

## Kollaboratives Tagging zur inhaltlichen Beschreibung von Lern- und Wissensressourcen

XML-Tage Berlin, 24.09.2007

**Ronald Maier & Stefan Thalmann**

Institut für Wirtschaftsinformatik,  
Produktionswirtschaft und Logistik

Tätigkeitsbereich Wirtschaftsinformatik

stefan.thalmann@uibk.ac.at

- Metadaten
- Metadatengenerierung
- Kollaboratives Tagging
- Design des Experiments
- Ergebnisse
- Zusammenfassung und Ausblick

- Daten über Daten
- Beschreibung der Charakteristika einer Dateninstanz [Tannenbaum, 2002]
- Teilweise nicht entscheidbar (Kontext der Erstellung bzw. der geplanten Verwendung) [Nilsson et al., 2002]
  - Dienen der Verwaltung (Suche, Nutzung und Management)
  - Zentrales Element bei der Integration von Ressourcen

## Metadatenkategorien: [Maier&Thalman, 2007]

- |                                     |                          |
|-------------------------------------|--------------------------|
| 1. Administrativer Bereich          | (ID, Version)            |
| 2. Beschreibender Bereich           | (Titel, Schlagworte)     |
| 3. Technischer Bereich              | (Format, Größe)          |
| 4. Pädagogisch-Didaktischer Bereich | (Schwierigkeitsniveau)   |
| 5. Nutzungsspezifischer Bereich     | (Nutzungsrechte, Kosten) |

- Beschreibung der Ressourcen traditionell durch
    - Fest vorgegebenes Vokabular
    - Professionelle Metadatenautoren
    - Fachexperten oder Themenverantwortliche [Greenberg, 2003]
  - Stark ansteigende Anzahl von Ressourcen
    - Kürzere Lebenszeiten der Ressourcen
    - Größere Anzahl von Autoren
    - Leichtere Einbindung externer Ressourcen [Greenberg, 2004]
- Aufwand für die traditionelle Art der Beschreibung ist wirtschaftlich und organisatorisch nicht vertretbar. [Cardinaels et al., 2005]
- Automatische Metadatengeneration
    - Analyse der Inhalte (Bildanalysen, Textanalysen)
    - Analyse des Kontextes (Autorendaten, Harvesting)
    - Analyse der Nutzung (Analyse der Nutzungshistorie)

- Ein mit einer Ressource verbundenes Schlagwort
- Gespeichert in einem Tripel aus:
  - Benutzer (Stefan Thalmann)
  - Ressource (Vortrag XML Tage)
  - Schlagwort (Tagging)
- Statistisch auswertbar und grafisch darstellbar in:
  - Tag-Wolken
  - Tag-Clustern
  - Tag-Netzwerken
- Einfache Nutzungsmöglichkeiten
- Individuell: Organisation des eigenen Arbeitsbereiches
- Kollaborativ: Strukturierung des gemeinsamen Bereiches

- Warum ist die Menge schlauer als der Einzelne? [Surowiecki, 04]
  - Kollektive Intelligenz
  - Breites Spektrum an Ideen und Erfahrungen
  - Verschiedene Perspektiven und Verknüpfungen
- Kritische Menge an Taggern wird benötigt
  - Netzwerktheorie [Hofmann, 2001]
- Es lassen sich 2 Mengen von Tags unterscheiden: [Golder & Huberman, 2005]
  - Gemeinschaftlich anerkannte Tags
  - Spezifische (persönliche) Tags
- Anforderungen für die Nutzung in Organisationen
  - Konvergenz gegen eine stabile Menge von Tags
  - Zuverlässige Separierung der Tag-Mengen
  - Beeinflussbarkeit der Beschreibung durch Organisation

- Mehrrunden-Laborexperiment
- Mai – Juni 2007 an der Universität Innsbruck
- Studierende der Wirtschaftswissenschaften im ersten Abschnitt
- Experimente während Proseminaren
- 10 Ressourcen (Videos, Folien und Grafiken)
- Nutzung eines Umfragetools
- Anzeige und direkte Auswahl der 5 am häufigsten ausgewählten Schlagworte möglich
- Berechnung der Top-5 nach jeder Übungsgruppe
- Freie Eingabe von Schlagworten

- Fragestellung:

**1. Commitment:** Kann die Menge von gemeinschaftlich anerkannten und die Menge an spezifischen Tags zuverlässig separiert werden?

