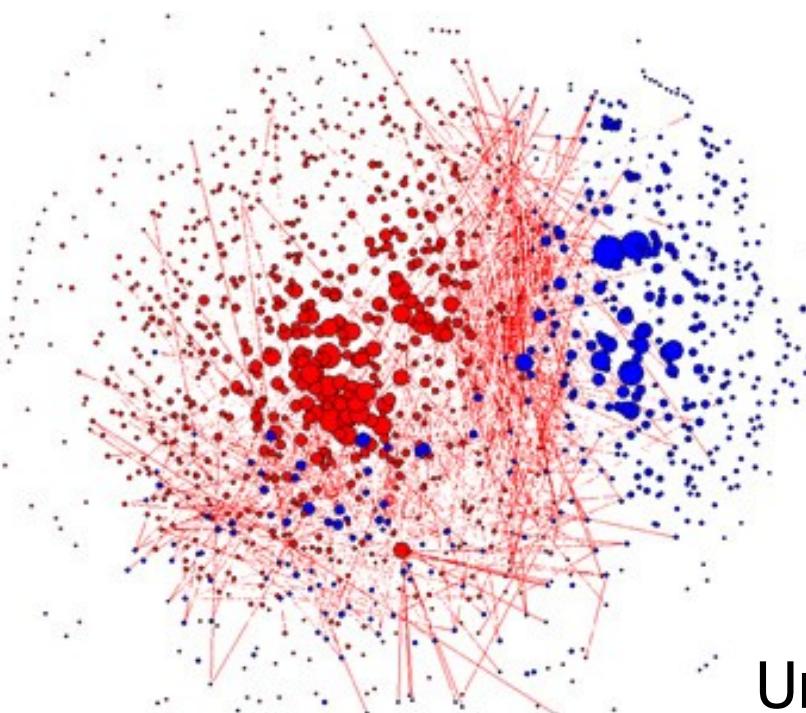




Intelligenz im Chaos: Perspektiven und Probleme der Sacherschließung durch Collaborative Tagging



Christine Krätzsch
Christof Niemann
Universitätsbibliothek Mannheim



Themen

- Fluten, Ströme und Massen: Über die Fassbarkeit des Wissens
- Collaborative Tagging. Gemeinsames Erschließen
- Von Gärtnern und Beeten
- Intelligenz im Chaos?
- Chancen für Bibliotheken



Fluten, Ströme und Massen: Die Fassbarkeit des Wissens

Der Information Overload

- Problem: Rasanter Zuwachs an Datenmengen bei stark eingeschränkter Nutzbarkeit (was ist relevant?)
- Lösung: Verbesserung von Suchalgorithmen, Social Tagging (Nutzerannotationen), manuelle Erstellung von Web-Ontologien

→ *Neue Probleme! - begrenzte „Mathematisierbarkeit“ von Inhalten, „Wildwuchs“ von Schlüsselwörtern oder hoher Bereitstellungsaufwand*



Collaborative Tagging. Gemeinsames Erschließen

Collaborative Tagging: Sacherschließung im Web 2.0

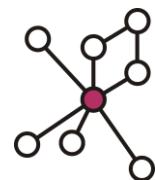
- Freie Vergabe von Schlagwörtern für Artikel, Monographien, Bilder, Videos, Webseiten usw.
- Hohes Innovationspotenzial
- Visualisierung der Verschlagwortung durch Tag-Clouds
- Annahme: die besten Deskriptoren (Tags) werden sich durchsetzen

Das sind nicht zwangsläufig die besten im bibliothekarischen Sinne.



Collaborative Tagging. Gemeinsames Erschließen

Die Tagging-Funktionen im Online-Katalog der Universitätsbibliothek Mannheim



BibSonomy

Collaborative Tagging. Gemeinsames Erschließen

RVK-Notationen • QV 200
RVK-Notationen • QV 202
Schlagwörter
SFX: Volltextsuche
Scandienst
Signatur / Bestell
Signatur / Bestellen
Lok. Notation • MZES ⓘ
Lok. Notation • MZ-15 03.00.00 Structure of Employment
Lok. Notation • MZ-15 09.00.00 Labour Market
Lok. Notation • MZ-15 12.00.00 Organization of Work
Lok. Notation • MZ-18 03.09.00 Modernization
Lok. Notation • MZ-42 83.00.00 UN-Countries / World
Exporte
Meinungen

