

707.000

## Web Science and Web Technology „Metadata, Tagging and Folksonomies“

How can we acquire, organize, analyze and  
make use of **Data about Data** on a  
participatory web?

**Markus Strohmaier**

Univ. Ass. / Assistant Professor  
Knowledge Management Institute  
Graz University of Technology, Austria

e-mail: [markus.strohmaier@tugraz.at](mailto:markus.strohmaier@tugraz.at)  
web: <http://www.kmi.tugraz.at/staff/markus>

# Overview

## Agenda

- Metadata
- Tagging
- Folksonomies

How can we acquire, organize, analyze and make use of **Data about Data**?

Based in part on slides prepared by M. Lux, Multimedia Information Systems

<http://mathias.lux.googlepages.com/multimediaminformationsystems>

# *OpenData*

# What is Metadata?

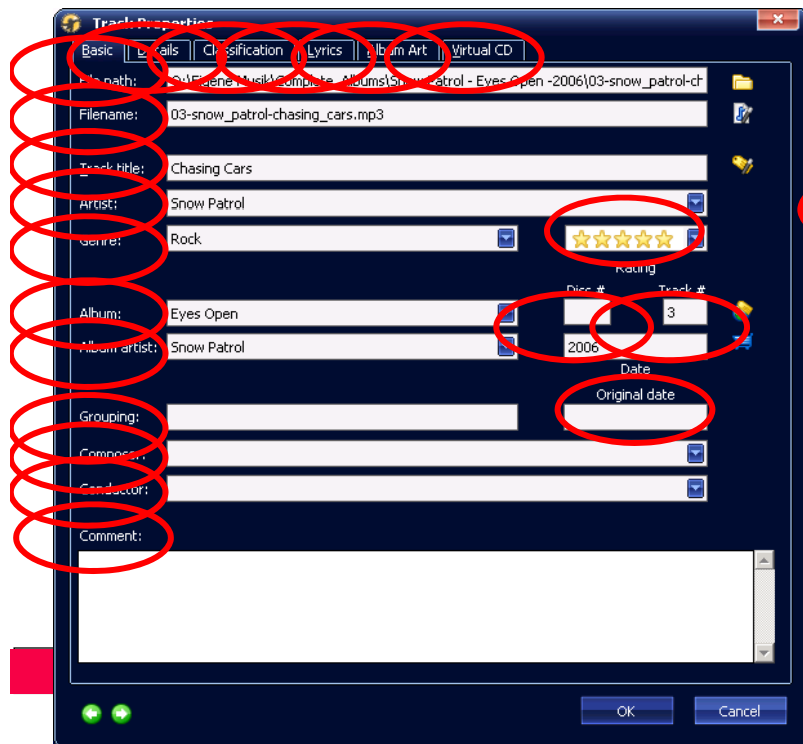
*Metadata is Data about Data*

*Meta<sup>2</sup> data is data about metadata*

*What is metadata used for?  
What is metadata useful for?*

# Aspects of Metadata

- Content Description
- Administrative Aspects
- Quality Metadata
- Legal Metadata
- Technical Metadata



# Classification Systems in Library Science

## The ACM Computing Classification System (1998)

- A. General Literature
  - A.0 GENERAL
    - *Biographies/autobiographies*
    - *Conference proceedings*
    - *General literary works (e.g., fiction, plays)*
  - A.1 INTRODUCTORY AND SURVEY
  - A.2 REFERENCE (e.g., dictionaries, encyclopedias, glossaries)
  - A.m MISCELLANEOUS
- B. Hardware
  - B.0 GENERAL
  - B.1 CONTROL STRUCTURES AND MICROPROGRAMMING (D.3.2)
    - B.1.0 General
    - B.1.1 Control Design Styles
      - *Hardwired control* [\*\*]
      - *Microprogrammed logic arrays* [\*\*]
      - *Writable control store* [\*\*]
    - B.1.2 Control Structure Performance Analysis and Design Aids

## Overview of the Dewey Decimal Classification

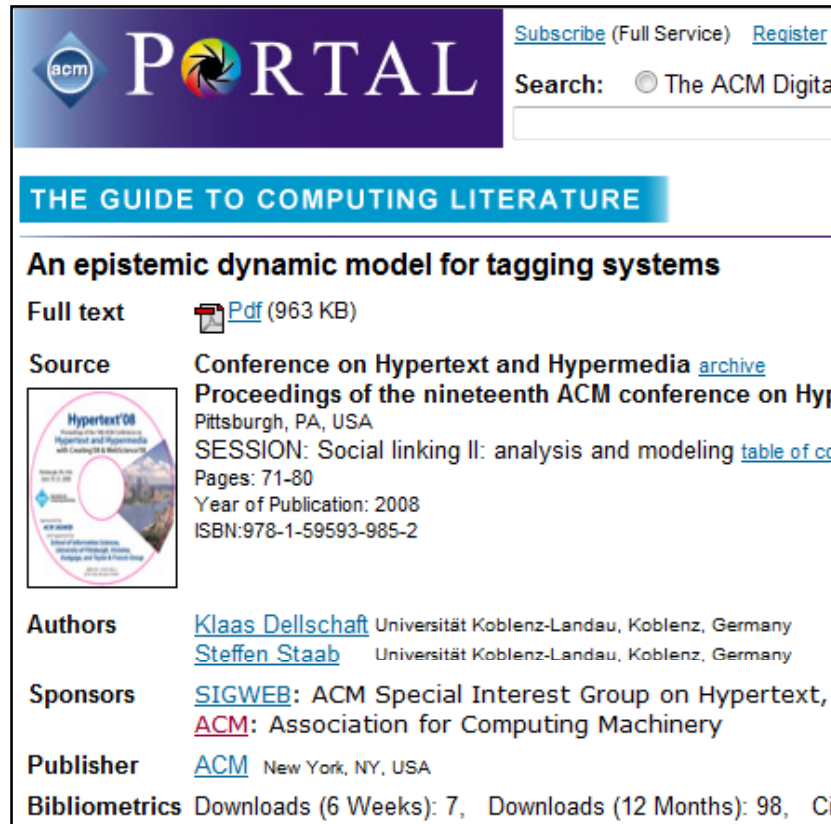
The ten main classes are:

000	Computers, information & general reference
100	Philosophy & psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography

# Indexing Resources

## Categories vs. Keywords

### The ACM Digital Library:




**acm PORTAL** [Subscribe](#) (Full Service) [Register](#)


Search:  The ACM Digital

**THE GUIDE TO COMPUTING LITERATURE**

**An epistemic dynamic model for tagging systems**

Full text  [Pdf](#) (963 KB)

