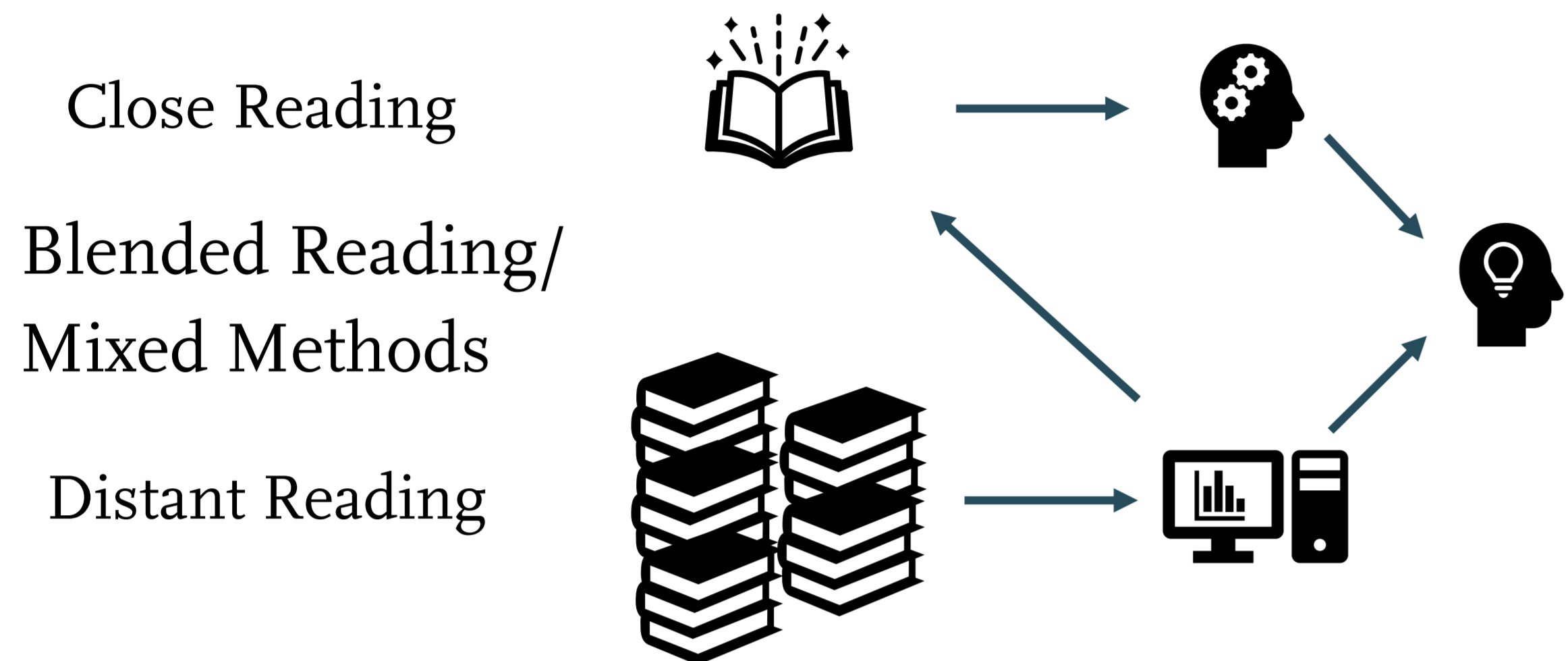


Methoden der automatischen Sprachverarbeitung und Zugänge zu Big Data für den Fachunterricht

Aaron Lang, Andreas Henrich

Otto-Friedrich-Universität Bamberg | aaron.lang@uni-bamberg.de, andreas.henrich@uni-bamberg.de

Zugänge zu (Big) Data



Fragestellungen

Untersucht werden kann beispielsweise:

- Welche Unterschiede und Gemeinsamkeiten existieren zwischen verschiedenen **politischen Parteien**? Wie verändern sich die Themen von Parteien im Zeitverlauf?
- Welche Charaktere treten häufig in **Märchen** auf? Welche Eigenschaften werden ihnen zugeschrieben?
- Wie verändert sich der Gebrauch von **Wörtern** und deren Schreibweisen im Zeitverlauf und welche **zeitpolitischen Aspekte** könnten daran beteiligt sein?