BibSonomy Export

MZES ⓘ
• MZ-15 03.00.00 Structure of Employment
• MZ-15 09.00.00 Labour Market
• MZ-15 12.00.00 Organization of Work
• MZ-18 03.09.00 Modernization
• MZ-42 83.00.00 UN-Countries / World

BibSonomy BibTeX ⓘ

Rezensionen einblenden/schreiben

Schreiben Sie die erste Online-Rezension zu diesem Titel und teilen Sie anderen Kritikern mit!
Bitte berücksichtigen Sie dabei unsere [Rezensionsrichtlinien](#)!

Tagging Tags ausblenden ⓘ

BibSonomy Tags

Arbeit Arbeitsgesellschaft **Arbeitslosigkeit** Arbeitsmarkt Automatisierung
das_ende_der_klassischen_erwerbsarbeit Dritte_Industrielle_Revolution Dritter_Sektor Erwerbsarbeit Fortschritt
Gesellschaft Globalisierung Industriezeitalter Informationsgesellschaft Informationszeitalter Non-Profit-Organisations
Unterbeschäftigung wirtschaftstheorie **wismasys0809** Wissensarbeiter Zukunft Überproduktion

Weitere Optionen

SPEICHERN/SENDEN

Tag-Cloud BibSonomy

Collaborative Bibliography **einzelnes Erschließen**

Letzte Lesezeichen

Letzte BibTeX-Einträge

Publikationen (0)

Tag-Cloud

Das Debian-System
Martin F. Krafft *Open Source Press, München, (2006)*
als class="text3" LINUX Debian von pertinax am 14.01.2009, 15:08
merken | bearbeiten | löschen | URL | BibTeX | Spam

Hacking exposed Linux
Brian Hatch *Osborne, Berkeley, Calif.; London, 3rd ed. / ISECOM Edition, (2008)*

Es gibt offenbar Muster bei der Tag-Vergabe. Wie erklärt sich das?

Linux-Netzwerke
Stefan Fischer *Millin, Poing, (2005)*
als LINUX von pertinax am 24.10.2008, 14:28

filter:
Relationen
(alle anzeigen | alle verbergen)
Tags
(Alpha | Freq) (Wolke | Liste)
(Minfreq 1 | 2 | 5)

Antike article Befehle
Betriebssystem bibsonomy
bokmarking class="text3"
classification collaborative
connotea Debian
DebianDebian del.icio.us
Dienstleistung Firewall
folksonomy GNULINUX
GNU/LINUX imported
Java Kuntgeschichte
LINUX Linux NT ontology
Referenz Rom social
Suchmaschine tagging tags

Lesezeichen und Referenzen teilen - in blau!
Home meinBibSonomy Lesezeichen eintragen Publikation eintragen Tags Relationen Gruppen Beliebt
Lesezeichen (0) RSS XML bearbeiten << < 1 > >>
Suchmaschinen-Marketing
als Suchmaschine von pertinax am 03.11.2008, 10:55
bearbeiten | löschen | Spam

Ontology is overrated
als tags article classification bibsonomy folksonomy ontology
del.icio.us connotea von pertinax und 31 weiteren Benutzern am 22.10.2008, 14:30
bearbeiten | löschen | Spam

Einfluss
Der Einfluss der Bordleistungen auf die Kundenzufriedenheit Flugreisen unter besonderer Berücksichtigung des Flug als Dienstleistung von pertinax am 16.10.2008, 14:48
bearbeiten | löschen | Spam

Wiki Straßenkarte
als Wik von pertinax und 1 weiterem Benutzer am 15.10.2008, 16:12
bearbeiten | löschen | Spam

Siever. Ellen: *Linux in a nutshell*. 2008



Collaborative Tagging. Gemeinsames Erschließen

Emergierende Ordnung = Schwarmintelligenz?

