(19
2008-10-28	11:21:09	CT_SSQ	493	received IS-Notification: SsqCtrlIsClient::onIrsAsyncAbo
2008-10-28	11:21:09	CT_ITH	53	Network error.
2008-10-28	11:21:09	CT SSQ	211	SsqCtrl: Imaging system automatically reset.
2008-10-28	11:21:09	CT ISV	10	Abort message arrived from IRS.
2008-10-28	11:21:09	CT ISV	32	Reset Indicator IRS_RESET_DEMANDED has been detected in

Figure 10: Log file excerpt with semantic block ISV32/SSQ211/ITH1

accuracy rate because no final spare part examination data was available. A sample analysis of three incorrect classifications justifies the assumption, that some spare parts have been replaced although they had no defects.

5 Conclusion and Future Work

Data Mining is often described as an adequate method to find patterns in big, unstructured databases. From a practical point of view, the application of data mining algorithms often fails due to a lack of understanding of how those algorithms can be applied in a concrete situation. The method shown therefore resembles a cooking recipe, which consists of list of ingredients (statistical methods, data mining algorithms) and a description how and when to mix these ingredients (process activities). The process we describe showed its usefulness and applicability in practice - we successfully identified several patterns for different healthcare service objects which are used for computer tomography, magnetic resonance tomography and vascular surgery.

Our future work will focus on an autonomous software agent which is able to correlate those patterns and integrates into commonly existing service management environments. Though we started by analyzing healthcare service equipment, we anticipate that the process shown is not limited to this sector and therefore we are going to analyze its applicability into other service segments as well.

References

[1] Aboelela/Douligeris (1999): Fuzzy Temporal Reasoning Model for Event Correlation in Network Management, IEEE CS, in: Proc. of the 24th Annual IEEE Conf. on Local Computer Networks ,Page:150

[2] Flatin/Jakobson/Lewis (2007): Event Correlation in Integrated Management: Lessons Learned and Outlook, in: Journal of Network and Systems Management ,Volume 15, Issue 2,Page(s):481-502

[3] Hanemann/Marcu (2008): Algorithm design and application of service-oriented event correlation, in: Proc. of the 3rd IFIP/IEEE Intern. Workshop on Business-Driven IT Management,Vol. 7, Issue 7 ,Page(s):61 - 70

[4] Hegering/Abeck/Neumair (1999): Integriertes Management vernetzter Systeme: Konzepte, Architekturen und deren betrieblicher Einsatz, 1st edition, dpunkt, Heidelberg

[5] Kätker/Geihs(1997): A Generic Model for Fault Isolation in Integrated Management Systems, Springer, in: Journal of Network and Systems Management,Volume 5, Number 2,Page(s):109-130

[6] Li/Zhou/Choubey/Sievenpiper (1997): Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures, Taylor and Francis Ltd, in: IEE Transactions, Volume 39, No. 3, Page(s): 303-317

[7] Lindsay/Cox (2005): Effective Probability Forecasting for Time Series Data Using Standard Machine Learning Techniques , Springer, in: Pattern Recognition and Data Mining

[8] Maimon/Rokach (2005): Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers Kluwer Academic Publishers

[9] Miloucheva/Anzaloni/Müller (2003): A practical approach to forecast Quality of service parameters considering outliers, in: Inter-domain performance and simulation Workshop

[10] Pavlou/Mykoniatis/Sanchez (1996): Distributed intelligent monitoring and reporting facilities in: Distributed Systems Engineering, Vol. 3, No. 2, Page(s):124-135

[11] Srikant/Agrawal (1996): Mining Sequential Patterns: Generalizations and Performance Improvements, in: Proceedings of 5th Int. Conf. Extending Database Technology, Volume 1057, Page(s): 3-7

[12] Thottan/Ji (1999): Statistical Detection of Enterprise Network Problems, in: Journal of Network and Systems Management, Volume 7 , Issue 1,Page(s): 27-45

