Information Retrieval für die Digital Humanities

Tim Gollub und Benno Stein Bauhaus-Universität Weimar webis.de

Workshop Algorithmische Diskursanalyse · Weimar · 26. Oktober 2018

Prozessorientierte Diskursanalyse



Projektgruppe:





Henning Schmidgen Weimar



Bernhard Dotzler Regensburg



Johannes Hess Weimar

Computer Science



Benno Stein Weimar

Tim Gollub

Weimar

Information Science



Frank Simon-Ritz Weimar



Gernot Deinzer Regensburg



Colin Sippl Regensburg

- Aufgabe der Informatik: Technologien für die Diskursanalyse.
- Zentrales Ziel: Entwicklung einer Diskurssuchmaschine.

Diskurssuchmaschine Diskursanalytische Fragestellungen

Exemplarische Fragen an den Diskurs der Lebenswissenschaften um 1900:

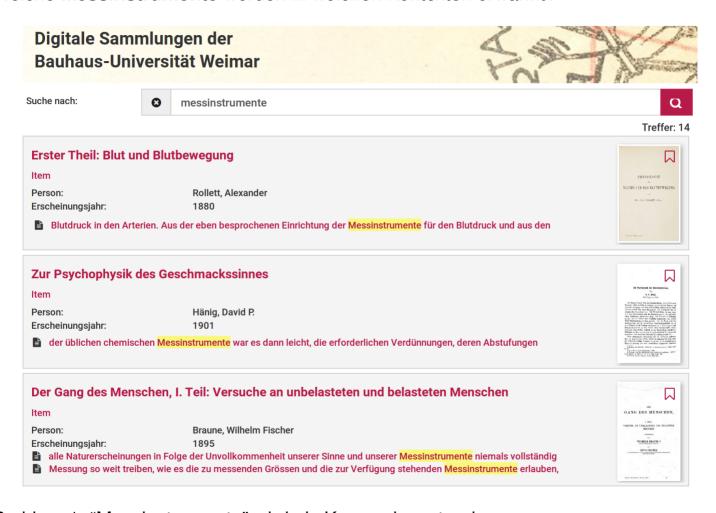
□ Welche **Messinstrumente** werden im Diskurs erwähnt und in welchem Kontext stehen diese Erwähnungen?

Welchen Einfluss hat die Entstehung von Laboratorien auf die Struktur, den Stil und die Argumentationsweise in wiss. Veröffentlichungen?

Welche Konzepte von **Medium** und **Medialität** prägen sich durch das Aufkommen der Laborforschung aus?

Diskurssuchmaschine Baseline Standard-Suchmaschine

Welche Messinstrumente werden in welchen Kontexten erwähnt?



□ Problem 1: "Messinstrumente" wird als Keyword verstanden.

Diskurssuchmaschine Erweiterung 1: Facettensuche zur Anfragemodellierung

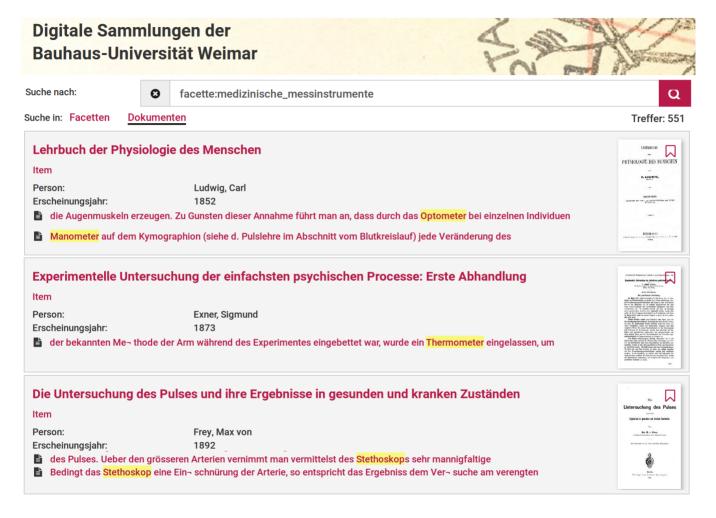
□ Welche **Messinstrumente** werden in welchem Kontext erwähnt?



Lösung 1: Suche nach dem Keyword "Messinstrumente" in einer Facettendatenbank.
Verwenden der Facette "Medizinische Messinstrumente" als Anfrage.