Source [Conference on Hypertext and Hypermedia archive](#)  
 Proceedings of the nineteenth ACM conference on Hypertext and Hypermedia, Pittsburgh, PA, USA  
 SESSION: Social linking II: analysis and modeling [table of contents](#)  
 Pages: 71-80  
 Year of Publication: 2008  
 ISBN:978-1-59593-985-2



**Authors** [Klaas Dellschaft](#) Universität Koblenz-Landau, Koblenz, Germany  
[Steffen Staab](#) Universität Koblenz-Landau, Koblenz, Germany

**Sponsors** [SIGWEB](#): ACM Special Interest Group on Hypertext, Hypermedia, and Web  
[ACM](#): Association for Computing Machinery

**Publisher** [ACM](#) New York, NY, USA

**Bibliometrics** Downloads (6 Weeks): 7, Downloads (12 Months): 98, Citation Count: 5

↑ **INDEX TERMS**

**Primary Classification:**

- H. [Information Systems](#)
  - ↳ H.5 [INFORMATION INTERFACES AND PRESENTATION \(I.7\)](#)
    - ↳ H.5.3 [Group and Organization Interfaces](#)
      - ↳ **Subjects:** [Collaborative computing](#)

**Additional Classification:**

- H. [Information Systems](#)
  - ↳ H.5 [INFORMATION INTERFACES AND PRESENTATION \(I.7\)](#)
    - ↳ H.5.3 [Group and Organization Interfaces](#)
      - ↳ **Subjects:** [Theory and models](#)

**Keywords:**

[complex systems](#), [stochastic modeling](#), [tagging](#), [temporal evolution](#)

→ **Categories and keywords serve different functions**

# Finding Resources

## *Searching vs. Browsing*

Example: Journal of Universal Computer Science



**Keywords** are used to facilitate *search*

### Articles by Topics

- [Topic A. - General Literature](#)
- [Topic B. - Hardware](#)
- [Topic C. - Computer Systems Organization](#)
- [Topic D. - Software](#)
- [Topic E. - Data](#)
- [Topic F. - Theory of Computation](#)
- [Topic G. - Mathematics of Computing](#)
- [Topic H. - Information Systems](#)

**Categories** are used to facilitate *browsing*



# Metadata Applications

## Retrieval & Browsing

- No need to download / view the whole video
- Push vs. Pull

## Management & Organization

- Rights, Billing, Ordering, Classification

## Adaptation

- Transformation to appropriate representation

## Service Description

- Orchestration, Harmonization, Access
- On technical and semantic level

# Aspects of Metadata: Content Description

## Agenda

- Overview about a presentation or a sequence of information to a particular topic

## Table of Contents

- A list of all segments and their position

## Abstract

- Describes the topic of a content within a few sentences.

## Preface

- Some words of the author

## Structure

- For consumption & navigation

And many others, such as Key words & Index, Summary, Literature reference & footnotes, Comments, Categories, Languages, Associated persons, History of Changes, Unique identifier, Versions

# Aspects of Metadata: Quality Aspects

## Weight

- Prioritization of segments

## Expiration Date

- Time period of validity of the content.

## Reviews

- Opinions, arguments from others.

## Process description & history

- Who corrected, translated and approved the content eg. within an workflow.

## Quality Assessment

- Rating of the (e.g. visual) quality of the content

# Aspects of Metatdata: Legal Metadata

## Copyright

- Person or company legally permitted to sell or trade with the content.

## Publish Date

- Date when the content has been released to public.

## License Model

- This is the mode how consumers are allowed to reuse the content



**creativecommons**

**Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 2.0 Österreich**

**Sie dürfen:**

-  das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen
-  Bearbeitungen des Werkes anfertigen

**Zu den folgenden Bedingungen:**

-  **Namensnennung.** Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).
-  **Keine kommerzielle Nutzung.** Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.
-  **Weitergabe unter gleichen Bedingungen.** Wenn Sie dieses Werk bearbeiten oder in anderer Weise umgestalten, verändern oder als Grundlage für ein anderes Werk verwenden, dürfen Sie das neu entstandene Werk nur unter Verwendung von Lizenzbedingungen weitergeben, die mit denen dieses Lizenzvertrages identisch oder vergleichbar sind.

# Aspects of Metadata: Technical Metadata

## Standards:

- Description of the standardized structure in which the content and the metadata are stored.

## Application/System

- application the content and metadata can be / has been processed.
- Resolution, compression of pictures or video clips.

## Encryption Method

- In case of encrypted content

## Storage Media

- on which the content has been stored e.g. CDs, tapes, MO, paper etc.

## Logs

- Technical history

# Media Production: Dublin Core

## Aims to provide

- Common denominator for metadata
- Simple yet powerful schema

## Dublin Core Metadata Initiative defined

- 15 elements (author, date, title, type, ...)
- Further refinements (creation date, extent, ...)

## Dublin Core does not provide

- A schema for storage
- A schema for data types (e.g. dates)

# Dublin Core

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

<b>Inhalt (Content)</b>	
<b>Title</b>	<b>Name</b> der WR (=Wissensressource); vergeben vom Erzeuger oder Herausgeber
<b>Subject &amp; Keywords</b>	<b>Thema &amp; Gegenstand</b> der WR; typischerweise wird das Subject durch Schlüsselwörter/ <b>keywords</b> , die den Inhalt <b>beschreiben</b> , repräsentiert Schlüsselwörter sollten aus einem standardisierten Set stammen (Thesaurus, etc.)
<b>Description</b>	<b>textuelle Beschreibung</b> der WR; <b>Abstracts</b> (bei Textdokumenten) oder <b>Inhaltsbeschreibung</b> bei visuellen Ressourcen
<b>Source</b>	<b>eindeutige Identifizierung der Quelle</b> , aus der diese WR stammt (wenn zutreffend); z.B. ISBN Nummer des Buches aus dem die PDF-Version der WR stammt
<b>Language</b>	<b>Sprache</b> der WR; wenn möglich konform mit RFC 1766
<b>Relation</b>	<b>Beziehung</b> der WR zu <b>anderen WRs</b> ; beschreibt die formalen Beziehungen von wissensobjektmäßig getrennten aber inhaltlich zusammengehörenden WRs; z.B. Bilder in Dokumenten, Kapitel in einem Buch
<b>Coverage</b>	räumliche/temporale Charakterisierung der WR
<b>Urheberschaft (Intellectual Property)</b>	
<b>Creator</b>	die für den intellektuellen Inhalt dieser WR <b>primär verantwortliche Person</b> oder <b>Organisation</b>
<b>Publisher</b>	<b>Herausgeber</b> der WR; z.B. Verlag, Universität, etc.
<b>Contributor</b>	<b>Person</b> oder <b>Organisation</b> die <b>sekundär</b> zu dieser WR <b>beigetragen</b> hat (und nicht im Creator-Feld genannt wird); Z.B. Übersetzer, Illustrator, etc.
<b>Rights</b>	Beschreibung der <b>Copyrights</b> auf diese WR
<b>Intsanzierung (Instantiation)</b>	
<b>Date</b>	<b>Datum</b> , an dem diese WR verfügbar gemacht wurde, empfohlenes Format: YYYY-MM-DD
<b>Type</b>	<b>Kategorie/Typ</b> der Ressource: Arbeitspapier, technical report, Erzählung, Homepage, etc.; standardisierte Namen erwünscht (z.B. <a href="http://sunsite.berkeley.edu/Metadata/types.html">http://sunsite.berkeley.edu/Metadata/types.html</a> )
<b>Format</b>	Datenformat der WR
<b>Resource Identifier</b>	Zeichenkette, die die WR <b>eindeutig identifiziert</b> ; z.B. URL, ISBN,...