**2. Convergence:** Konvergiert die Menge an vergebenen Tags?

**3. Coordination:** Kann die Beschreibungsrichtung durch organisatorische Maßnahmen beeinflusst werden?

- Eckdaten:

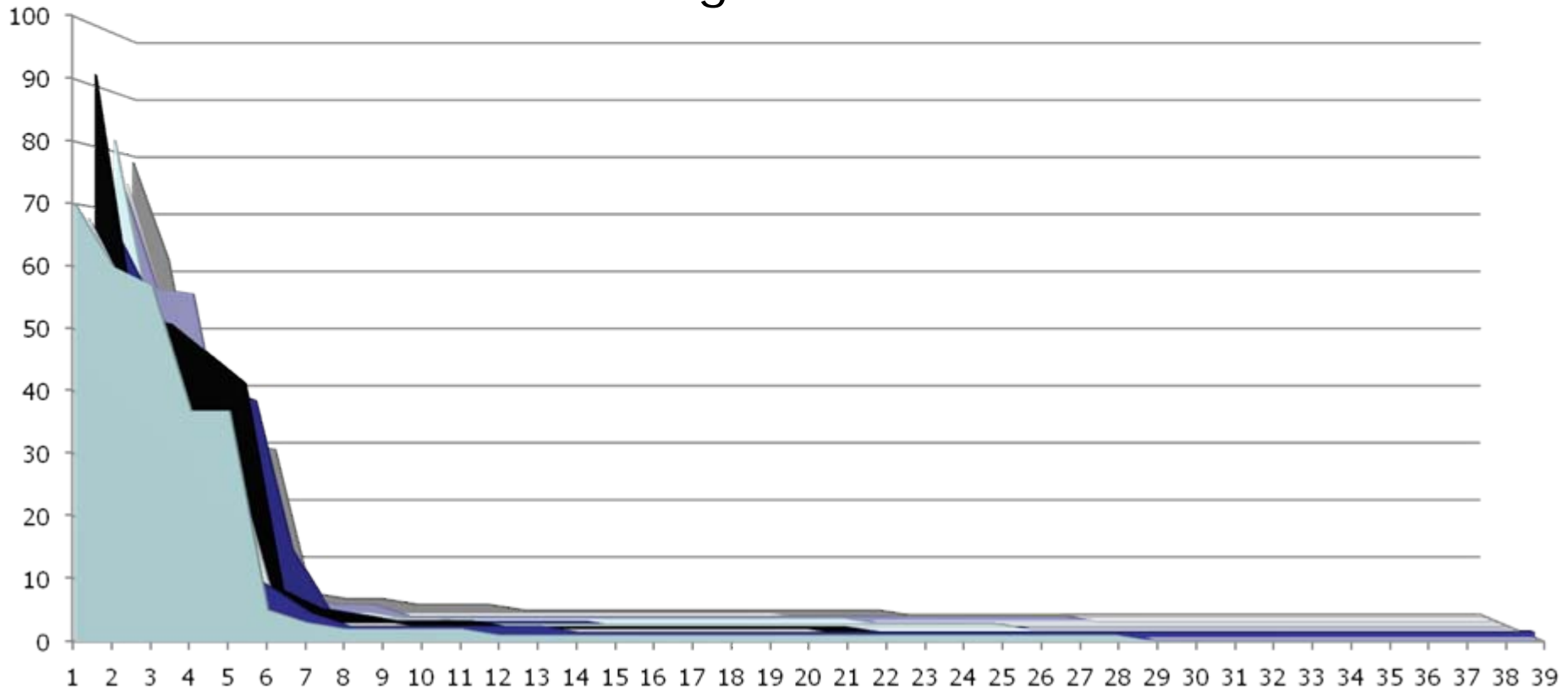
- 404 verschiedene Schlagworte
- 174 Studierende
- 4.246 Annotationen