- Vorbild Natur: Selbstorganisation
- Relativ unintelligente Einzelelemente bilden übergreifende Ordnung
- Flexibel, robust, dezentral, fehlertolerant
- Zielgerichtete Aufgabenerfüllung, Lösung von Optimierungsproblemen
- Mensch: Schwarmverhalten lediglich erste Stufe der kollektiven Intelligenz



Collaborative Tagging. Gemeinsames Erschließen

Kollektive Intelligenz

- Netzwerke: nicht hierarchische Kommunikation
- Intelligente Einzelemente
- Unkontrollierte (unkontrollierbare) Dynamik
- Anything goes: Massenproduktion von Anregungen und Denkbarem





Collaborative Tagging. Gemeinsames Erschließen

Kollektive Intelligenz II: Die Dynamik zähmen

- Problem: Chaos in den Daten
- Lösung: klare Rahmenbedingungen, Distanz zu den Einzelobjekten, Strukturierung der Anregungen

→ *Hierarchisches Handeln zur Beurteilung und Umsetzung der Ideen notwendig*





Von Gärtnern und Beeten

Wildwuchs im Tagging-Beet

- Problem: Chaos in den Tag-Daten mindert deren Nutzbarkeit
- „Erste-Hilfe-Maßnahmen“: Tag-Clouds, Tag-Relationen, Kookkurenzen, automatische Empfehlungen
- Unzureichend für wachsende Größe der Systeme





Von Gärtnern und Beeten

Wildwuchs im Tagging-Beet

- Neue Lösung: „tag gardening“ - Maßnahmen zur Editierung und zur Organisation von Tags
- Manuell durch Nutzer
- Automatisierte Unterstützung
- Differenzierung der Editionslevels (persönlich/kollaborativ, einzelne Quelle/alle Quellen, intra/cross-plattform)





Von Gärtnern und Beeten

Maßnahmen beim Tag Gardening

- Formatierungsempfehlungen
- „weeding“ (Rechtschreibung, Schreibweisen und Wortformen korrigieren)
- „garden design“ (Kategorienbildung, Clustering-Verfahren, Erstellung navigierbarer Taxonomien)
- Invertierte Tag-Clouds, „preferred terms“
- Kombination der Tag-Daten mit Thesauri, Ontologien usw.





Intelligenz im Chaos?

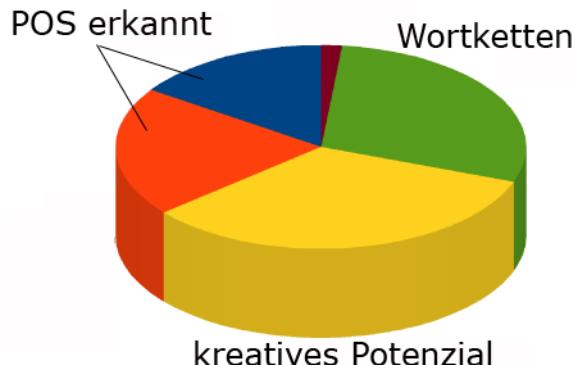
Explorative Analyse der BibSonomy-Daten

- BibSonomy-Datensatz: 97762 Tags
- Analyse durch Natural Language Processing (NLP), die Software „Semtinel“ sowie intellektuelle Erschließung
- Annahme: Strukturen und Inventionen in den Daten bereichern kontrollierte Vokabulare und das Retrieval

Intelligenz im Chaos?

Analyse durch Natural Language Processing

- Problem: große Datenmenge muss sinnvoll reduziert werden
- Lösung: Part-of-Speech Tagging
→ *in etwa Drittelung der Daten*



POS Analyse

Input	Anzahl Tags
Deutsch erkannt	97762
Englisch erkannt	17872
Kreatives Potenzial	23719
Wortketten	38337
Zeichen, Zahlen etc.	33285
	1891