# Library of Congress

- It takes 2 years of training to being able to use the LoC classification system
- It costs ~50 USD to classify a book
  
- What are alternatives?





[apparently](#) [apple](#) [asahi](#) [asks](#) [autopia](#) [batteries](#) [behest](#) [bittorrent](#) **[blog](#)** [case](#) [chris](#)  
[kohler](#) [community](#) [compete](#) [computer](#) [crankshaft](#) [cult](#) [disastrous](#) [discovery](#) [download](#) [engine](#)  
[fuel](#) [functional car](#) [game life](#) [gearbox](#) [geek](#) [giant](#) [global](#) [google](#) [help](#) [idea](#) [intel](#) [intellectual](#)  
[property](#) [ipod](#) [janna glasper](#) [john sculley](#) [kohler](#) [lepton](#) [lead](#) [leader](#) [macintosh](#) [mail](#) [media](#)  
**[microsoft](#)** [models](#) [money](#) [moving parts](#) [nasa](#) [new york](#) [notebook](#) [open source](#) [p2p](#)  
[personal](#) [pistons](#) [popular](#) [portables](#) [powerbook](#) [presence](#) [rods](#) [rootkit](#) [running](#) [sabah](#) [scientists](#)  
[search](#) [service](#) [sex drive](#) [slashdot](#) [sony](#) [space](#) [state](#) [story](#) [university](#) [v8 engine](#) [video](#)  
[games](#) [wired magazine](#) [xbox 360](#)

# Metadata and Social Software

## What is Social Software / web2.0?

Many views and definitions

Some common aspects of social software:

*“unprecedented emphasis on voluntary participation, user-control, emergent structures, self-organization and the facilitation of social interactions and social activities “*

# Metadata in the context of social software / web2.0

In the context of social software metadata

Is bottom up

- In contrast to controlled vocabularies
- In contrast to quality ensured content creation processes

Represents a superimposed structure

- Instead of using predefined hierarchies
- Through heavy use of linking / interrelation

Is huge and fuzzy

- Unimaginable mass of links & tags
- Lots of redundant information

Is being spammed

- Just starting ...

# Folksonomies

Definition & Description

Advantages and Disadvantages of Folksonomies

# Folksonomies

A folksonomy is a **user-generated classification, emerging through bottom-up consensus** [1]

- Network of Tags, Users and URLs
- Users describe resources
- By using (multiple) tags

Examples:

Social bookmarking, media sharing, etc.

[1] <http://www.iskoi.org/doc/folksonomies.htm>

## Folksonomies: The Structure

- User *tags* resource (URL)
- 1+ words or phrases (graz, „markus strohmaier“)
- No controlled vocabulary, taxonomy
- No quality control
- No constraints (language, length, number)

# A Simple Tag Ontology

[Tom Gruber, International Journal on Semantic Web & Information Systems, 3(2), 2007.]

*Expressing tagging relationships:*

Tagging(object, tag)

*Considering the user:*

Tagging(object, tag, tagger)

identifying vocabulary of users

*Considering namespaces:*

Tagging(object, tag, tagger, source)

identifying vocabulary of applications

*Considering positive and negative tags:*

Tagging(object, tag, tagger, source, + or -)

e.g. dealing with spam  
(„not X“)

# Folksonomies: Structure

## Tagging(object, tag, tagger)

The screenshot shows a Del.icio.us profile for 'joshua' by Joshua Schachter. The page lists various saved items with their titles, descriptions, and the number of people who saved them. A 'tags' sidebar on the right lists various categories with their respective counts. Three callout boxes are overlaid on the page: 'User' points to the user's name, 'URL' points to a specific item title, and 'Tags' points to the 'tags' sidebar.

**User**

**URL**

**Tags**

del.icio.us / joshua / by Joshua Schachter

All joshua's items (10127)

popular | recent

er | help

page 1 of 102

tags

- 20 3d
- 2 770
- 1 adobe
- 5 ads
- 43 advertising
- 47 aero
- 2 agile
- 105 ai
- 28 ajax
- 11 algo
- 10 alife
- 1 anime
- 4 apache
- 7 api
- 1 appengine
- 102 apple
- 1 arch
- 5 archaeology
- 23 architecture
- 229 art
- 56 astro
- 1 asymmetry
- 2 atari
- 1 attention
- 66 audio
- 1 auth
- 3 automation
- 3 av
- 43 backup
- 24 bay
- 1 biblio