Diskurssuchmaschine Erweiterung 1: Facettensuche zur Anfragemodellierung

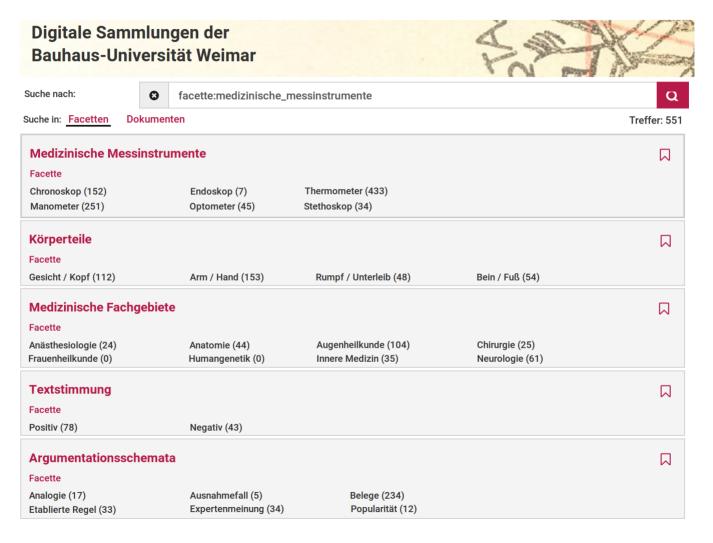
□ Welche Messinstrumente werden in welchem **Kontext** erwähnt?



□ Problem 2: Der Kontext ist aus den Dokumentsuchergebnissen nur schwer systematisch zu erfassen.

Diskurssuchmaschine Erweiterung 2: Relevante Facetten als Suchresultat

□ Welche Messinstrumente werden in welchen Kontexten erwähnt?



□ Lösung 2: Anzeige mit den Resultaten assoziierter Facetten in den Facettensuchergebnissen.

Diskurssuchmaschine Facettengenerierung und -interpretation

Quellen und Interpretationstechnologien für Facetten:

Quelle	Facetten (Beispiele)	Technologie (Beispiele)
Bib. Metadaten	Autor, Publikationsjahr, Verlag, Referenzen	Author Profiling
Knowledge Graphs	Listen von Entitäten	Entity Linking
Wikipedia	Themen, Ereignisse, Handlungen	Semantic Similarity
NLP-Forschung	Argumentation, Stil, Genre, Stimmung	Machine Learning
Geisteswissenschaft	Körperteile, Medialität(?), Diskursregeln(?)	Term Recommendation

Diskurssuchmaschine Feldversuch Philosophische Körper

- Wie häufig tauchen Körperteile in den Werken von Philosophen auf?
- □ Welches "Körperbild" ergibt sich daraus jeweils?

	kant	hegel	nietzsche	freud	heidegger	benjamin	luhmann	wikipedia
Autor	*1724	*1770	*1844	*1856	*1889	*1892	*1927	*2001
Wörter	883,901	4,212,120	469,768	1,629,032	7,211,627	2,681,734	3,079,998	1,093,538,292
Körperteile	424	1,381	699	2,355	2,165	3,403	789	636,186
	0.05%	0.03%	0.15%	0.14%	0.03%	0.13%	0.03%	0.06%
Gesicht / Kopf	112	594	330	780	1,156	1,324	306	225,704
	26.42%	43.01%	47.21%	33.12%	53.39%	38.91%	38.78%	35.48%
Arm / Hand	153	79	144	532	552	1,145	422	118,232
	36.08%	5.72%	20.60%	22.59%	25.50%	33.65%	53.49%	18.58%
Rumpf /	48	334	136	650	243	500	27	127,748
Unterleib	11.32%	24.19%	19.46%	27.60%	11.22%	14.69%	3.42%	20.08%
Bein / Fuß	54	71	6	171	72	175	15	88,084
	12.74%	5.14%	0.86%	7.26%	3.33%	5.14%	1.90%	13.85%
Allgemeines	57	303	83	222	142	259	19	83,442
	13.44%	21.94%	11.87%	9.43%	6.56%	7.61%	2.41%	13.12%

Diskurssuchmaschine Feldversuch Philosophische Körper

- Wie häufig tauchen Körperteile in den Werken von Philosophen auf?
- Welches "Körperbild" ergibt sich daraus jeweils?