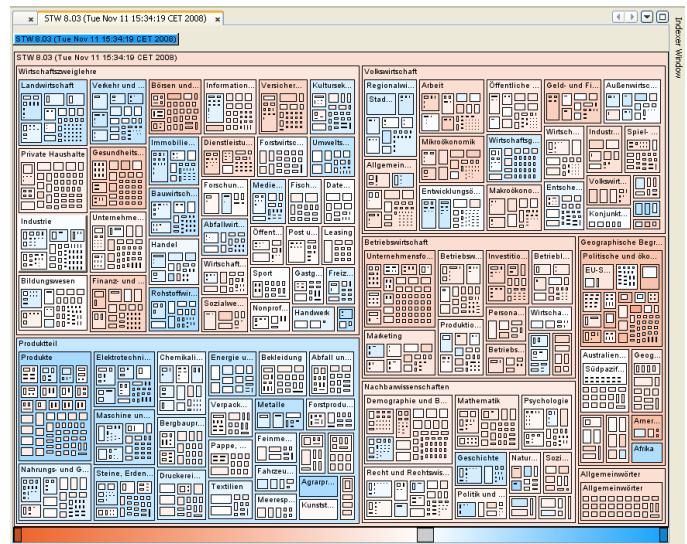


Intelligenz im Chaos?

Lexikalisch erkannte Begriffe

- Problem: Struktur der Tag-Daten bezogen auf Klassifikationen
- Lösung: Matching, Sentinel

- *Matching der Tags und ihrer Häufigkeiten auf RVK, STW usw.*
- *Visualisierung und Analyse der Treffer mittels Sentinel*



Bsp.: Graphische Treemap des STW im „Sentinel“

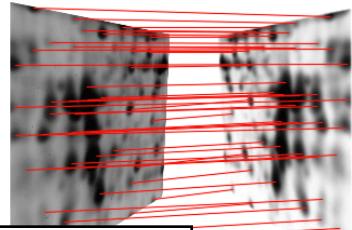
„work in progress“

Intelligenz im Chaos?

Bsp.: Graphische Treemap des STW

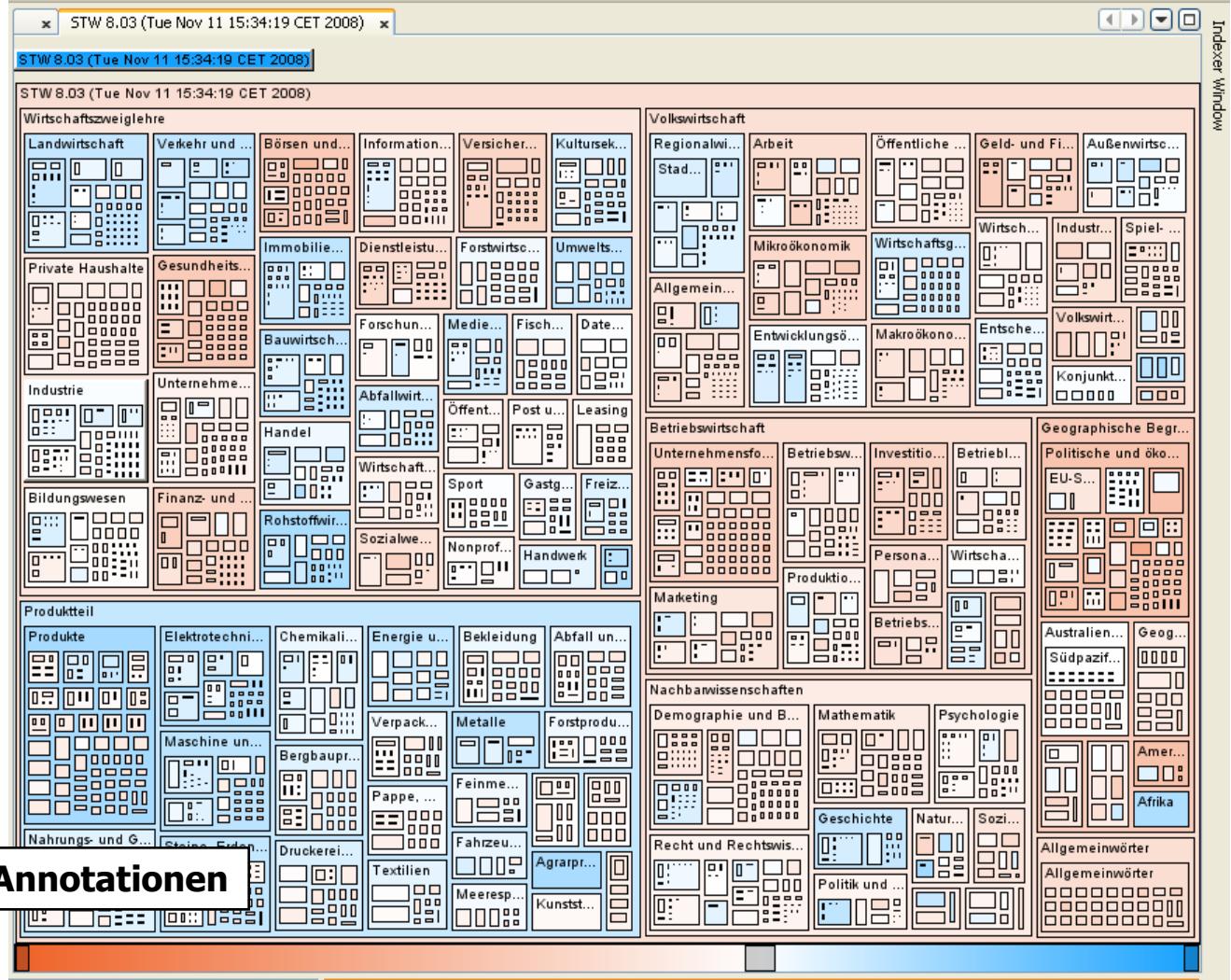
A 1991 Juergen
 1762 analysis
 1889 wismasy50809
 1844 book
 1771 data
 1607 model
 1480 theory
 1463 learning
 1446 nlp
 1395 ecology
 of
 2000
 1395 web
 1395 dev
 1395