[13] Weiss (1999): Time Weaver: a Genetic Algorithm for Identifying Predictive Patterns in Sequences of Events, Morgan Kaufmann, San Francisco, in: Proc. of the Genetic and Evolutionary Computation Conf.,Page(s):718-725

[14] Weiss (2009): The Impact of Small Disjuncts on Classifier Learning

[15] Wietgrefe/Tuchs/Jobmann/Carls/Fröhlich/Nejdl/Steinfeld (1997): Using Neural Networks for Alarm Correlation in Cellular Phone Networks, in: Proceedings of the International Workshop on Applications of Neural Networks in Telecommunications

[16] IT Process Maps (2009): IT Process Maps - event management; wiki.de.it-processmaps.com/index.php/Event_Management

[17] Salfner (2008): Event-based Failure Prediction-An Extended Hidden Markov Model Approach, Ph.D. Thesis, Computer Architecture and Communication Group, Dept. of CS, Humboldt-University, Berlin

A List of previous University of Bamberg reports

Bamberger Beiträge zur Wirtschaftsinformatik

Nr. 1 (1989)	Augsburger W., Bartmann D., Sinz E.J.: Das Bamberger Modell: Der Diplom-Studiengang Wirtschaftsinformatik an der Universität Bamberg (Nachdruck Dez. 1990)
Nr. 2 (1990)	Esswein W.: Definition, Implementierung und Einsatz einer kompatiblen Datenbankschnittstelle für PROLOG
Nr. 3 (1990)	Augsburger W., Rieder H., Schwab J.: Endbenutzerorientierte Informationsgewinnung aus numerischen Daten am Beispiel von Unternehmenskennzahlen
Nr. 4 (1990)	Ferstl O.K., Sinz E.J.: Objektmodellierung betrieblicher Informationsmodelle im Semantischen Objektmodell (SOM) (Nachdruck Nov. 1990)
Nr. 5 (1990)	Ferstl O.K., Sinz E.J.: Ein Vorgehensmodell zur Objektmodellierung betrieblicher Informationssysteme im Semantischen Objektmodell (SOM)
Nr. 6 (1991)	Augsburger W., Rieder H., Schwab J.: Systemtheoretische Repräsentation von Strukturen und Bewertungsfunktionen über zeitabhängigen betrieblichen numerischen Daten
Nr. 7 (1991)	Augsburger W., Rieder H., Schwab J.: Wissensbasiertes, inhaltsorientiertes Retrieval statistischer Daten mit EISREVU / Ein Verarbeitungsmodell für eine modulare Bewertung von Kennzahlenwerten für den Endanwender
Nr. 8 (1991)	Schwab J.: Ein computergestütztes Modellierungssystem zur Kennzahlenbewertung
Nr. 9 (1992)	Gross H.-P.: Eine semantiktreue Transformation vom Entity-Relationship-Modell in das Strukturierte Entity-Relationship-Modell
Nr. 10 (1992)	Sinz E.J.: Datenmodellierung im Strukturierten Entity-Relationship-Modell (SERM)
Nr. 11 (1992)	Ferstl O.K., Sinz E. J.: Glossar zum Begriffsystem des Semantischen Objektmodells
Nr. 12 (1992)	Sinz E. J., Popp K.M.: Zur Ableitung der Grobstruktur des konzeptuellen Schemas aus dem Modell der betrieblichen Diskurswelt
Nr. 13 (1992)	Esswein W., Locarek H.: Objektorientierte Programmierung mit dem Objekt-Rollenmodell
Nr. 14 (1992)	Esswein W.: Das Rollenmodell der Organsiation: Die Berücksichtigung aufbauorganisatorische Regelungen in Unternehmensmodellen
Nr. 15 (1992)	Schwab H. J.: EISREVU-Modellierungssystem. Benutzerhandbuch
Nr. 16 (1992)	Schwab K.: Die Implementierung eines relationalen DBMS nach dem Client/Server-Prinzip
Nr. 17 (1993)	Schwab K.: Konzeption, Entwicklung und Implementierung eines computergestützten Bürovorgangssystems zur Modellierung von Vorgangsklassen und Abwicklung und Überwachung von Vorgängen. Dissertation