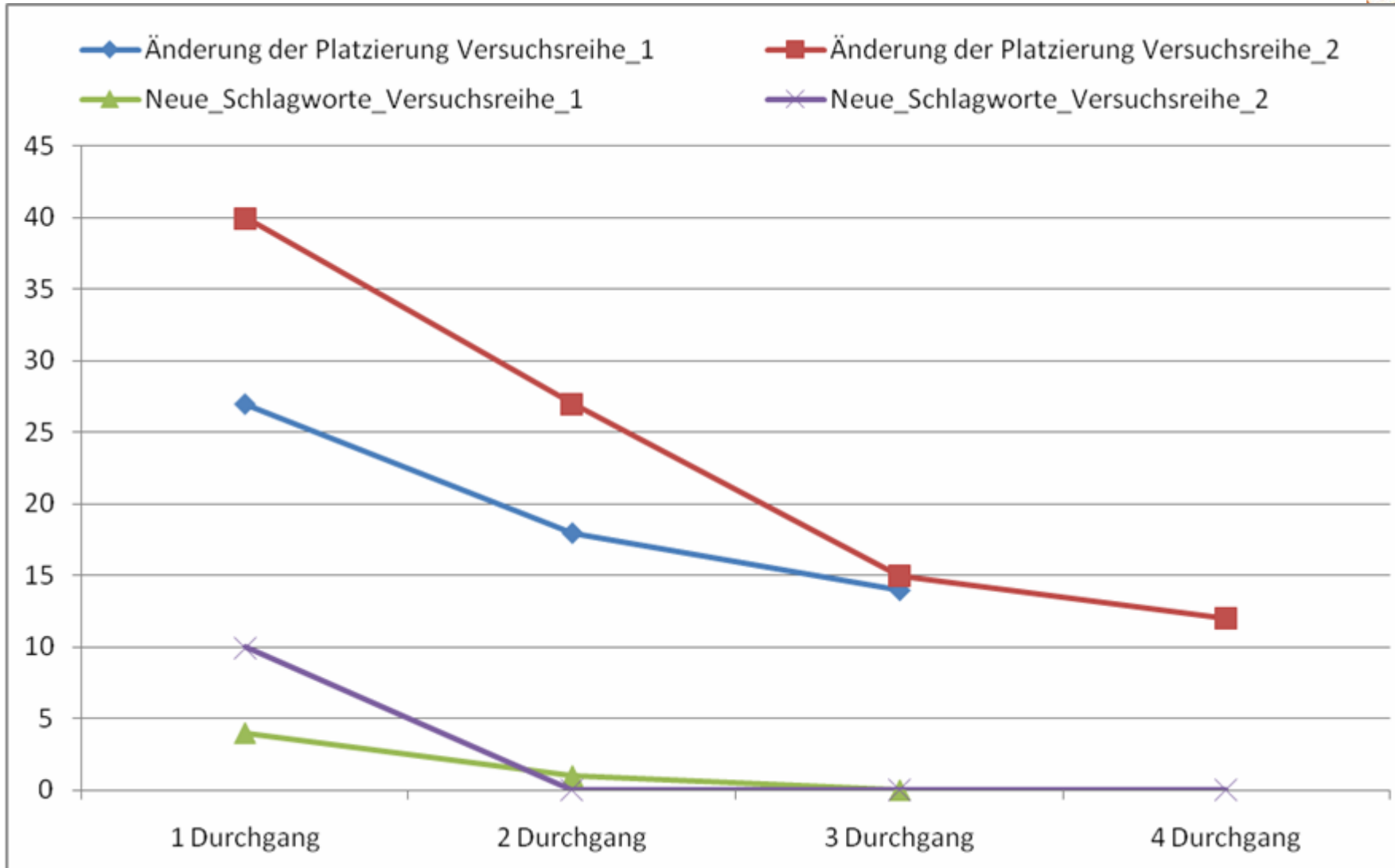
	Versuchsreihe 1	Versuchsreihe 2
1 Durchgang	28 Teilnehmer	27 Teilnehmer
2 Durchgang	20 Teilnehmer	23 Teilnehmer
3 Durchgang	23 Teilnehmer	27 Teilnehmer
4 Durchgang	%	26 Teilnehmer