Ubiquitous: Free dynamic graph visualization software save this

vis graph ... saved by 52 other people ... 2 hours ago

Persevere save this

distributed arch via json persistence

json ... saved by 46 other people ... 2 hours ago

save this

interfaces to amazon aws stuff

non s3 ec2 ... saved by 352 other people ... 2 days ago

Ajaxian » Persevere: JSON Storage / Application Server save this

to todo dev ... saved by 8 other people ... 2 days ago

Open GPS Tracker save this

neat

to gis gps sms dev hw ... saved by 588 other people ... 2 days ago

20 Useful Tools to Make Web Development More Efficient save this

to web dev js css ... saved by 2075 other people ... 2 days ago

Doctype save this

documenting the use of html and associated tech

to dox web dev ... saved by 2192 other people ... 2 days ago

pymeta save this

parser generator

to python dev ... saved by 30 other people ... 5 days ago

Understanding Games save this

to games ... saved by 19 other people ... 5 days ago



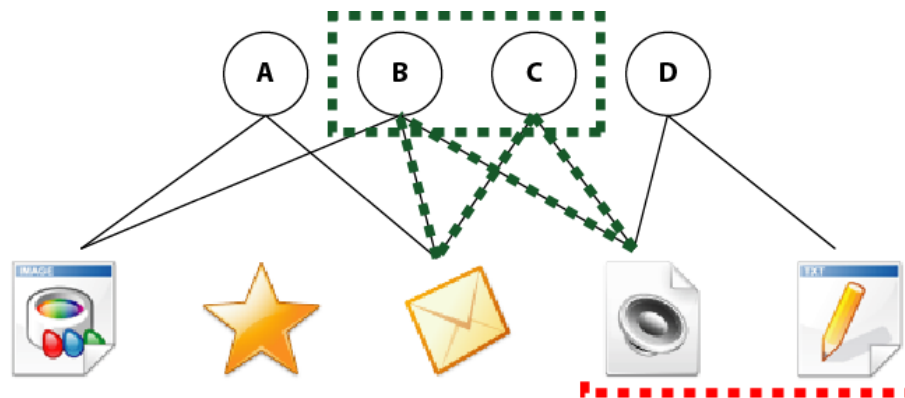
## Folksonomies: Structure

Tag to URL is a n:m relation

Superimposed structure through bidirectional links

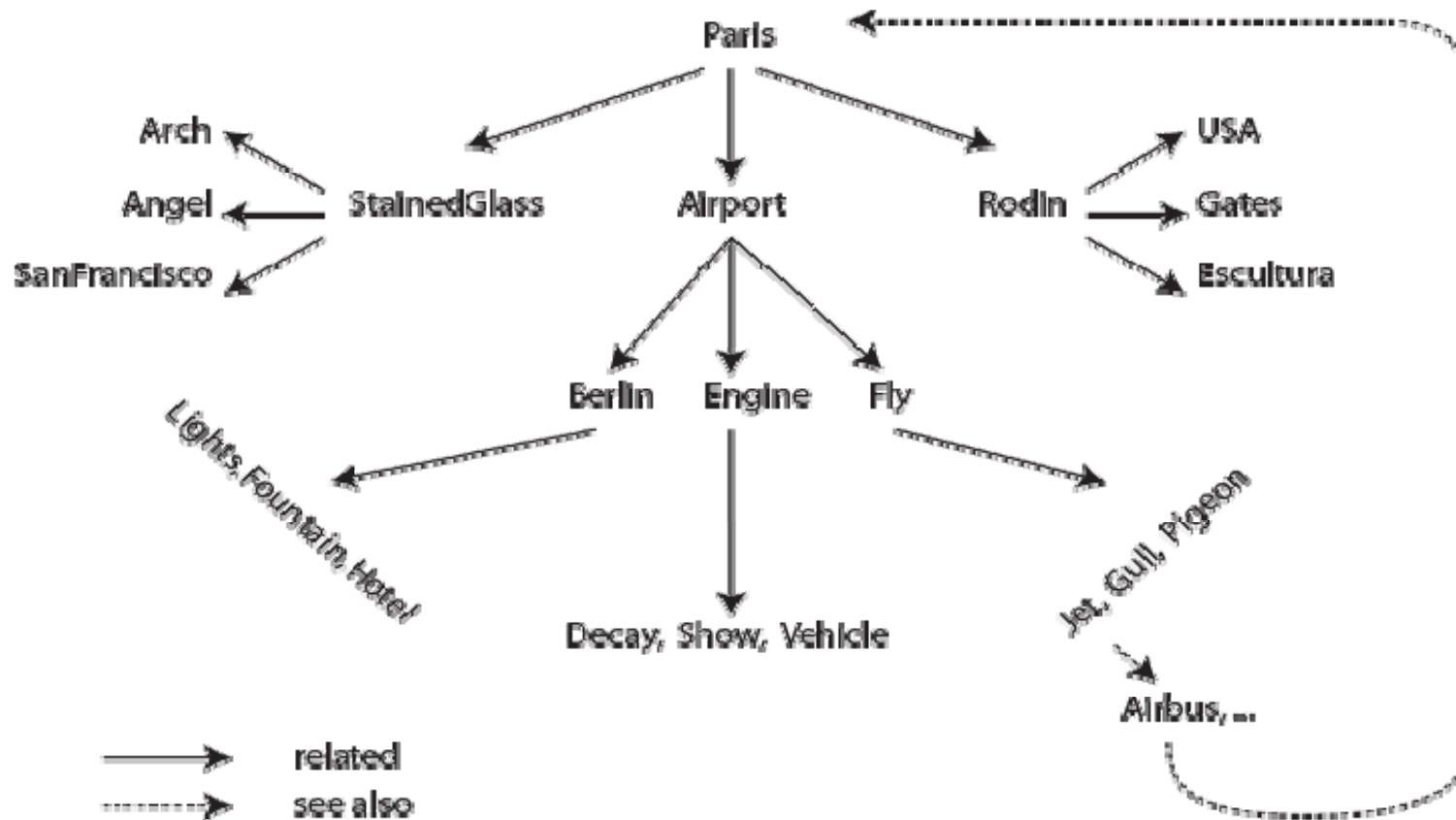
Structure is called „folksonomy“

Tagging(object, tag)



*Two-mode networks in  
social network analysis  
terms*

# Folksonomy Example: Flickr



# Folksonomy Example: Technorati



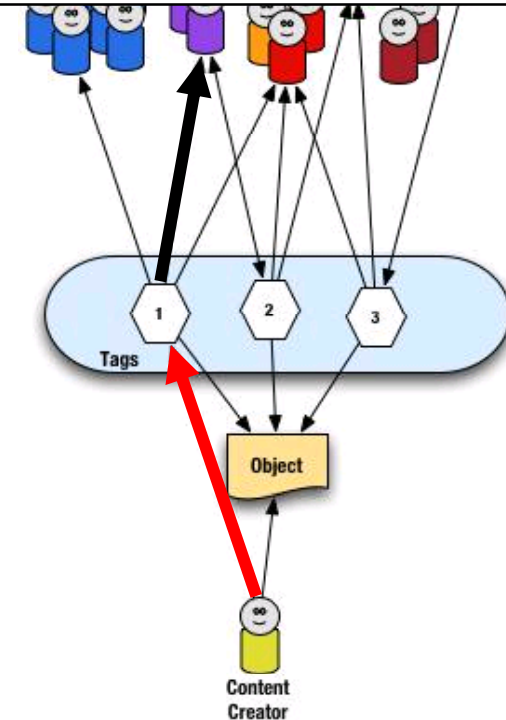
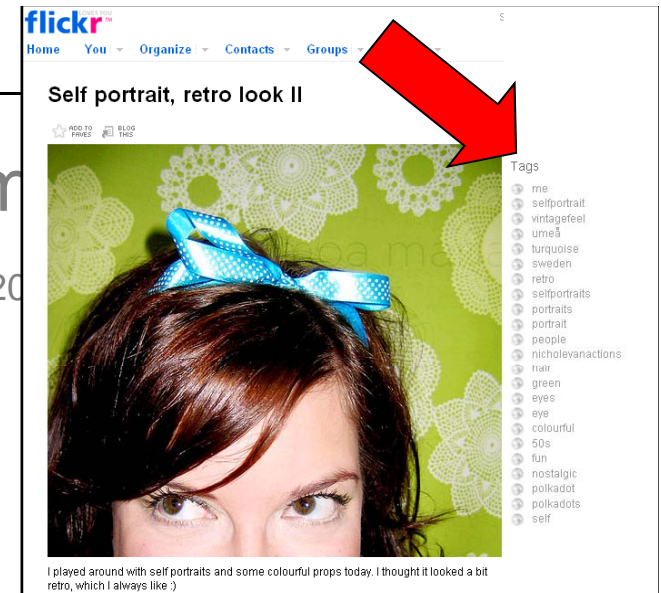