Nr. 18 (1993)	Ferstl O.K., Sinz E.J.: Der Modellierungsansatz des Semantischen Objektmodells
Nr. 19 (1994)	Ferstl O.K., Sinz E.J., Amberg M., Hagemann U., Malischewski C.: Tool-Based Business Process Modeling Using the SOM Approach
Nr. 20 (1994)	Ferstl O.K., Sinz E.J.: From Business Process Modeling to the Specification of Distributed Business Application Systems - An Object-Oriented Approach -. 1st edition, June 1994
	Ferstl O.K., Sinz E.J. : Multi-Layered Development of Business Process Models and Distributed Business Application Systems - An Object-Oriented Approach -. 2nd edition, November 1994
Nr. 21 (1994)	Ferstl O.K., Sinz E.J.: Der Ansatz des Semantischen Objektmodells zur Modellierung von Geschäftsprozessen
Nr. 22 (1994)	Augsburger W., Schwab K.: Using Formalism and Semi-Formal Constructs for Modeling Information Systems
Nr. 23 (1994)	Ferstl O.K., Hagemann U.: Simulation hierarischer objekt- und transaktionsorientierter Modelle
Nr. 24 (1994)	Sinz E.J.: Das Informationssystem der Universität als Instrument zur zielgerichteten Lenkung von Universitätsprozessen
Nr. 25 (1994)	Wittke M., Mekinic, G.: Kooperierende Informationsräume. Ein Ansatz für verteilte Führungsinformationssysteme
Nr. 26 (1995)	Ferstl O.K., Sinz E.J.: Re-Engineering von Geschäftsprozessen auf der Grundlage des SOM-Ansatzes
Nr. 27 (1995)	Ferstl, O.K., Mannmeusel, Th.: Dezentrale Produktionslenkung. Erscheint in CIM-Management 3/1995
Nr. 28 (1995)	Ludwig, H., Schwab, K.: Integrating cooperation systems: an event-based approach
Nr. 30 (1995)	Augsburger W., Ludwig H., Schwab K.: Koordinationsmethoden und -werkzeuge bei der computergestützten kooperativen Arbeit
Nr. 31 (1995)	Ferstl O.K., Mannmeusel T.: Gestaltung industrieller Geschäftsprozesse
Nr. 32 (1995)	Gunzenhäuser R., Duske A., Ferstl O.K., Ludwig H., Mekinic G., Rieder H., Schwab H.-J., Schwab K., Sinz E.J., Wittke M: Festschrift zum 60. Geburtstag von Walter Augsburger
Nr. 33 (1995)	Sinz, E.J.: Kann das Geschäftsprozeßmodell der Unternehmung das unternehmensweite Datenschema ablösen?
Nr. 34 (1995)	Sinz E.J.: Ansätze zur fachlichen Modellierung betrieblicher Informationssysteme - Entwicklung, aktueller Stand und Trends -
Nr. 35 (1995)	Sinz E.J.: Serviceorientierung der Hochschulverwaltung und ihre Unterstützung durch workflow-orientierte Anwendungssysteme
Nr. 36 (1996)	Ferstl O.K., Sinz, E.J., Amberg M.: Stichwörter zum Fachgebiet Wirtschaftsinformatik. Erscheint in: Broy M., Spaniol O. (Hrsg.): Lexikon Informatik und Kommunikationstechnik, 2. Auflage, VDI-Verlag, Düsseldorf 1996

Nr. 37 (1996) Ferstl O.K., Sinz E.J.: Flexible Organizations Through Object-oriented and Transaction-oriented Information Systems, July 1996

Nr. 38 (1996) Ferstl O.K., Schäfer R.: Eine Lernumgebung für die betriebliche Aus- und Weiterbildung on demand, Juli 1996

Nr. 39 (1996) Hazebrouck J.-P.: Einsatzpotentiale von Fuzzy-Logic im Strategischen Management dargestellt an Fuzzy-System-Konzepten für Portfolio-Ansätze

