

Tag Recommendations

in an Enterprise 2.0 Tool

Stefan Deser

8. November 2010

Inhalt

Hintergrund

Tagging und Folksonomies
Enterprise 2.0

Ziele der Arbeit

Tag Recommender Systems

TRS: Verschiedene Herangehensweisen
Probleme von TRSs

Zeitplan

Hintergrund

Taxonomie und Folksonomie, Enterprise 2.0.

Klassifikation

Klassische Taxonomien (Klassifikationsschemata) wie z.B.
Internationale Klassifikation der Krankheiten (ICD),
Biologische Systematik oder
Bibliothekssystematiken (Bsp.: Dewey Decimal Classification)

zeichnen sich aus durch

- ▶ **(Mono-)Hierarchie** und
- ▶ ein **kontrolliertes Vokabular**.

Nachteile solcher Schemata:

- ▶ Oft kaum möglich Schemata vorab festzulegen
- ▶ Eindeutige Einordnung nicht immer möglich
- ▶ (Lange) Einarbeitungszeit für Benutzer nötig

Religion in *Dewey's Decimal Classification*

Beispiel für die Problematik der genannten Systematiken

- ▶ Dewey, 200: Religion
 - 210 Natural theology
 - 220 Bible
 - 230 Christian theology
 - 240 Christian moral & devotional theology
 - 250 Christian orders & local church
 - 260 Christian social theology
 - 270 Christian church history
 - 280 Christian sects & denominations
 - 290 Other religions

„How much is this not the categorization you want in the 21st century?“ – C. Shirky, 2005

Collaborative Tagging und Folksonomies

In Web 2.0-Anwendungen entstehen durch kollaboratives Tagging (Verschlagworten) **Folksonomies**:

- ▶ Wort-Neuschöpfung aus *folks* und *taxonomy*
- ▶ Prosumer (Producer/Consumer) indexieren Inhalte selbst
- ▶ Keine zentrale Kontrolle
 - ▶ „grassroot classification“ (A. Mathes)
 - ▶ „bottum-up categorization“ (T. Vander Wal)
- ▶ Folksonomies bilden Gegensatz zu klassischen Systematiken:
 - ▶ Fehlende Hierarchie durch **flat namespace**
 - ▶ **Unkontrolliertes Vokabular**
 - ▶ Keine Einarbeitungszeit
- ▶ Tags in Folksonomies folgen oft der **Power Law**-Verteilung

Narrow und Broad Folksonomies

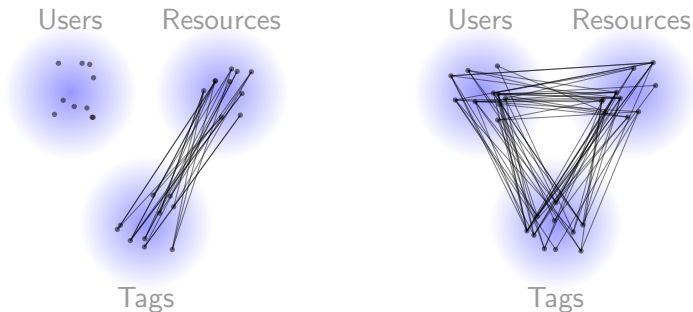


Abb.: Narrow Folksonomy, z.B.:
<http://www.flickr.com>

Abb.: Broad Folksonomy, z.B.:
<http://del.icio.us>

Vorteile von Folksonomies

Tagging resp. Folksonomies eignen sich sehr gut für

- ▶ Suche, Navigation und *Serendipity*
- ▶ Informationsorganisation
- ▶ Entdecken und Kommunizieren von Informationen
- ▶ Entdecken von Trends und Themen in Communities
- ▶ Aufgabenorganisation („todo“, „toread“ usw.)