Tag-Daten



Matching

Häufigkeit der Annotationen





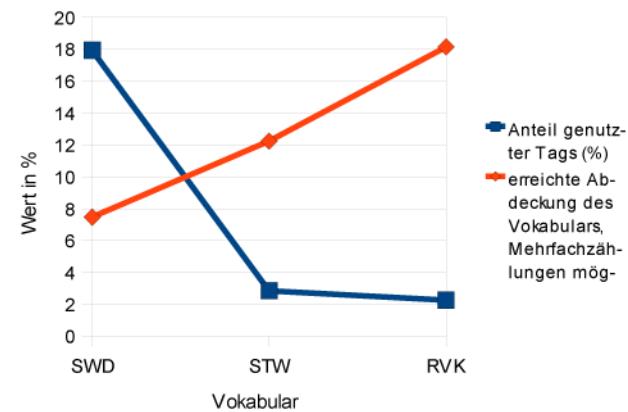
Intelligenz im Chaos?

Matching-Ergebnisse Vokabulare

Vokabular	SWD	STW	RVK
Struktur	flach, wenige Relationen	preferred terms	Hierarchie (Aufstellungs-systematik)
N/Deskriptoren	540166	22800	795000

Matching

N/Treffer	17520	2790	2210
Anteil genutzter Tags (%)	17,92	2,85	2,26
erreichte Abdeckung des Vokabulars, Mehrfachzählungen möglich (%)	7,47	12,23	18,14



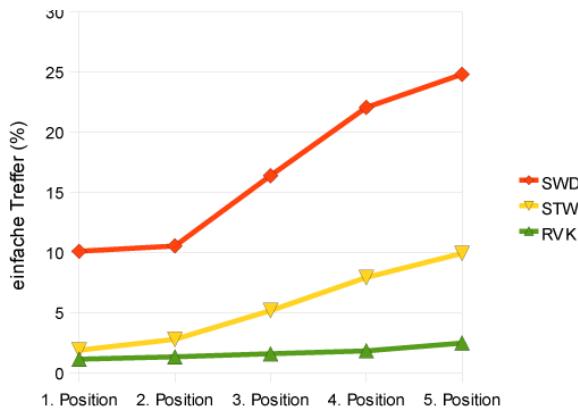
→ STW, RVK: bessere Abdeckung durch englische Begriffe bzw. Begriffswiederholungen

Intelligenz im Chaos?

Matching-Ergebnisse Ketten

Beispiel: Deutschland_Erneuerbare_Energien_Gesetz

Wortketten (N=29830)	SWD	STW	RVK
Einfachzählung Treffer (%)			
1. Position	10,13	1,92	1,16
2. Position	10,57	2,81	1,35
3. Position	16,4	5,2	1,61
4. Position	22,08	7,95	1,86
5. Position	24,82	9,95	2,51



- Je später ein Begriff in der Kette, desto höher ist seine Trefferwahrscheinlichkeit („Qualität“)
- Die meisten Ketten enthalten keine Hierarchie



Intelligenz im Chaos?