Nr. 40 (1997) Sinz E.J.: Architektur betrieblicher Informationssysteme. In: Rechenberg P., Pomberger G. (Hrsg.): Handbuch der Informatik, Hanser-Verlag, München 1997

Nr. 41 (1997) Sinz E.J.: Analyse und Gestaltung universitärer Geschäftsprozesse und Anwendungssysteme. Angenommen für: Informatik '97. Informatik als Innovationsmotor. 27. Jahrestagung der Gesellschaft für Informatik, Aachen 24.-26.9.1997

Nr. 42 (1997) Ferstl O.K., Sinz E.J., Hammel C., Schlitt M., Wolf S.: Application Objects – fachliche Bausteine für die Entwicklung komponentenbasierter Anwendungssysteme. Angenommen für: HMD – Theorie und Praxis der Wirtschaftsinformatik. Schwerpunkheft ComponentWare, 1997

Nr. 43 (1997): Ferstl O.K., Sinz E.J.: Modeling of Business Systems Using the Semantic Object Model (SOM) – A Methodological Framework - . Accepted for: P. Bernus, K. Mertins, and G. Schmidt (ed.): Handbook on Architectures of Information Systems. International Handbook on Information Systems, edited by Bernus P., Blazewicz J., Schmidt G., and Shaw M., Volume I, Springer 1997

Ferstl O.K., Sinz E.J.: Modeling of Business Systems Using (SOM), 2nd Edition. Appears in: P. Bernus, K. Mertins, and G. Schmidt (ed.): Handbook on Architectures of Information Systems. International Handbook on Information Systems, edited by Bernus P., Blazewicz J., Schmidt G., and Shaw M., Volume I, Springer 1998

Nr. 44 (1997) Ferstl O.K., Schmitz K.: Zur Nutzung von Hypertextkonzepten in Lernumgebungen. In: Conradi H., Kreutz R., Spitzer K. (Hrsg.): CBT in der Medizin – Methoden, Techniken, Anwendungen -. Proceedings zum Workshop in Aachen 6. – 7. Juni 1997. 1. Auflage Aachen: Verlag der Augustinus Buchhandlung

Nr. 45 (1998) Ferstl O.K.: Datenkommunikation. In. Schulte Ch. (Hrsg.): Lexikon der Logistik, Oldenbourg-Verlag, München 1998

Nr. 46 (1998) Sinz E.J.: Prozeßgestaltung und Prozeßunterstützung im Prüfungswesen. Erschienen in: Proceedings Workshop „Informationssysteme für das Hochschulmanagement". Aachen, September 1997

Nr. 47 (1998) Sinz, E.J.:, Wismans B.: Das „Elektronische Prüfungsamt". Erscheint in: Wirtschaftswissenschaftliches Studium WiSt, 1998

Nr. 48 (1998) Haase, O., Henrich, A.: A Hybrid Respresentation of Vague Collections for Distributed Object Management Systems. Erscheint in: IEEE Transactions on Knowledge and Data Engineering

Nr. 49 (1998) Henrich, A.: Applying Document Retrieval Techniques in Software Engineering Environments. In: Proc. International Conference on Database and Expert Systems

Applications. (DEXA 98), Vienna, Austria, Aug. 98, pp. 240-249, Springer, Lecture Notes in Computer Sciences, No. 1460

Nr. 50 (1999) Henrich, A., Jamin, S.: On the Optimization of Queries containing Regular Path Expressions. Erscheint in: Proceedings of the Fourth Workshop on Next Generation Information Technologies and Systems (NGITS'99), Zikhron-Yaakov, Israel, July, 1999 (Springer, Lecture Notes)

Nr. 51 (1999) Haase O., Henrich, A.: A Closed Approach to Vague Collections in Partly Inaccessible Distributed Databases. Erscheint in: Proceedings of the Third East-European Conference on Advances in Databases and Information Systems – ADBIS'99, Maribor, Slovenia, September 1999 (Springer, Lecture Notes in Computer Science)

