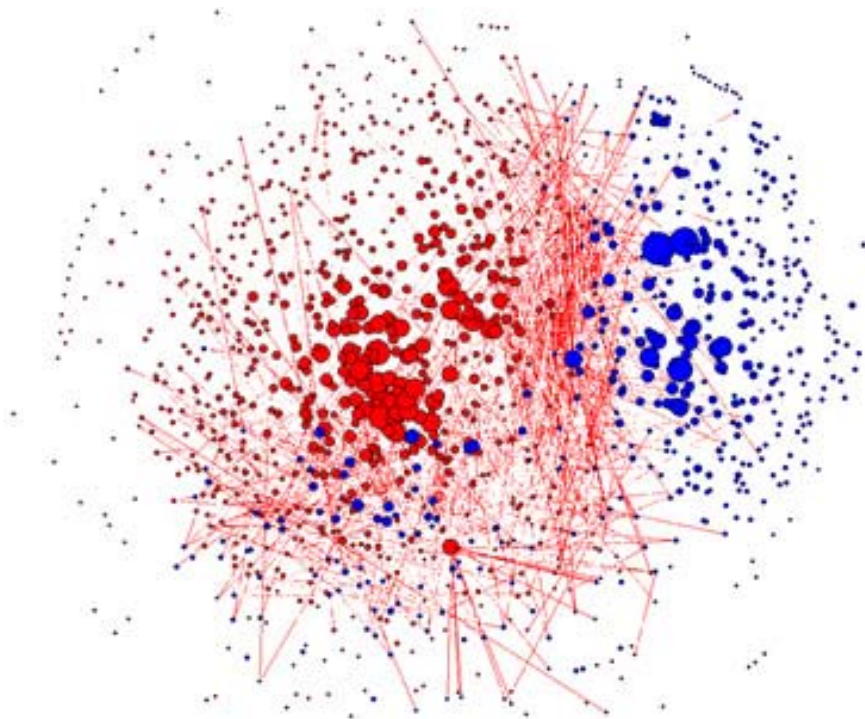




Intelligenz im Chaos: Collaborative Tagging als neue Form der Sacherschließung



Christof Niemann



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



Collaborative Tagging. Gemeinsames Erschließen

RVK-Notationen • QV 200

RVK-Notationen • QV 202

Schlagwörter

SFX: Volltextche

Scandienst

Signatur / Best

Signatur / Bestellen

Lok. Notation

Lok. Notation

Lok. Notation

Lok. Notation

Lok. Notation

Exporte

Meinungen

Tagging

Weitere Optionen

SPEICHERN/SENDEN

BibSonomy Export

MZES i

- 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 03.00.00 UN-Countries / World

BibSonomy BibTeX ?

Rezensionen einblenden/schreiben

Schreiben Sie die erste Online-Rezension zu diesem Titel und teilen Sie anderen K... nit!

Bitte berücksichtigen Sie dabei unsere [Rezensionsrichtlinien!](#)

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

Tag-Cloud BibSonomy

Collaborative Tagging - Gemeinsames Erschließen

Letzte
Lesezeichen

Letzte BibTeX-
Einträge

Tag-Cloud

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

Home | meinBibSonomy | Lesezeichen eintragen | Publikation eintragen | Tags | Relationen | Gruppen | Belieb

Lesezeichen (5) | RSS | XML | Publikationen (10)

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 Flugzeugs
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

Das Debian-System
Martin F. Krafft *Open Source Press, München, (2006)*
als class="text3" GNU/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)*

Linux-Netzwerke
Stefan Fischer *Millin, Poing, (2005)*
als LINUX von pertinax am 24.10.2008, 14:28
merken | bearbeiten | löschen | BibTeX | Spam

filter:

Relationen
(alle anzeigen | alle verbergen)

Tags
(Alpha | Freq) (Wolke | Liste)
(Minfreq 1 | 2 | 5)

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

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 Einzelelemente
- 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

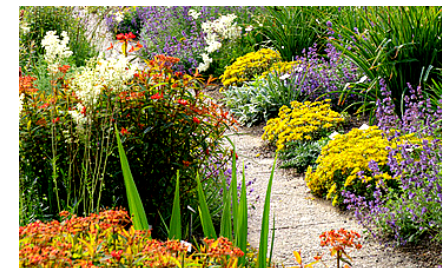
- Neue Lösung: „tag gardening“ - Maßnahmen zur Editierung und zur Organisation von Tags
- Manuell durch Nutzer
- Automatisierte Unterstützung
- Differenzierung der Editionslevels (persönlich/kollaborativ, einzelne Quelle/alle Quellen, intra/cross-plattform)



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?