Neologismen und kreatives Potenzial

- Problem: Aktualität von kontrollierten Vokabularen
- Lösung: fachliche Prüfung häufig vorkommender Tags

- *Beispiele für interessante Begriffe: Folksonomie, BibTeX, Web 2.0, Semantic Web, E-Learning*
- *Recherche kryptischer Begriffe*

M	N	O
4294	JRR	
4294	CCP	
1889	wismasys0809	
1446	nlp	
1200	devo	
1182	pfc	
1089	Imtm-Bibliothek	
1079	mrefs	
1065	threemode	
1060	folksonomy	
1045	statphys23	
939	system:unfiled	
872	myown	
802	bg	
784	bibtex-import	
732	dblp	
717	diss	
612	high-tc	
582	cogn	
561	ltp	
536	web2.0	
532	nnets	
526	toread	
487	fca	
486	prolearn	
478	Humans	
408	inaki	
398	mythesis	
372	htsct	
364	semanticweb	
361	ownprivate	
331	lang:en	
323	elearning	



Element	Optionen
Allgemeines	
Häufigkeit	Häufigkeit des Tags im Tagging-System
Tag	String
Wortart	NN (Nomen), NE (Eigennamen), AK (Akronym), ADJ (Adjektive), VV (Verb), CD (Zahlen), XY (Sonderzeichen, Nichtwort)
Wortlänge	Anzahl der Buchstaben
Nutzer	Anzahl der Tagger
Intellektuelle Analyse	
Entschlüsselung	Wenn Akronym, Schreibweise untypisch, Wortdrehung etc.: ausgeschriebene (normalisierte) Form
Komposition	Durch Trennzeichen (Kamelschreibweise usw.) erkennbar; Kürzel der jeweiligen Wortart, Komma separiert
Fachbereich	Physik, Informatik etc. (ermittelt aus Journaltiteln der Publikationen)
Sprache	de, en, lat, fr, es, sonst, amb
Fehlergrund	NL (Neologismus), NK (Neukomposition), ORT (Rechtschreibung), FKT (Funktionswort, systemspezifische Wortschöpfungen), NV (nicht im POS-Vokabular), FS (Fremdsprache), XY (Sonderzeichen, Nichtwort), (WK) Wortkette, (TT) Treetagger
Tag Eigenschaften	
bibliometrisch	Autor (A), Publikationsmedium (Pm), Datum (D), Publikationsart (Pa), Herkunft (H)
deskriptiv	konkret (kon), abstrakt (abs)
funktional	Vermeidung, Funktional, Gruppe
Reliabilität	
Belastbarkeit	1=gesichert, 2=begründet vermutet, 3=vermutet
Anmerkungen	

Neologismen und kreatives Potenzial: Kategorien



Chancen für Bibliotheken

Fazit für Bibliotheken

- Das kreative Potenzial der Tagging-Systeme kann zur Pflege und Erweiterung kontrollierter Vokabulare hinzugezogen werden
- Die drei Teilmengen der Tag-Daten sollten unter Berücksichtigung ihrer Besonderheiten genutzt werden
- Die fachliche Aufsicht bei der Aggregation und Integration der Daten ist unabdingbar
- Meta-Tagging-Systeme (Tag Gardening) werden sich zunehmend durchsetzen



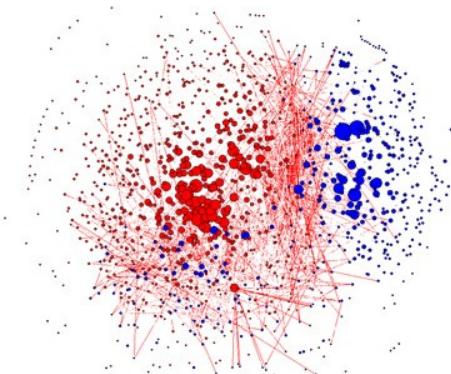
Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

christine.kraetzsch@bib.uni-mannheim.de

christof.niemann@bib.uni-mannheim.de

www.ub.uni-mannheim.de/485.html



Gefördert durch