Nr. 52 (1999) Sinz E.J., Böhnlein M., Ulbrich-vom Ende A.: Konzeption eines Data Warehouse-Systems für Hochschulen. Angenommen für: Workshop „Unternehmen Hochschule" im Rahmen der 29. Jahrestagung der Gesellschaft für Informatik, Paderborn, 6. Oktober 1999

Nr. 53 (1999) Sinz E.J.: Konstruktion von Informationssystemen. Der Beitrag wurde in geringfügig modifizierter Fassung angenommen für: Rechenberg P., Pomberger G. (Hrsg.): Informatik-Handbuch. 2., aktualisierte und erweiterte Auflage, Hanser, München 1999

Nr. 54 (1999) Herda N., Janson A., Reif M., Schindler T., Augsburger W.: Entwicklung des Intranets SPICE: Erfahrungsbericht einer Praxiskooperation.

Nr. 55 (2000) Böhnlein M., Ulbrich-vom Ende A.: Grundlagen des Data Warehousing. Modellierung und Architektur

Nr. 56 (2000) Freitag B, Sinz E.J., Wismans B.: Die informationstechnische Infrastruktur der Virtuellen Hochschule Bayern (vhb). Angenommen für Workshop "Unternehmen Hochschule 2000" im Rahmen der Jahrestagung der Gesellschaft f. Informatik, Berlin 19. - 22. September 2000

Nr. 57 (2000) Böhnlein M., Ulbrich-vom Ende A.: Developing Data Warehouse Structures from Business Process Models.

Nr. 58 (2000) Knobloch B.: Der Data-Mining-Ansatz zur Analyse betriebswirtschaftlicher Daten.

Nr. 59 (2001) Sinz E.J., Böhnlein M., Plaha M., Ulbrich-vom Ende A.: Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen. Angenommen für: WI-IF 2001, Augsburg, 19.-21. September 2001

Nr. 60 (2001) Sinz E.J., Wismans B.: Anforderungen an die IV-Infrastruktur von Hochschulen. Angenommen für: Workshop „Unternehmen Hochschule 2001" im Rahmen der Jahrestagung der Gesellschaft für Informatik, Wien 25. – 28. September 2001

Änderung des Titels der Schriftenreihe *Bamberger Beiträge zur Wirtschaftsinformatik* in *Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik* ab Nr. 61

Note: The title of our technical report series has been changed from *Bamberger Beiträge zur Wirtschaftsinformatik* to *Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik* starting with TR No. 61

Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik

Nr. 61 (2002) Goré R., Mendler M., de Paiva V. (Hrsg.): Proceedings of the International Workshop on Intuitionistic Modal Logic and Applications (IMLA 2002), Copenhagen, July 2002.

Nr. 62 (2002) Sinz E.J., Plaha M., Ulbrich-vom Ende A.: Datenschutz und Datensicherheit in einem landesweiten Data-Warehouse-System für das Hochschulwesen. Erscheint in: Beiträge zur Hochschulforschung, Heft 4-2002, Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung, München 2002

Nr. 63 (2005) Aguado, J., Mendler, M.: Constructive Semantics for Instantaneous Reactions

Nr. 64 (2005) Ferstl, O.K.: Lebenslanges Lernen und virtuelle Lehre: globale und lokale Verbesserungspotenziale. Erschienen in: Kerres, Michael; Keil-Slawik, Reinhard (Hrsg.); Hochschulen im digitalen Zeitalter: Innovationspotenziale und Strukturwandel, S. 247 – 263; Reihe education quality forum, herausgegeben durch das Centrum für eCompetence in Hochschulen NRW, Band 2, Münster/New York/München/Berlin: Waxmann 2005

Nr. 65 (2006) Schönberger, Andreas: Modelling and Validating Business Collaborations: A Case Study on RosettaNet

Nr. 66 (2006) Markus Dorsch, Martin Grote, Knut Hildebrandt, Maximilian Röglinger, Matthias Sehr, Christian Wilms, Karsten Loesing, and Guido Wirtz: Concealing Presence Information in Instant Messaging Systems, April 2006