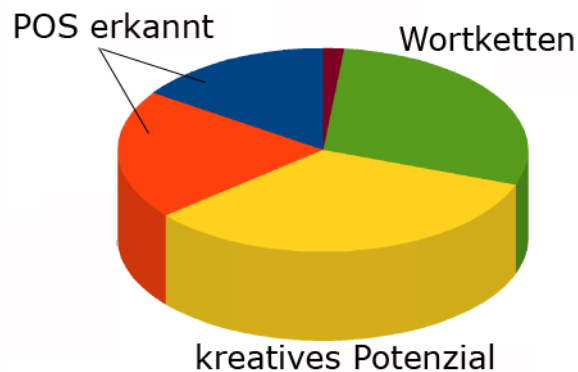
Explorative Analyse der BibSonomy-Daten

- 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*

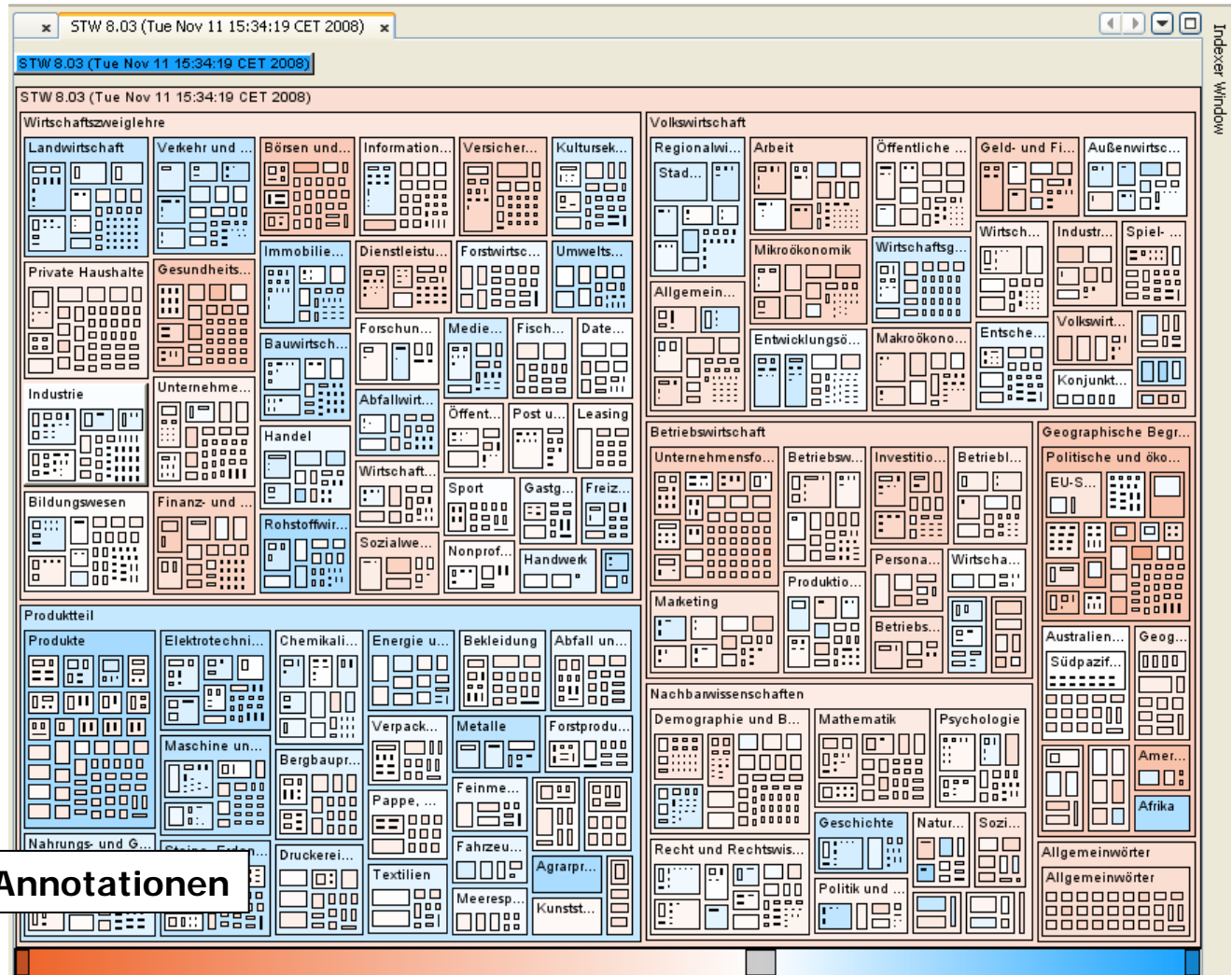


POS Analyse

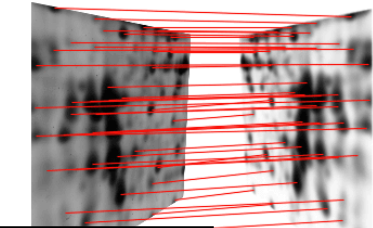
	Anzahl Tags
Input	97762
Deutsch erkannt	17872
Englisch erkannt	23719
Kreatives Potenzial	38337
Wortketten	33285
Zeichen, Zahlen etc.	1891

Intelligenz im Chaos?

Bsp.: Graphische Treemap des STW



Tag-Daten



Matching

Häufigkeit der Annotationen

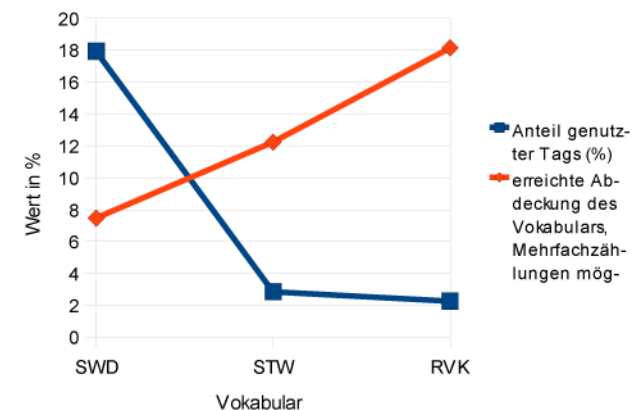
Intelligenz im Chaos?

Matching-Ergebnisse Vokabulare

Vokabular	SWD	STW	RVK
Struktur	flach, wenige Relationen	preferred terms	Hierarchie (Aufstellungssystematik)