# Types of Folksonom

[Thomas Vander Wal <http://www.personalinfocloud.com/20>

## Narrow folksonomies

- tagging objects that are **not easily searchable** or have no other means of using text to describe or find the object
- done by **one or a few people** providing tags that the person uses to get back to that information.
- The tags, unlike in the broad folksonomy, are **singular in nature**
- tags are **directly associated with the object**.
- Example: Flickr

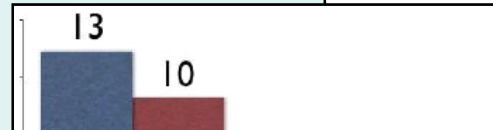
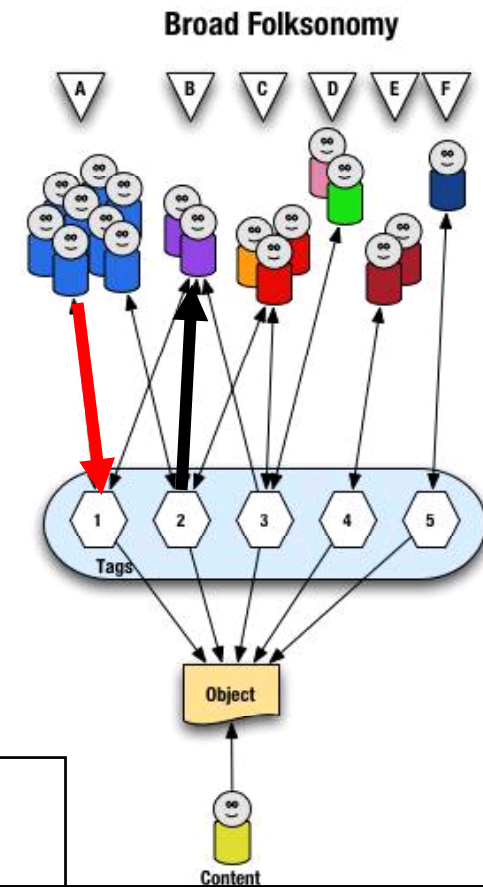


# Types of Folksonomies

[Thomas Vander Wal [http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)]

## Broad folksonomies

- many people **tagging the same object** and
- every person can **tag the object with their own tags** in their own vocabulary
- Example: Social bookmarking
- The broad folksonomy provides a means to see trends in how a broad range of people are tagging one object.
- power law curves and long-tail are relevant phenomena



The Top 20 Ways to Come Up With Amazing Ideas [save this](#) **144** people

first posted by BlogNavigator | [creativity](#) | [ideas](#) | [productivity](#) | [writing](#) | [lifhacks](#) | [tags](#)

Del.icio.us

Markus Strohmaier

Tag Tag Tag Tag Tag

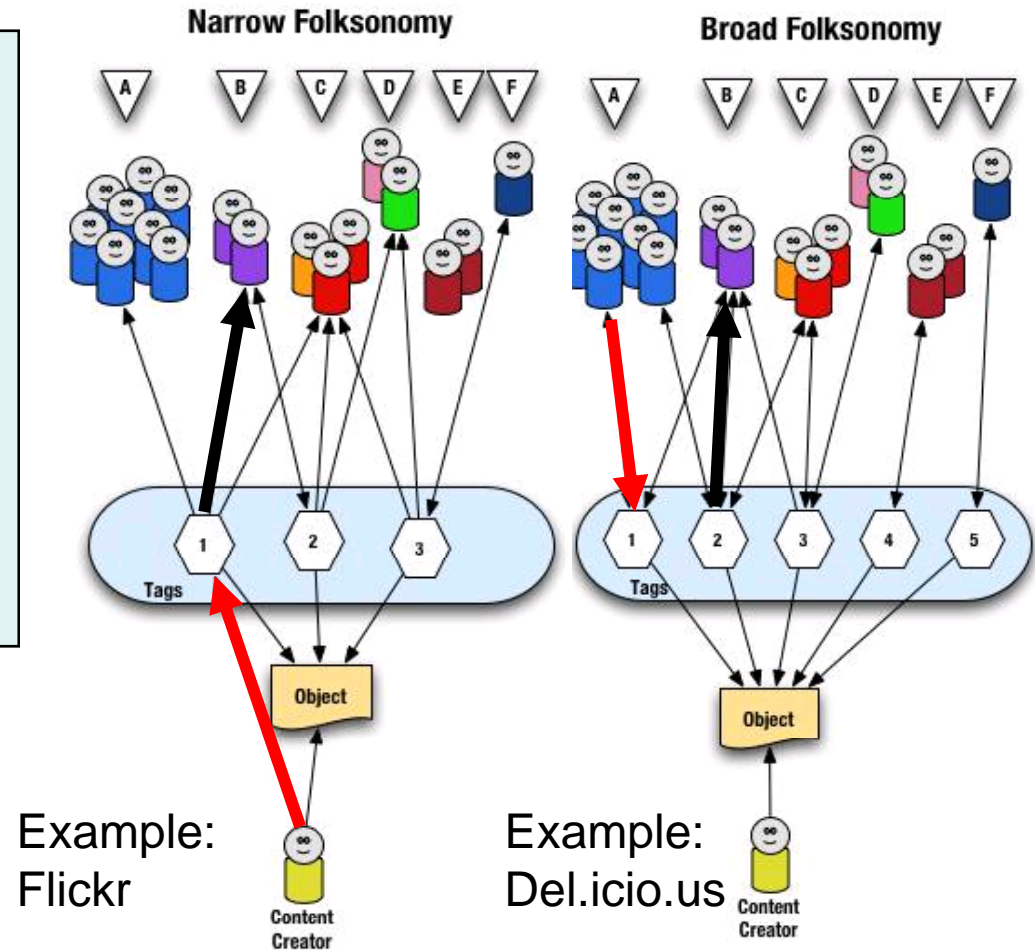


# Types of Folksonomies

[Thomas Vander Wal [http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)]