Nr. 67 (2006) Marco Fischer, Andreas Grünert, Sebastian Hudert, Stefan König, Kira Lenskaya, Gregor Scheithauer, Sven Kaffille, and Guido Wirtz: Decentralized Reputation Management for Cooperating Software Agents in Open Multi-Agent Systems, April 2006

Nr. 68 (2006) Michael Mendler, Thomas R. Shiple, Gérard Berry: Constructive Circuits and the Exactness of Ternary Simulation

Nr. 69 (2007) Sebastian Hudert: A Proposal for a Web Services Agreement Negotiation Protocol Framework . February 2007

Nr. 70 (2007) Thomas Meins: Integration eines allgemeinen Service-Centers für PC-und Medientechnik an der Universität Bamberg – Analyse und Realisierungs-Szenarien. Februar 2007

Nr. 71 (2007) Andreas Grünert: Life-cycle assistance capabilities of cooperating Software Agents for Virtual Enterprises. März 2007

Nr. 72 (2007) Michael Mendler, Gerald Lüttgen: Is Observational Congruence on μ-Expressions Axiomatisable in Equational Horn Logic?

Nr. 73 (2007) Martin Schissler: out of print

Nr. 74 (2007) Sven Kaffille, Karsten Loesing: Open chord version 1.0.4 User's Manual. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 74, Bamberg University, October 2007. ISSN 0937-3349.

Nr. 75 (2008) Karsten Loesing (Hrsg.): Extended Abstracts of the Second *Privacy Enhancing Technologies Convention* (PET-CON 2008.1). Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 75, Bamberg University, April 2008. ISSN 0937-3349.

Nr. 76 (2008) Gregor Scheithauer and Guido Wirtz: Applying Business Process Management Systems? A Case Study. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 76, Bamberg University, May 2008. ISSN 0937-3349.

Nr. 77 (2008) Michael Mendler, Stephan Scheele: Towards Constructive Description Logics for Abstraction and Refinement. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 77, Bamberg University, September 2008. ISSN 0937-3349.

Nr. 78 (2008) Gregor Scheithauer and Matthias Winkler: A Service Description Framework for Service Ecosystems. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 78, Bamberg University, October 2008. ISSN 0937-3349.

Nr. 79 (2008) Christian Wilms: Improving the Tor Hidden Service Protocol Aiming at Better Performancs. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 79, Bamberg University, November 2008. ISSN 0937-3349.

Nr. 80 (2009) Thomas Benker, Stefan Fritzemeier, Matthias Geiger, Simon Harrer, Tristan Kessner, Johannes Schwalb, Andreas Schönberger, Guido Wirtz: QoS Enabled B2B Integration. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 80, Bamberg University, May 2009. ISSN 0937-3349.

Nr. 81 (2009) Ute Schmid, Emanuel Kitzelmann, Rinus Plasmeijer (Eds.): Proceedings of the ACM SIGPLAN Workshop on *Approaches and Applications of Inductive Programming* (AAIP'09), affiliated with ICFP 2009, Edinburgh, Scotland, September 2009. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 81, Bamberg University, September 2009. ISSN 0937-3349.

Nr. 82 (2009) Ute Schmid, Marco Ragni, Markus Knauff (Eds.): Proceedings of the KI 2009 Workshop *Complex Cognition*, Paderborn, Germany, September 15, 2009. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 82, Bamberg University, October 2009. ISSN 0937-3349.

Nr. 83 (2009) Andreas Schönberger, Christian Wilms and Guido Wirtz: A Requirements Analysis of Business-to-Business Integration. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 83, Bamberg University, December 2009. ISSN 0937-3349.

Nr. 84 (2010) Werner Zirkel and Guido Wirtz: A Process for Identifying Predictive Correlation Patterns in Service Management Systems. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 84, Bamberg University, February 2010. ISSN 0937-3349.

Nr. 85 (2010) Jan Tobias Mühlberg und Gerald Lüttgen: Symbolic Object Code Analysis. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 85, Bamberg University, February 2010. ISSN 0937-3349.