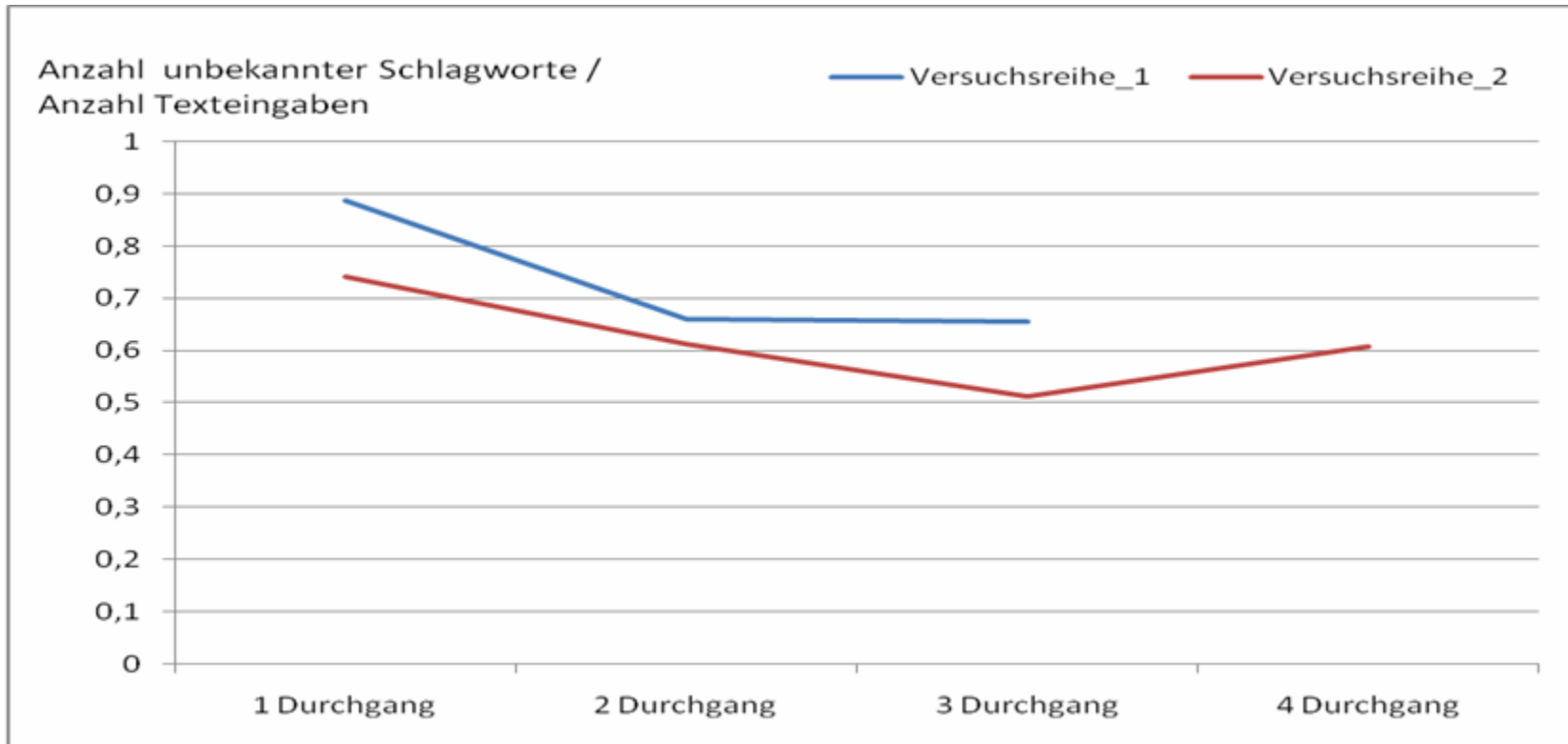
	Ressource 1		Ressource 2		Ressource 3		Ressource 4		Ressource 5	
Platz 1	70	24%	58	22%	68	29%	92	29%	69	28%
Platz 2	60	20%	55	21%	56	24%	52	17%	47	19%
Platz 3	57	19%	53	20%	40	17%	51	16%	44	18%
Platz 4	37	13%	34	13%	25	11%	46	15%	37	15%
Platz 5	37	13%	12	5%	25	11%	41	13%	22	9%
Platz 6	5	2%	8	3%	3	1%	7	2%	3	1%
Platz 7	3	1%	4	2%	2	1%	4	1%	3	1%
Restliche	25	9%	38	15%	15	6%	19	6%	21	9%
Summe	<b>294</b>	<b>100%</b>	<b>262</b>	<b>100%</b>	<b>234</b>	<b>100%</b>	<b>312</b>	<b>100%</b>	<b>246</b>	<b>100%</b>
	Ressource 6		Ressource 7		Ressource 8		Ressource 9		Ressource 10	
Platz 1	69	25%	82	35%	76	29%	75	34%	79	31%
Platz 2	57	21%	51	22%	57	21%	54	25%	62	24%
Platz 3	46	17%	33	14%	56	21%	31	14%	30	12%
Platz 4	40	15%	30	13%	25	9%	30	14%	30	12%
Platz 5	38	14%	22	9%	23	9%	15	7%	29	11%
Platz 6	13	5%	3	1%	4	2%	1	0%	4	2%
Platz 7	3	1%	2	1%	3	1%	1	0%	3	1%
Restliche	8	3%	13	6%	22	8%	12	5%	20	8%
Summe	<b>274</b>	<b>100%</b>	<b>236</b>	<b>100%</b>	<b>266</b>	<b>100%</b>	<b>219</b>	<b>100%</b>	<b>257</b>	<b>100%</b>