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



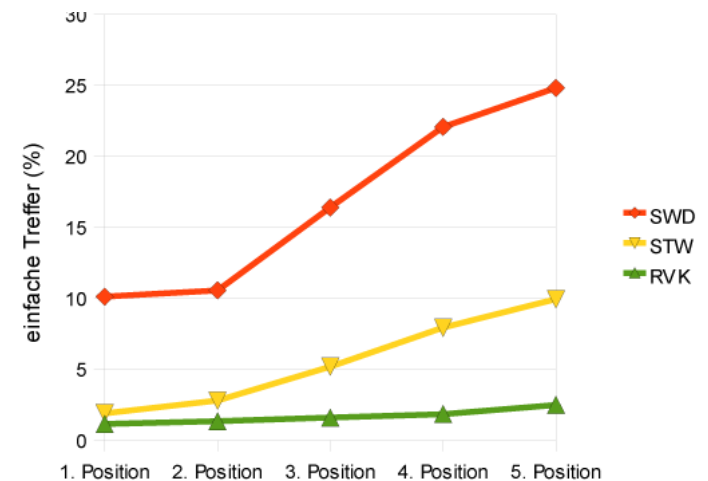
→ *STW, RVK: bessere Abdeckung durch englische Begriffe bzw. Begriffswiederholungen*

Intelligenz im Chaos?

Matching-Ergebnisse Ketten

Beispiel: Deutschland_Erneuerbare_Energien_Gesetz

Wortketten	29830	29830	29830
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	



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

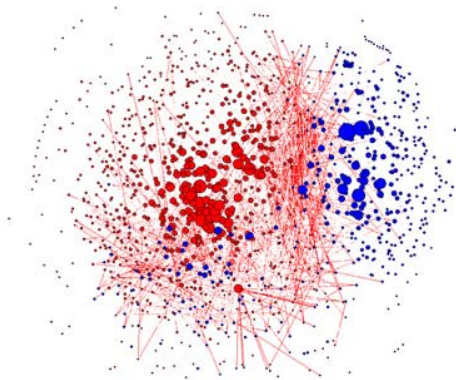


Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

christof.niemann@mail.uni-mannheim.de

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



Gefördert durch