## Differences

- Number of people tagging a single object
- Narrow folksonomies are more sparse
- Purpose
- Narrow ones allow for enhanced metadata for an object



## Why do tagging systems work?

This was topic of a panel at CHI 2006, following conclusions were drawn:

### Tagging has a benefit for the user

- Similar to bookmarking, integrated apps
- Benefit of accessibility from everywhere in the internet

### Tagging allows social interaction

- Connecting a user to a community through tags
- People can subscribe your stream



# Benefits of Tagging

## Tags are useful for retrieval

- Synonyms and typos vanish in the mass of tags
- Communities can retrieve “their” stuff (e.g. by special tag)

## Tagging Systems have a low participation barrier

- Apps are easy to use, intuitive, responsive
- Free text is used to do the tagging
- Requires no previous considerations & training

# Analyzing Folksonomies

Mika P. (2004) *“Ontologies are us: A unified model of social networks and semantics”*

**How can meaning/semantics emerge from folksonomies?**

Ontologies contain instances  $I$  and concepts  $C$   
*(cf. Tag ontology consisting of [object, tags])*

## What are the fundamental constructs?

A third set besides  $C$  and  $I$  is needed

Agents  $A$  are those who specify

Agent defines

- which Concept  $C$  is
- assigned to Instance  $I$

⇒ A **tripartite model** can be identified

## A tripartite model

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics.  
International Semantic Web Conference, 522-536, Springer, 2005.

### 3 partitions: **A**, **C** & **I** (a three-mode network)

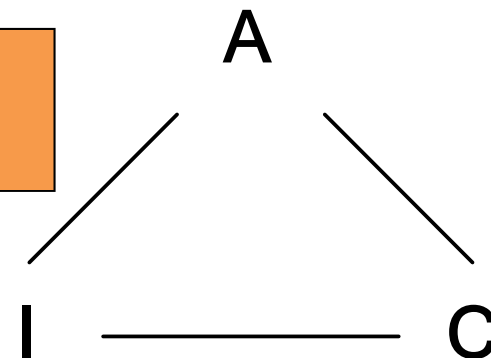
The set of vertices is partitioned into the three (possibly empty) disjoint sets  $A = \{a_1, \dots, a_k\}$ ,  $C = \{c_1, \dots, c_l\}$ ,  $I = \{i_1, \dots, i_m\}$  corresponding to the set of actors (users), the set of concepts (tags, keywords) and the set of objects annotated (bookmarks, photos etc.)

Hyperedges connect exactly one  $a \in A$  with one  $c \in C$  and  $i \in I$

Edge denotes that a user assigns a concept to a resource.

**But tripartite graphs are rather hard to understand and to work with!**

*What do you think can be done about that?*



# Folding the tripartite Model

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics.  
International Semantic Web Conference, 522-536, Springer, 2005.

Three possible two mode networks:

- A-C, C-I, A-I

Concepts are particularly interesting in the context of folksonomies  
Folding the two two-mode networks A-C, C-I into two one-mode  
networks

Co-Affiliation networks:

- ➔ Overlapping communities ( $O_{ac}$ ) and
- ➔ Overlapping sets of instances ( $O_{ci}$ )

# Folding

Folding allows to transform the Matrix to a one mode network

(also see the co-occurrence matrix in text retrieval)

$$M_P = M_{PC} * M_{PC}'$$

$$M_C = M_{PC}' * M_{PC}$$

*Commutativity!*

Result is a matrix connecting concepts to concepts

# Example: Folding

Two mode Network [excerpt]

	computer	pda	cellphone	wlan	network
i1	7	5	0	6	1
i2	7	1	1	1	2
i3	0	4	5	0	0
i4	8	0	0	0	6
i5	3	3	0	4	0

One mode Network [excerpt]

	computer	pda	cellphone	wlan	network
computer	111	62	20	62	60
pda	62	56	9	68	28
cellphone	20	9	41	0	12
wlan	62	68	0	100	24
network	60	28	12	24	34

## Other Association Matrices

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics.  
International Semantic Web Conference, 522-536, Springer, 2005

Based on A[C|I]-Graph the social network between agents can be analyzed

- Based on the AC-Graph
  - Bipartite agent to concept graph
  - Instances are used as weights
- Based on the AI-Graph
  - Bipartite agent 2 instance Graph
  - concepts are used as weights



## Broader / narrower term relations

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics.  
International Semantic Web Conference, 522-536, Springer, 2005

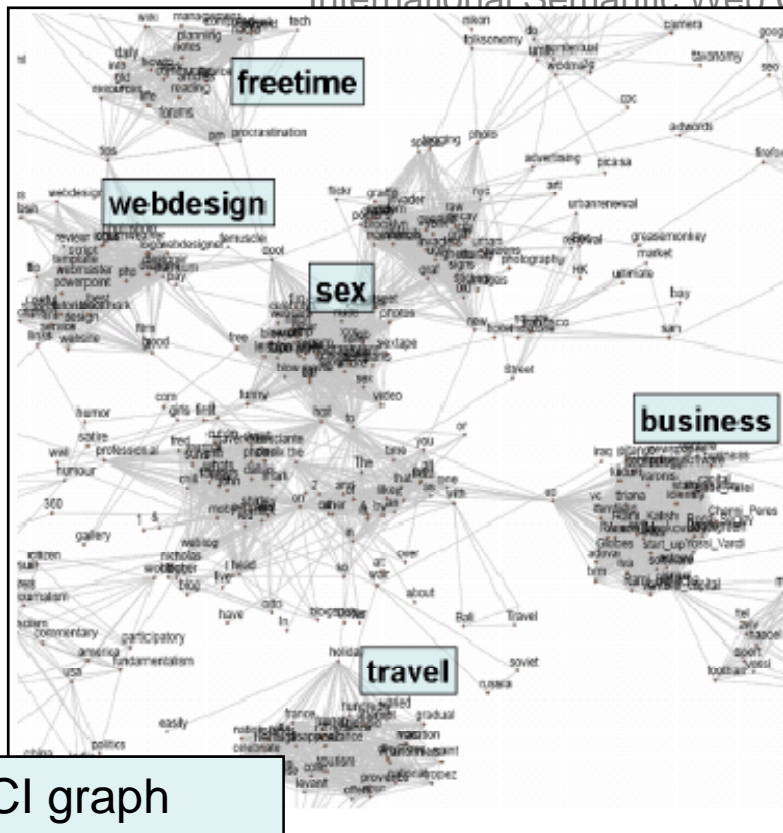
We now can think about extracting broader/narrower term relations typical of thesauri using set theory.

In an ideal situation, we would say that Concept A is a super concept of Concept B if the set of entities (persons or items) classified under B is a subset of the entities under A ( $B \in A \quad A \cap B = B$ ).