Daten Versuchsreihe 2

- 5% Kriterium zur Separierung  
→ 5-6 Ressourcen konnten separiert werden  
→ Diese entsprechen 90% aller Tags
- 404 verschiedene Schlagworte







- Texteingaben pro Teilnehmer zwischen 1,8 und 4 (Ø 2,8)
- Semantisch ähnliche Schlagworte traten häufig auf
- Auf Grund der vielen Kombinationsmöglichkeiten, Stichprobenumfang zu klein

- 8 Ressourcen mit unterschiedlichen Anfangsschlagworten in 2 Versuchsreihen zur Überprüfung der Beeinflussbarkeit
- davon eine Ressource mit ungeeigneten Anfangsschlagworten

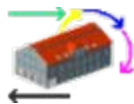


Versuchsreihe 1: den Inhalt beschreibend

Versuchsreihe 2: den Kontext beschreibend



- 2 Ressourcen mit gleichen Anfangsschlagworten



## Diffusions- oder Schnittmengen-Modell



Ronald Maier  
University of Innsbruck



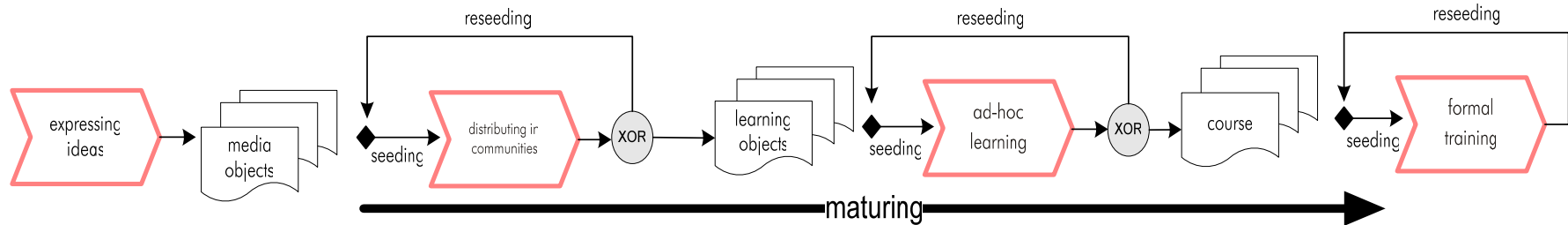
- Diffusionsmodell
- Schnittmengenmodell
- Wirtschaftsinformatik
- Forschungsbereich
- Teilbereich

- Universität\_Innsbruck
- Modell
- Übersichtsgrafik
- Maier
- Unterrichtsfolie

- Texteingaben zwischen beiden Versuchsreihen unterschieden sich zu 71%-93% (Ø 81%).
- Eindeutig mit der Ressource in Verbindung gebrachte Schlagworte setzten sich in beiden Reihen durch.
- Beschreibungsrichtung wurde bei den Texteingaben fortgeführt
- Texteingaben zu den beiden Ressourcen mit identischer Startmenge nahezu identisch, gleiche Texteingaben kamen in die Top-5
- Überdurchschnittlich viele Texteingaben zu den Ressourcen mit der ungeeigneten Startmenge.
- Es konnten sich trotzdem ungeeignete Schlagworte in den Top-5 behaupten.

- Commitment und Coordination unterstützt
- Convergence konnte nicht abschließend geklärt werden
- Tagging eignet sich zur Generierung von kontextabhängigen Metadaten
- Einsatz in Kombination mit automatischen Verfahren erscheint sinnvoll
- Kritisch sind u.a. die notwendige Partizipation durch viele Tagger und die mögliche negative Beeinflussung der Beschreibung
- Empfindlichkeit gegenüber neuen Strömungen nimmt im Zeitverlauf ab
- Einsatz der Ergebnisse im organisatorischen Umfeld

- Anwendung des SER-Modells (seeding, evolutionary growth and reseeding) [Fischer et al. 2001] zur Unterstützung der Wissensreifung.



- Weitere Untersuchungen im Rahmen des EU-IP FP07 MATURE



<http://mature-ip.eu/>