... und das mittels Distant Reading auch für **große Textmengen!**

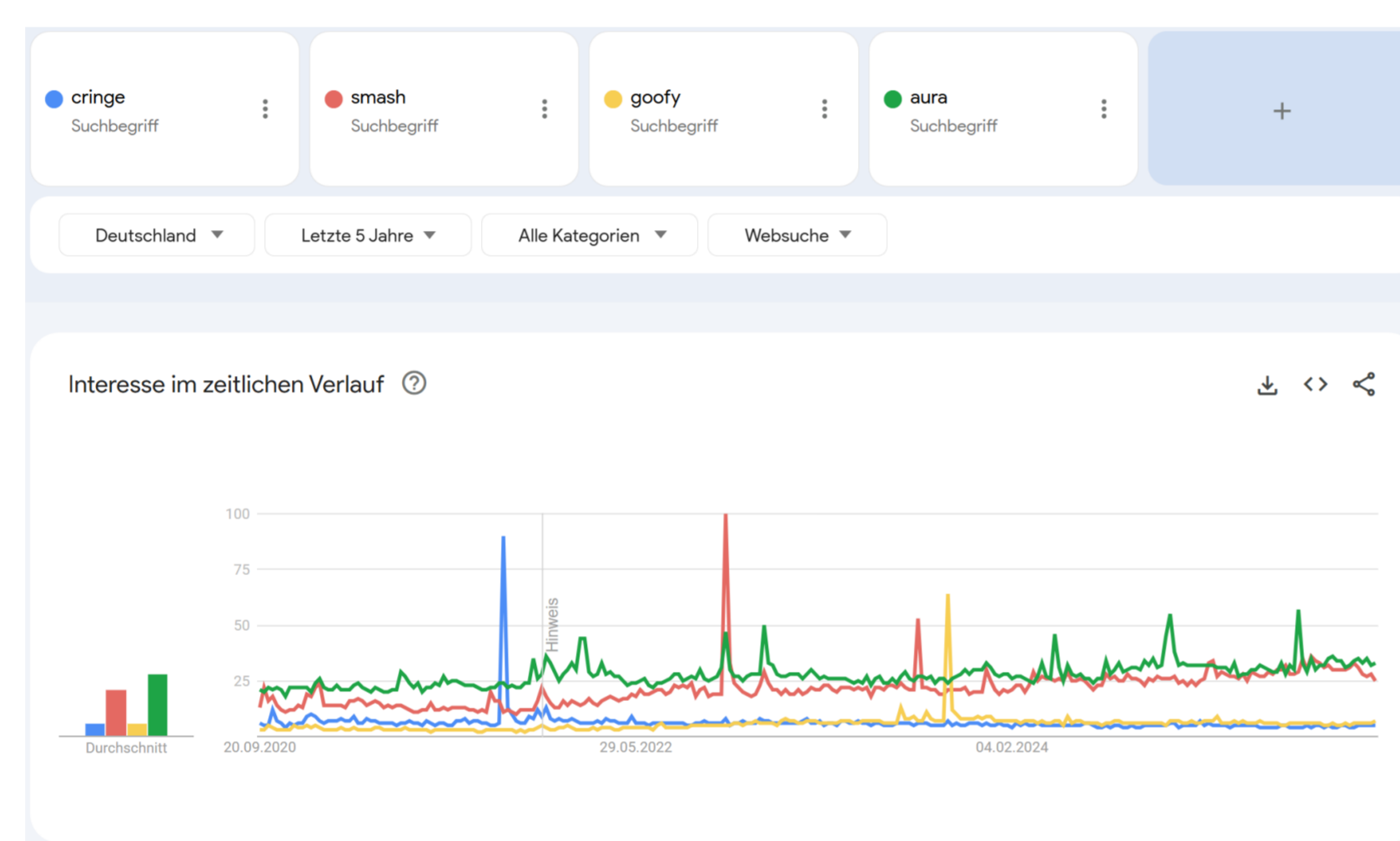
Methoden der automatischen Sprachverarbeitung