We might also add the criterium that the set of A should be significantly larger than the set of B, i.e.  $|B|/|A| < k$  for some value of k.

# Broader / narrower term relations

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. International Semantic Web Conference 522-536 Springer 2005



**Fig. 1.** The delicio.us tags associated through co-occurrence on items and the clusters emerging

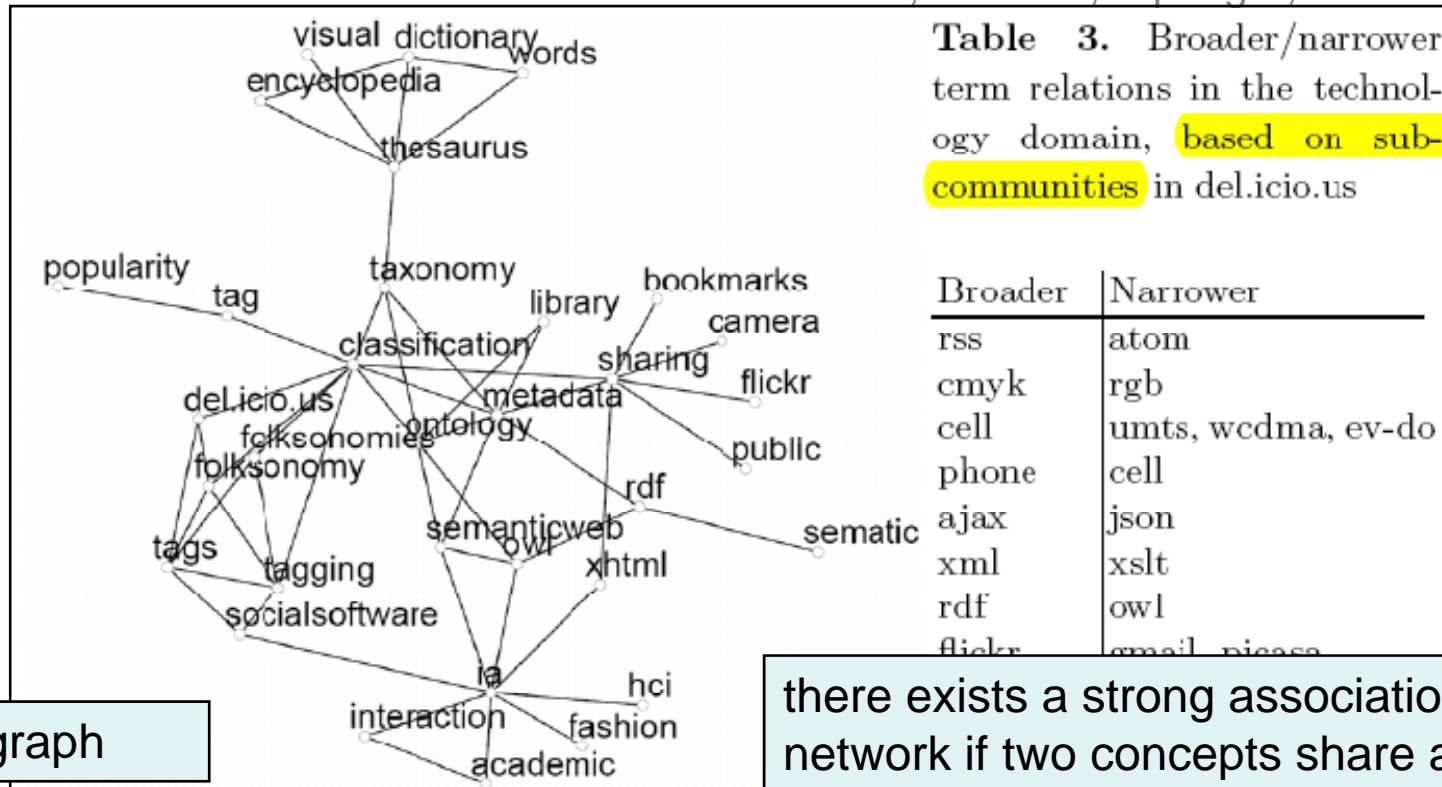
**Table 1.** The five main clusters of interest based on the **Concept-Object** network

travel	cote, provence, villa, azur, mas, holiday, vacation, tourism, france, heritage
business	venture_capital, enterprise, up, start, venture, newspaper, capital, Segev, pitango, vc
free time	procrastination, info, advice, gtd, life, notes, planning, daily, reading, forums

there exists a strong association between concepts if they share a large percentage of items, *independent of the number of users interested in them and regardless if these associations were added by the same users or not.*

# Broader / narrower term relations

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics.  
 International Semantic Web Conference, 522-536, Springer, 2005



AC graph

**Table 3.** Broader/narrower term relations in the technology domain, based on sub-communities in del.icio.us

Broader	Narrower
rss	atom
cmyk	rgb
cell	umts, wcdma, ev-do
phone	cell
ajax	json
xml	xslt
rdf	owl
flickr	email, nicasa

**Fig. 2.** Detail view of the del.icio.us tags associated through users: a 3-neighborhood of the term *ontology*. Note that the term *sematic* is correctly associated, despite the obvious typo.

there exists a strong association in the network if two concepts share a large fraction of the users among them, independent of the number of instances associated with them and regardless whether these terms were added to the same instances or not.

## Problems of the approach

Popularity vs. Generality

Tags have typos, synonyms

Tags have different intentions

- Abstract semantics (funny, sad, friendship)
- Media description (pdf, online, word, image)
- Rights and authors (persons names)
- Organizational (2read, todo, marker)
- etc.

*Faceted folksonomies,  
polyhierarchical organization of tags*

Example:

<http://www.bibsonomy.org/user/mstrohm>

## Problems of the approach

### Computational problems

- Big matrix multiplications are hard to compute

### Narrow folksonomies restrict tagging to the originating user:

- Flickr tags could historically only be assigned by the uploader
- YouTube similar restrictions

[Skip Case-Study](#)

# A Case Study

## Tag Gathering: del.icio.us

### Based on RSS feeds of del.icio.us

- Read main feed
- Get entries for each user

### Avoid spammers

- Use entries of URIs with a min. of 2 users

### Write to relational database

- In this case MySQL 5.1

# A Case Study

## Tag similarity

Tags are assigned to resources

Tags describe same URIs-> Similarity

- E.g. Javascript & Ajax
- E.g. Windows & Software
- E.g. Linux & Kernel

Tags never describe same URIs-> Dissimilarity

- E.g. Free & Shop
- E.g. Usability & SAP

# A Case Study

## Tag Merging: Objectives

Main problems within del.icio.us (and possibly in many folksonomies due to their nature)

- Synonyms
- Basic level variation

Encounter these problems by “merging” synonyms

- Different spellings: e.g. eLearning & e-Learning
- Typos & plurals



# A Case Study

## Tag Networks: Objectives

What is the conceptual structure within a community?