Erleichterte Partizipation bedeutet

- ▶ Folksonomy „reflects the vocabulary of users“
- ▶ und begünstigt **Aktualität** der Folksonomy

Nachteile und Probleme

Größter Nachteil: Gefahr eines **inkonsistenten Vokabulars**

Problem	Beispiel
Synonymie	synonym, gleichbedeutend
Mehrsprachigkeit	politik, politics, politica
Äquivokationen	Homonym, z.B. „Tau“: Morgendlicher Niederschlag, gr. Buchstabe: τ oder Seil? Polysem, z.B. „Pferd“: Tier oder Schachfigur?
Singular vs. Plural	tag, tags
Synkategorische Tags	todo, ich, später
Groß- und Kleinschreibung, Satzzeichen, Leerzeichen, Sonderzeichen etc.	Tag, tag, ;-), ***, semweb, semantic_web, behörde, behoerde
Abkürzungen, Akronyme etc.	ajax, orm, i18n

Enterprise 2.0

Enterprise 2.0 (McAfee, 2006):

Einsatz von Web 2.0-Software in Unternehmen zur

- ▶ Projektkoordination,
- ▶ zum Wissensmanagement und zur
- ▶ Innen- und Außenkommunikation

Veränderte Bedingungen – **Intranet** statt **Internet**:

- ▶ Kleinerer Benutzerkreis
 - ▶ „Kritische Masse“ möglicherweise schwerer zu erreichen
 - ▶ Usability wichtig
- ▶ Oft Narrow Folksonomy

Tag Recommendations in an Enterprise 2.0 Tool

Ziele der Arbeit.

Ziele der Arbeit

Konsolidierung des Vokabulars und Komfortabilität

1. Theoretischer Teil

- ▶ Festlegung konkreter Anforderungen
- ▶ Auswahl geeigneter Ansätze

2. Praktischer Teil

- ▶ Integration eines *Tag Recommender Systems* in Tricia
(Option: Vorschläge auch für *HybridWiki?*)

Tag Recommendations

Towards a consistent vocabulary.

Tag Recommender Systems

Ein *Tag Recommender System* unterstützt den Indizierenden

- (1) durch das Unterbreiten von Vorschlägen
- (2) in komfortablem User Interface.

Verschiedene Arten von TRSs möglich:

1. Tags als Wörter – *NLP (Natural Language Processing)*
 - ▶ Einsatz von Thesauri, Wörterbüchern wie z.B. *WordNet*
2. Extraktion der Information aus der bisherigen Struktur
 - ▶ „Most popular tags“ oder etwa
 - ▶ *FolkRank* Algorithmus in Anlehnung an *PageRank*
3. Berücksichtigung des zu verschlagwortenden Inhalts
 - ▶ Techniken des Machine Learning, wie z.B.:
Bayes-Klassifikator oder Neuronale Netze.

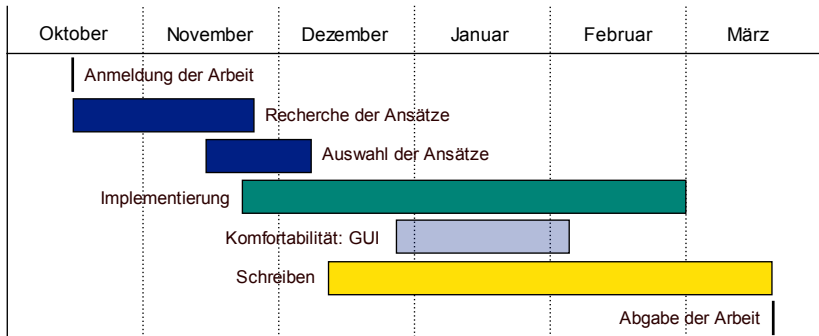
Probleme von TRSs

- ▶ **Cold-Start Problem.** Wie sollen für die ersten noch nicht getaggten Inhalte Vorschläge gemacht werden?
 - ▶ Einbeziehung des Inhalts oder von Eigenschaften der Ressource?
 - ▶ Problem: Viele verschiedenartige Ressourcen möglich!
- ▶ Anfragen an Webservice für Keyword Extraction (z.B. *OpenCalais*) oftmals nicht möglich (z.B. wegen Datenschutz)
- ▶ Wie misst man die **Effizienz** eines TRS?
 - ▶ Qualitativ? Quantitativ?
 - ▶ Welche Metriken gibt es?

Tag Recommendations in an Enterprise 2.0 Tool

Zeitplanung.

Zeitplan



Vielen Dank für die Aufmerksamkeit!