Wortfrequenzanalyse



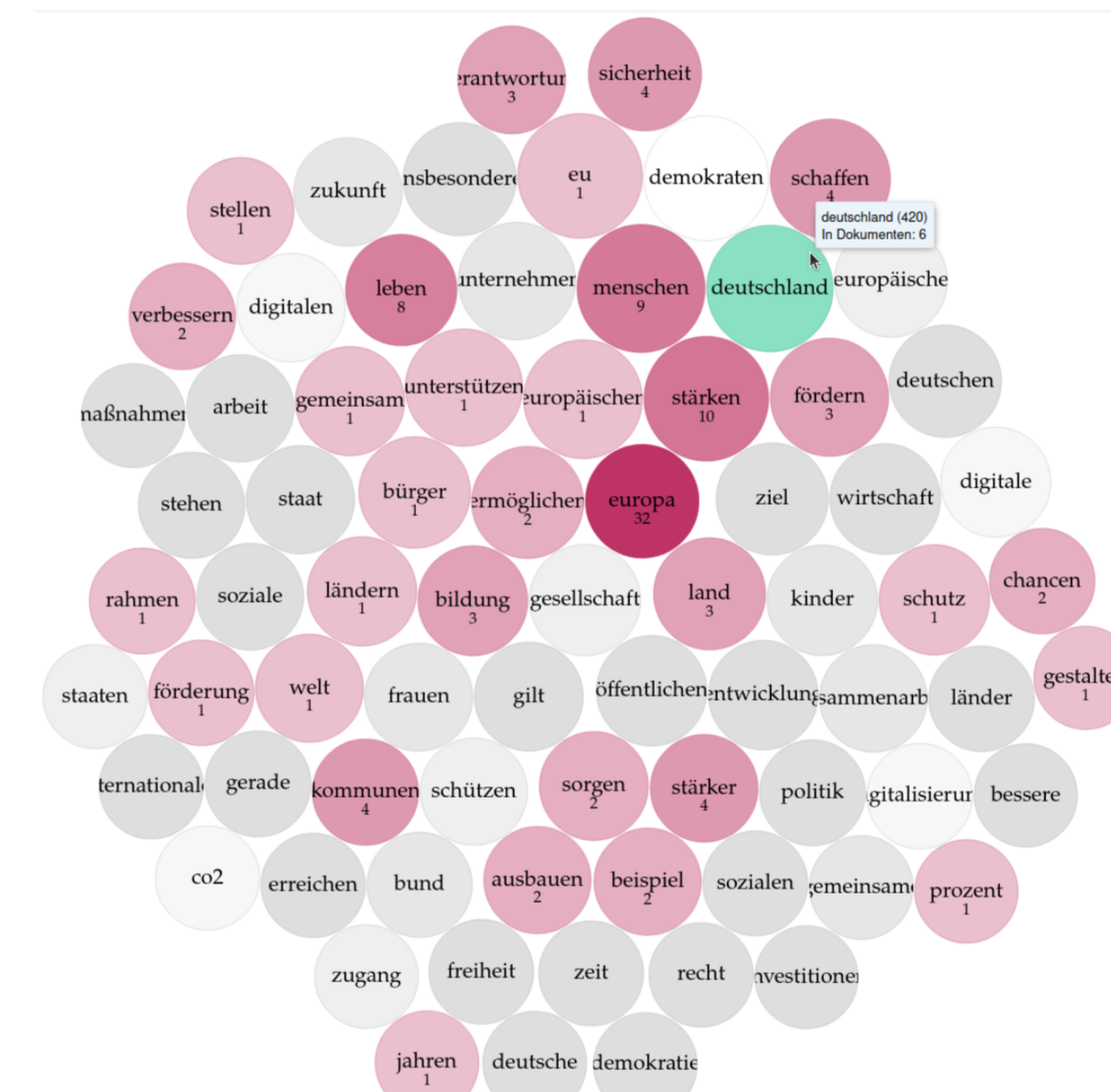
[1]

Trendanalyse



[2]

Kookkurrenzanalyse



[1]

Sentimentanalyse

Habe nun, ach! Philosophie,
Juristerei und Medicin,
Und **leider** auch Theologie!
Durchaus studirt, mit **heißem** Bemühn.
Da steh 'ich nun, ich **armer** Thor!
Und bin so **klug** als wie zuvor;
Heiße Magister, **heiße** Doctor gar,
Und ziehe schon an die zehen Jahr,
Herauf, herab und quer und **krumm**,
Meine Schüler an der Nase herum -- Und sehe, daß
wir nichts **wissen** können!
Das will mir schier das Herz verbrennen.
Zwar bin ich gescheidter als alle die Laffen,
Doctoren, Magister, Schreiber und Pfaffen;
Mich plagen keine Scrupel noch **Zweifel**,
Fürchte mich weder vor **Hölle** noch Teufel -- Dafür
ist mir auch alle Freud ' entrissen,
Bilde mir nicht ein was **Rechts** zu **wissen**,
Bilde mir nicht ein ich könnte was lehren.
Die Menschen zu **bessern** und zu bekehren.

[3]

Fortbildungsinhalte

Einführung



Vergleich Distant Reading vs. Close Reading
Mögliche Fragestellungen

Korpusfindung



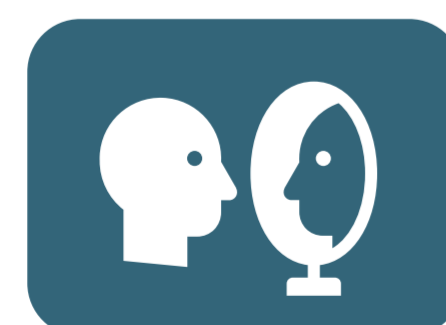
Kriterien zur Korpuswahl wie Repräsentativität, Metadaten
sowie rechtliche und technische Verfügbarkeit

Methoden



Praxisnahes Kennenlernen und Anwenden mittels gängiger
Tools wie Voyant Tools [1], SentText [3], Google Books Ngram
Viewer [4] und DWDS [5]