Which tags are similar / interconnected?

Direction of the connection?

Probability of transition for network edges?

Network Analysis?

- Hubs, central authorities
- Clusters

# A Case Study

## Tag Centrality: Objectives

Which are the most prominent nodes?

Based on different measures?

- In degree
- In Betweenness
- PageRank / HITS

The removal of central nodes would affect connectivity most!

# A Case Study

## Tag Clustering: Objectives

What are interesting conceptual clusters?

- {design, webdesign, graphics}
- {html, xhtml, css}
- {ajax, javascript, prototype, script.aculo.us}

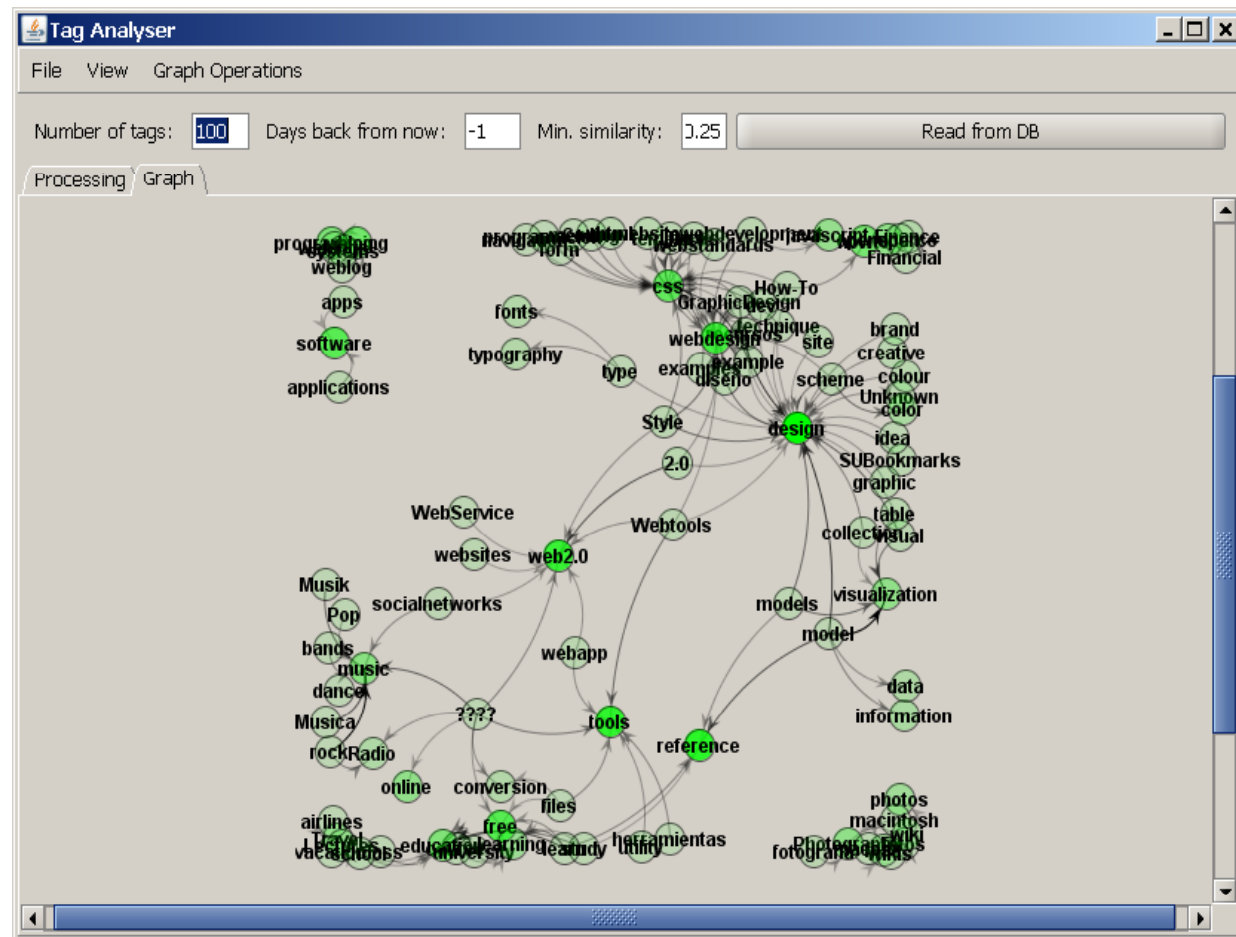
What is a meaningful disambiguation of a topic / tag?

### Clusters of tag programming

1. [systems+unix](#) (3,42)
2. [developer+development](#) (2,49)
3. [webdevelopment+javascript+webdev](#) (2,34)
4. [ebook+books+book](#) (2,19)
5. [Coding](#) (2,19)
6. [programacao+ruby](#) (2,14)
7. [script+ajax](#) (1,78)
8. [DotNet+.NET](#) (1,65)

# A Case Study

## Folksonomy Analysis Example



# Outlook

## An Experimental Goal-Tagging Social Software Application

Logged in as **mstrohm**

My Goals | My Friends | All Goals

my search

- develop firefox extensions
- do research on goals of web users
- find art related to Graz
- find books
- find restaurants in Graz
- find restaurants in graz
- find showtimes for movies in Graz
- order asian take-out food in Graz
- relax
- translate text
- understand linear algebra

---

News Ticker

13:03 *Chriskoerner* accepted your Request


17:24 Added *Mkroell* as Friend

16:52 New Entries in *Find Restaurants In Graz*

16:07 New Entries in *Understand Linear Algebra*

20:19 *Andreas* accepted your Request

Goal-based Bookmarking Community



The Goal-Cloud

To Join the community please Register using the sidebar

This community provides about 33 links for 28 goals

find showtimes for movies in Graz understand linear algebra include AJAX framework find restaurants in Graz translate text read rss feeds get information about tugraz create Firefox Sidebars find books do research on goals of web users read rss feeds online Understand Mozilla Applications go skiing search for publications find information about favelets order asian take-out food in Graz organise my bookmarks buy tickets for events in Austria find lolcat pictures Get to know SIMBA get coding related news buy dvds on the internet find a Job in Graz develop firefox extensions relax buy books on the internet get recommendations for music find art related to Graz

Haselsberger Andreas; Ruggenthaler Christoph; [Univ. Ass. Strohmaier Markus](#) @ [kmi.tugraz.at](mailto:kmi.tugraz.at) [Get our Firefox Plugin!](#)

Any further questions?