Reflexion



Grenzen von Distant Reading
Fragen-, Ideen- und Feedbacksammlung

Fokus: Vermittlung **allgemeiner** Kompetenzen mittels **praxisnaher** Hands-Ons

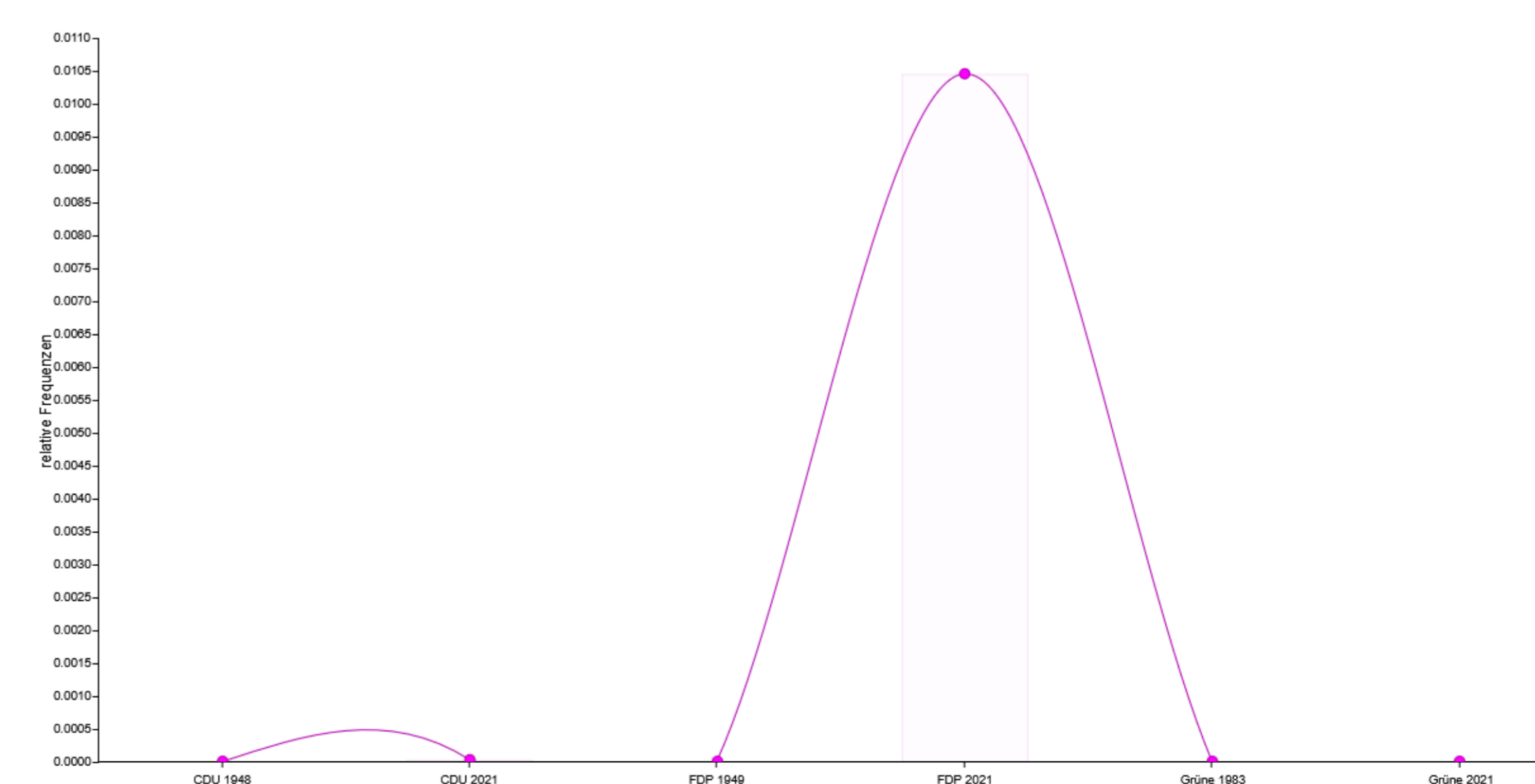
Auszug aus Hands-On zu Voyant Tools

Aufgabe: Die Zusammenfassung zeigt auch die häufigsten Begriffe im Korpus [aus sechs politischen Wahlprogrammen]. Entsprechen diese Ihren Erwartungen? Was wären Ihre Erwartungen gewesen? Untersuchen Sie eventuelle Auffälligkeiten und ermitteln Sie deren Ursprung.

Häufigste Begriffe:

Menschen (468); Deutschland (420); stärken (379); Demokraten (379); EU (374)

„Demokraten“ kommt fast ausschließlich im Wahlprogramm der FDP 2021 vor → Stoppwort!



[1]

Gefördert vom: