Folksonomy and Information Retrieval

Isabella Peters

Heinrich-Heine-University Düsseldorf, Dept. for Information Science, Universitätsstr. 1, D-40225 Düsseldorf, Germany peters@phil-fak.uni-duesseldorf.de

Wolfgang G. Stock (corresponding author)

Heinrich-Heine-University Düsseldorf, Dept. for Information Science, Universitätsstr. 1, D-40225 Düsseldorf, Germany stock@phil-fak.uni-duesseldorf.de

In Web 2.0 services "prosumers" - producers and consumers - collaborate not only for the purpose of creating content, but to index these pieces of information as well. Folksonomies permit actors to describe documents with subject headings, "tags", without regarding any rules. Apart from a lot of benefits folksonomies have many shortcomings (e.g., lack of precision). In order to solve some of the problems we propose interpreting tags as natural language terms. Accordingly, we can introduce methods of NLP to solve the tags' linguistic problems. Additionally we present criteria for tagged documents to create a ranking by relevance (tag distribution, collaboration and actor-based aspects). We would like to open the discussion concerning the following aspects: Which tag distributions seem to be characteristic for folksonomies and how can we use these distributions effectively in information retrieval? What are the problems of indexing by using tags, especially regarding indexing photos and videos? How may we use factors of collaborative indexing for relevance ranking?

Information Indexing Using "Collective Intelligence" in Web 2.0 Services

In the early years of the World Wide Web only few experts were able to distribute information via this new medium. The majority of the people dealing with the WWW solely acted as passive users. Yet, since the beginning of the 21st century some services are arising which are easy to handle and allow users to publish content on the Web in an uncomplicated way. The former (passive) user now becomes additionally a (active) Web author. The consumer of knowledge turns into a knowledge producer, a "prosumer" according to Toffler (1980). As the authors / prosumers edit and correct the documents reciprocally, we can speak of "collective intelligence" (Weiss, 2005, p. 16) in this context:

> **With content derived primarily by community contribution, popular and influential services like Flickr and Wikipedia represent the emergence of "collective intelligence" as the new driving force behind the evolution of the Internet.**

Figure 1: Tag cloud referring to the homepage of the book called "Web Information Retrieval" by Dirk

Lewandowski. Source:Del.icio.us.

"Collective intelligence" arises from joint efforts of a group of authors or users in so-called "collaborative services". Such services can be summarized under the tag "Web 2.0" (O'Reilly, 2005). They offer possibilities for keeping or searching diaries (Weblogs, Technorati), for the construction of encyclopedias (Wikipedia) and the management of bookmarks (Del.icio.us), photos (Flickr) or videos (YouTube). Sometimes several complementary services are combined and build new "mash-ups" (Weiss, 2005, p. 23), e.g. Housingmaps.com, a mash-up of real estate information from Craigslist and maps and satellite's photos from Google Maps. The collaboration does not stop with providing content but includes indexing of provided knowledge in some Web 2.0 services as well. Figure 1 shows a "tag cloud" of words that have been used to describe bookmarks and the Website's content respectively. The word frequency is represented with different font sizes. A lot of people have tagged this Website with *informationswissenschaft*, *searchengine*, **and** *Web*, **tags like** *e-book* **or** *german*, *information_retrieval*, *recherche*, *retrieval*, **and** *suchmaschine* **are less frequent, only few people tag the Website with** *bibliothekswissenschaft* **or** *informationsmanagement*.

An annotation system open for users to apply subject headings is called "folksonomy", the freely chosen subject headings are called "tags". The process of indexing by means of folksonomies is named "(social) tagging". Peter Merholz (2004) entitles this method as "metadata for the masses", the writer James Surowiecki calls it "the wisdom of the crowds" (McFedries, 2006, p. 80). The term "folksonomy", as a combination of "folk" and "taxonomy", was introduced by Thomas Vander Wal and cited in a blog post by Gene Smith (2004):

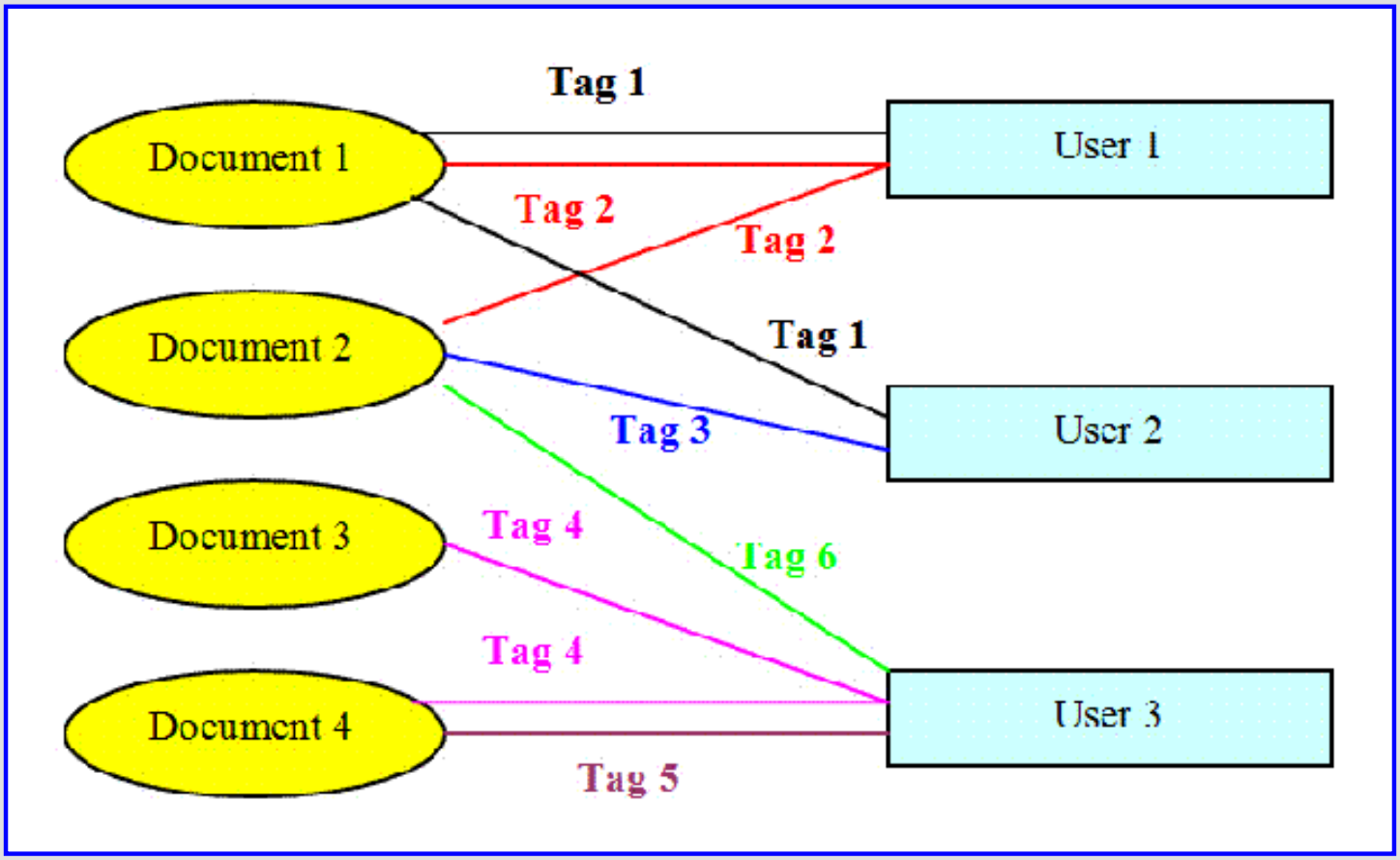
> **Last week I asked the AIfIA (i.e. the "Asilomar Institute for Information Architecture", the authors) member's list what they thought about the social classification happening at Furl, Flickr and Del.icio.us. In each of these systems people classify their pictures/bookmarks/web pages with tags ..., and then the most popular tags float on the top ...**
>
> **Thomas Vander Wal, in his reply, coined the great name for these informal social categories:** *a folksonomy*.
>
> **Still, the idea of socially constructed classification schemes (with no input from an information architect) is interesting. Maybe one of these services will manage to build a social thesaurus.**

Smith uses the term "classification" for paraphrasing folksonomies. This term arouses a misleading and wrong connotation. The same holds for the term "taxonomy". Folksonomies *are not* **classifications, because they neither work with notations nor relations. What is important for us, is Smith' hint that folksonomies can be used to build up thesauri collaboratively.**

In folksonomies we are confronted with three different aspects (Marlow, Naaman, Boyd, & Davis, 2006):

- **the documents to be described,**
- **the tags (words), which are used for description,**
- **the users (prosumers), who are indexing.**

Figure 2: Documents, tags and users in a folksonomy.



Users as well as documents are interconnected with each other in a social network environment in which the paths run along the tags. On the one hand, documents are linked *thematically* with each other when they have been indexed with the same tags. In figure 2 the documents 1 and 2 as well as 3 and 4 are thematically connected (documents 1 and 2 with tag 2; documents 3 and 4 with tag 4). Additionally, documents are related via users, so-called *shared users*. So the documents 1 and 2, 3 and 4 are linked as well as 2 and 4, considering the relation to the users.

On the other hand, users are linked with each other when they use the same tags for indexing or when they index the same documents. They are *thematically* related when they index with the same tags (see in the example user 1 and 2 with the tags 1 and 2); they are coupled with *shared documents* when users describe their content (users 1, 2 and 3 with document 2). The extent of commonality may be illustrated quantitatively with similarity rates like Cosinus, Jaccard-Sneath or Dice (Stock, 2007, p. 177).

Occasionally documents may be indexed with multiple tags which have been indexed with varying frequency. Document 1, for example, is annotated with two different tags, but one of them (tag 1) has been allocated twice. Depending on how often the tags are attached to a document we may gain document-specific tag distributions. Analogously it is possible to determine user-specific tag distributions (Huang, 2006).

Broad and Narrow Folksonomies

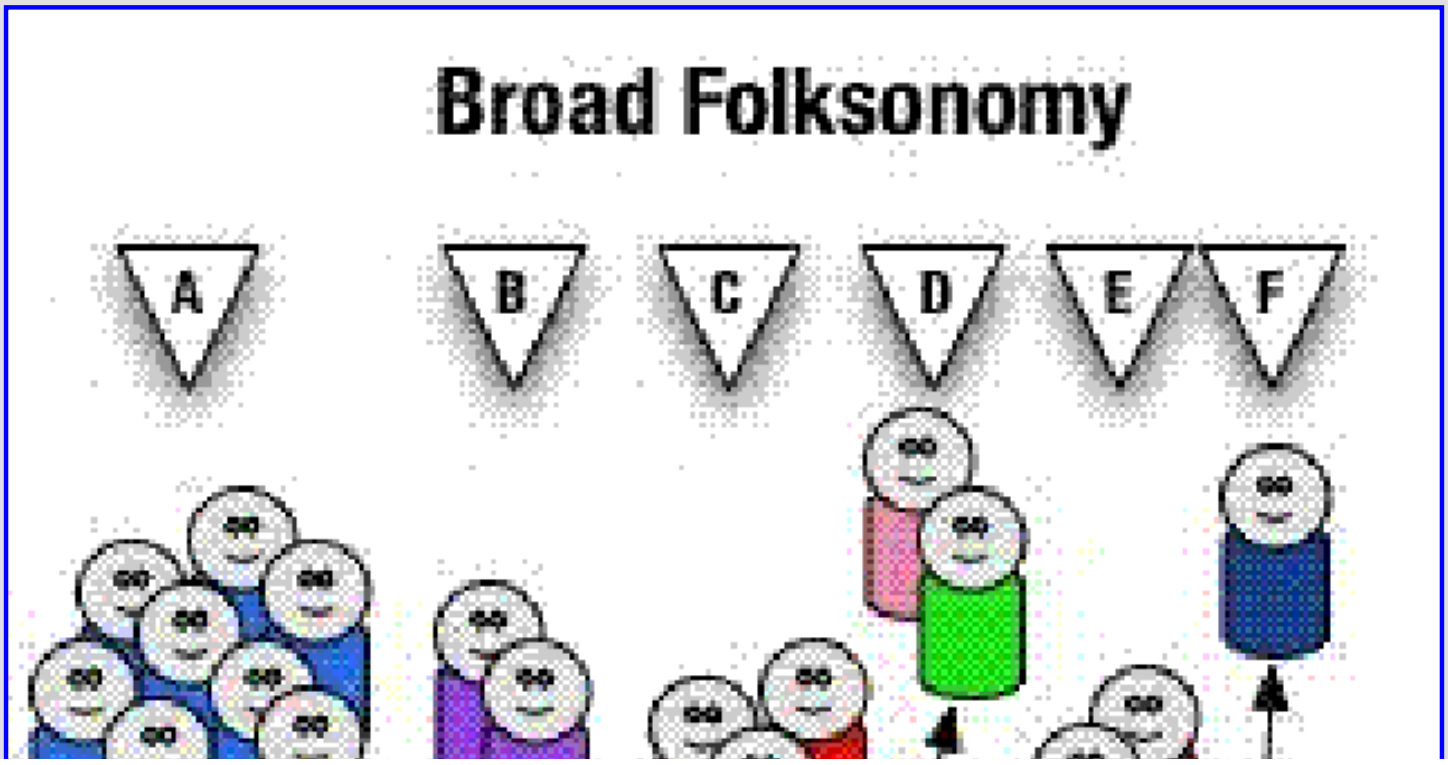
Folksonomy is a method of indexing documents on the Web using uncontrolled terms (Furnas et al., 2006; Peters, 2007). The producers or the users of the documents themselves become indexers. There is no authority which controls the terminology and the prosumer-indexers. Folksonomies find broad application in Web 2.0 services, but it is also possible to work with folksonomies in intranets of companies (Fichter, 2006), e.g. for indexing corporate blogs, podcasts and vodcasts (Peters, 2006a; Peters, 2006b; Peters & Stock, 2006), corporate bookmarking services (Millen, Feinberg, & Kerr, 2006) and message boards (Murison, 2005). Commercial online information suppliers (e.g., Engineering Village) can work with folksonomies as well. Spiteri (2006) discusses the use of folksonomies in public library catalogues.

Referring to Thomas Vander Wal (2005) there are two types of folksonomies: broad and narrow folksonomies (see also Dye, 2006). In a broad folksonomy (fig. 3) many different people (A to F in the figure) may tag the same document (called "object" by Vander Wal). The people describe the document's content from their point of view by using the same, similar or completely different tags (1 to 5). A well known example of a service using this sort of folksonomy is the social bookmarking tool Del.icio.us (for social bookmarking see Hammond, Hannay, Lund, & Scott, 2005; Lund, Hammond, Flack, & Hannay, 2005). Vander Wal (2005), Shirky (2005) and others state that the distribution of tags given to a document follows a Lotka-like power law (Egghe & Rousseau, 1990, p. 293). Such a distribution has the form

$f(x) = C / x^a$,

where C is a constant, x is the rank of the tag relative to the document, and a is a value ranging normally from about 1 to about 2. If this assumption is true, we see a curve with only few tags at the top of the distribution, and a "long tail" of numerous tags on the lower ranks on the right-hand side of the curve.

Figure 3: Folksonomy with multiple tag application ("broad folksonomy").Source: Vander Wal (2005).



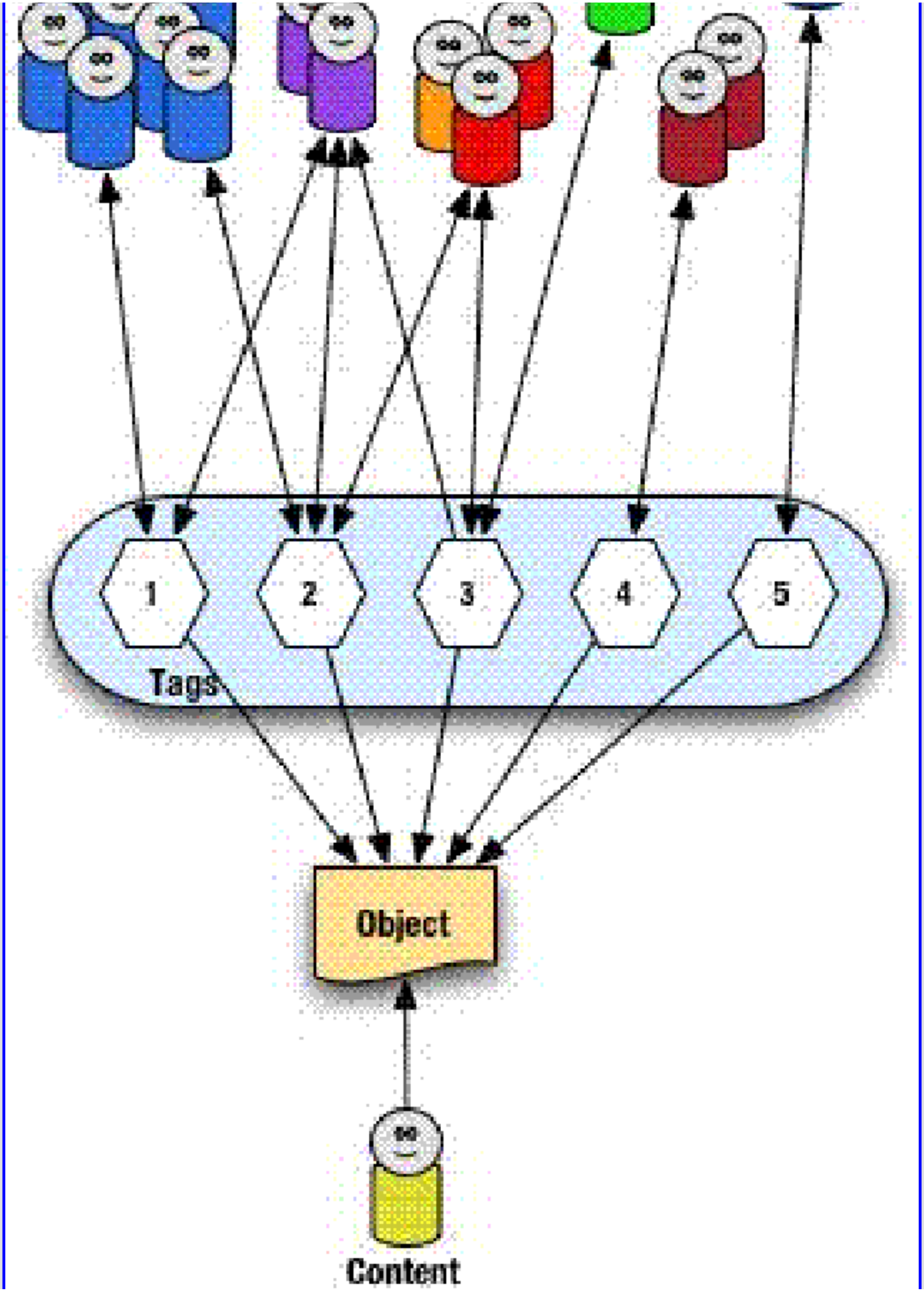
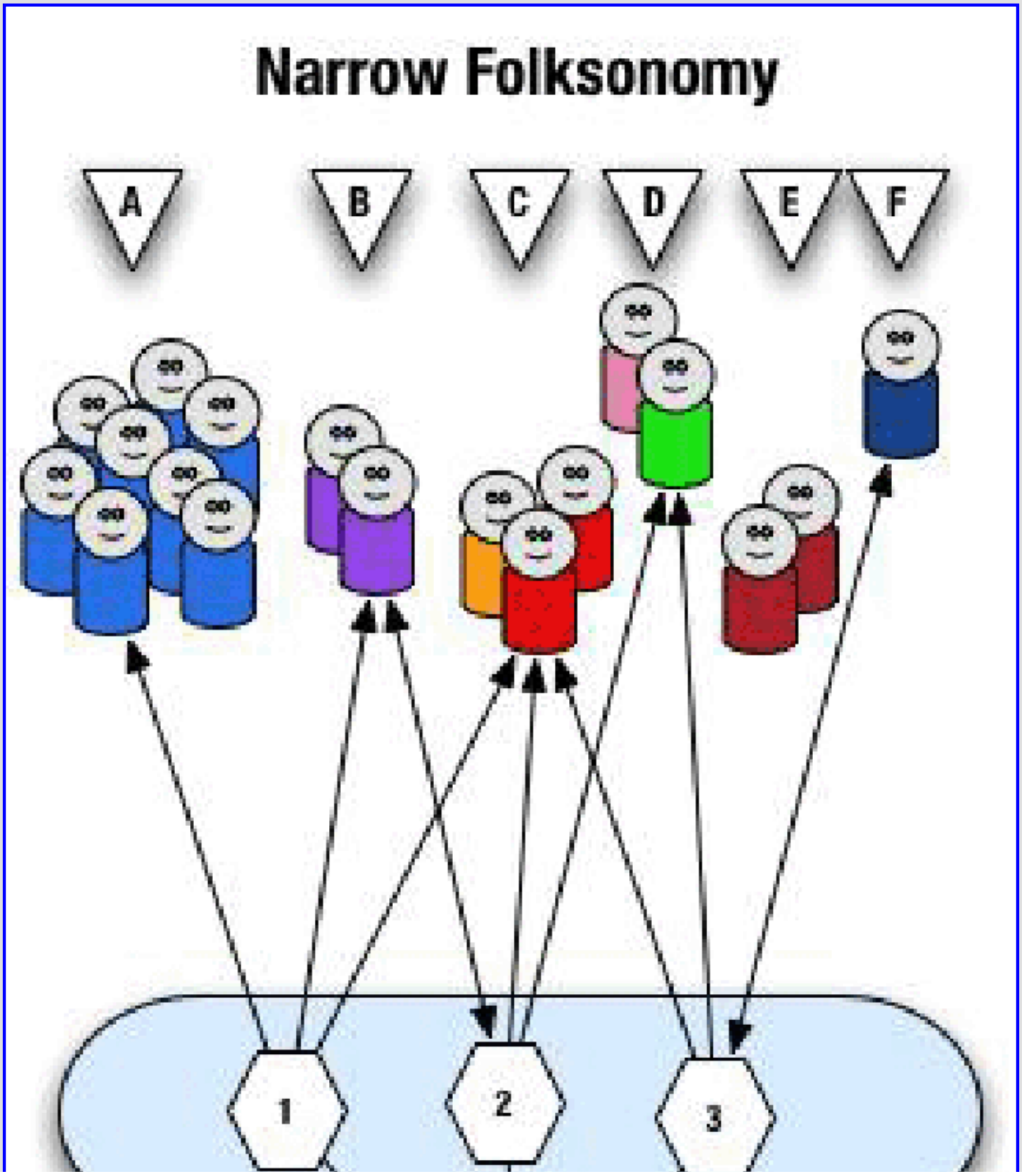
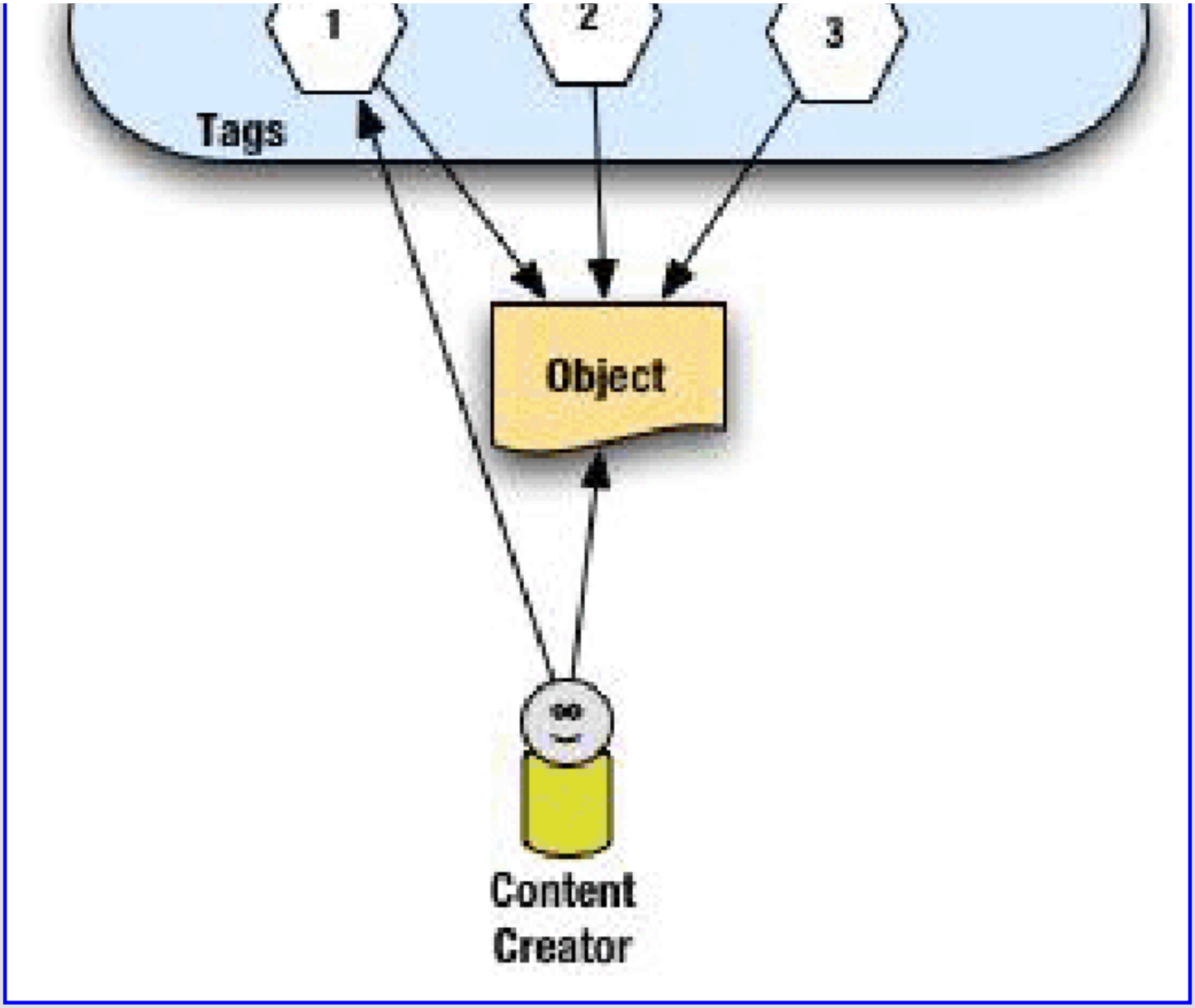

1
2
3
4
5
Tags
Object
Content

Figure 4: Folksonomy with single tag application ("narrow folksonomy").Source: Vander Wal (2005).

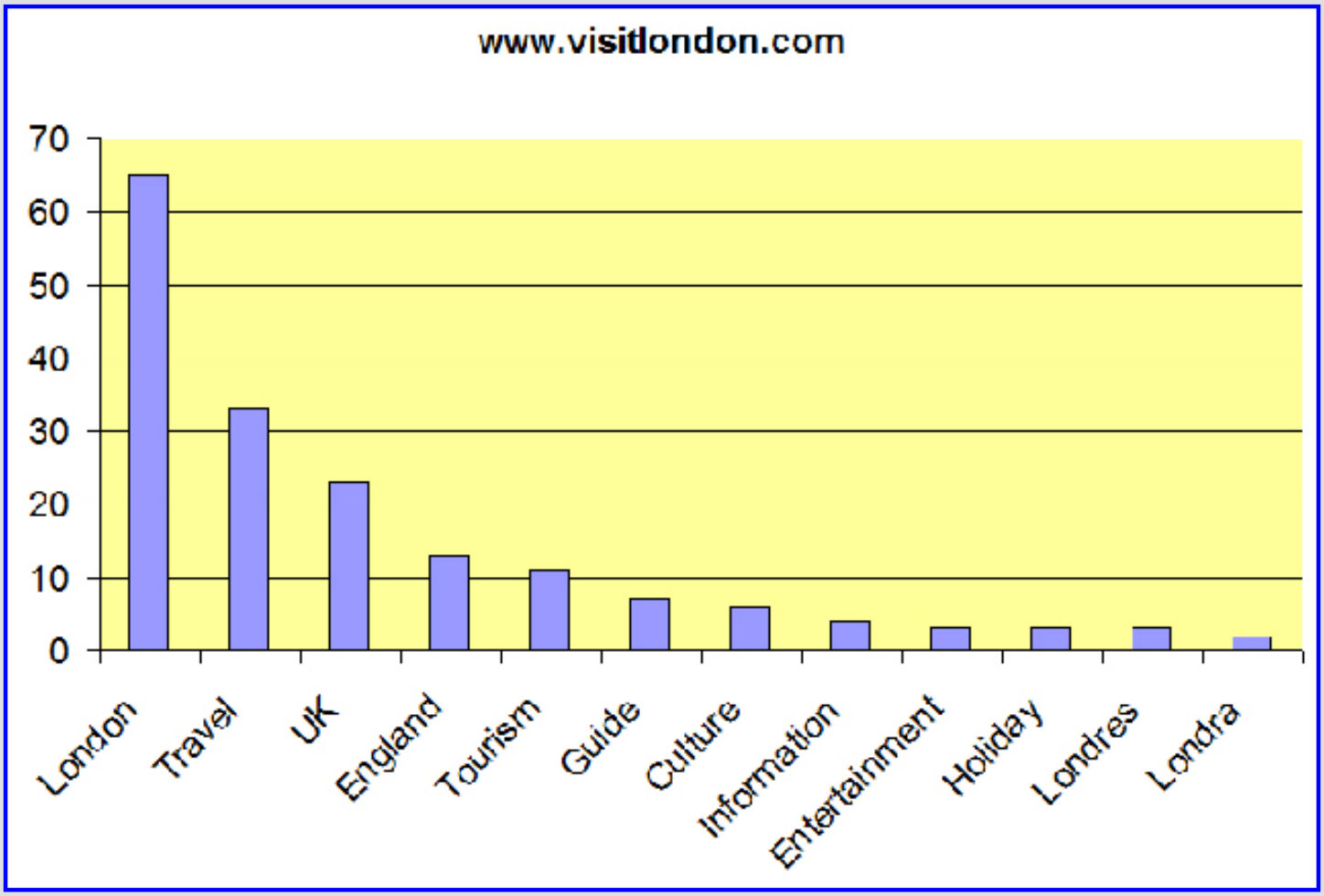


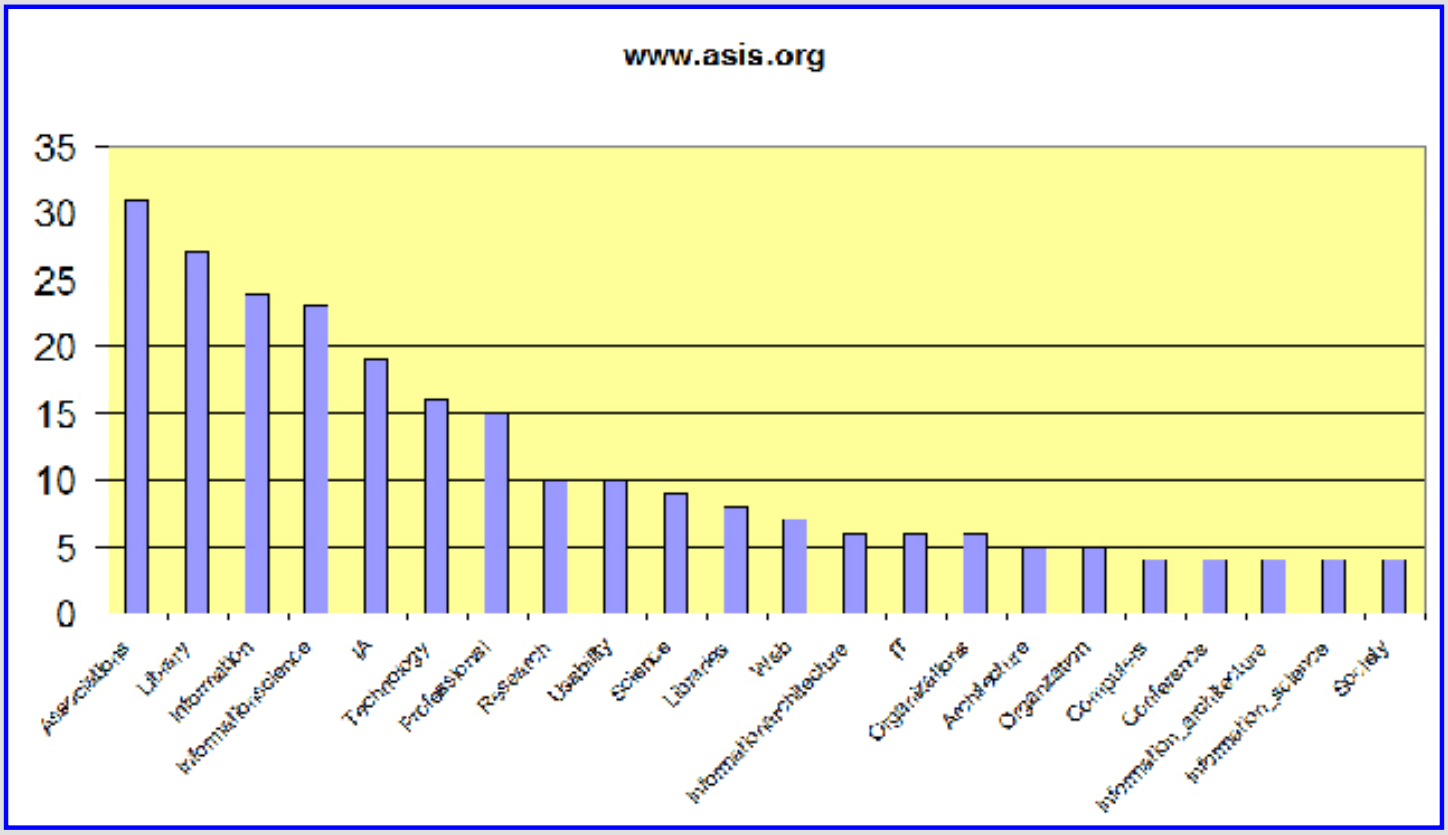


In narrow folksonomies, tags for a documented object are generally recorded just once. So only new tags can be applied and there is no possibility to count tag frequencies. Usually the documents' author (or content creator) provides the tags; occasionally other users are also allowed to add tags. This approach reminds of the known procedure in professional indexing with controlled terms of thesauri, classification systems and ontologies (of course, while using folksonomies the terms are not controlled). Collaborative Web services like Flickr (photos), Technorati (blog posts) or YouTube (videos) are working with narrow folksonomies. There are as well some interesting and fruitful patent applications in the technological area of narrow folksonomies (e.g., Butterfield et al. 2006; Butterfield, Schachter et al.: Yahoo!/Flickr; Ott, 2006: Yahoo! Video).

Narrow folksonomies cannot depict a special distribution of tags, because all tags are ranked equally. What is possible is to examine, whether a user finds the document by searching with tag 1, 2 or 3. According to this, a ranking of search-tags retrieving the document can be created. Compared to broad folksonomies, narrow ones differ in the number of assigned tags, the number is probably less. That is why it is not possible that a long tail arises. But the search-tags may form a specific distribution as well.

Figure 5: Different tag distributions. Source: Del.icio.us.



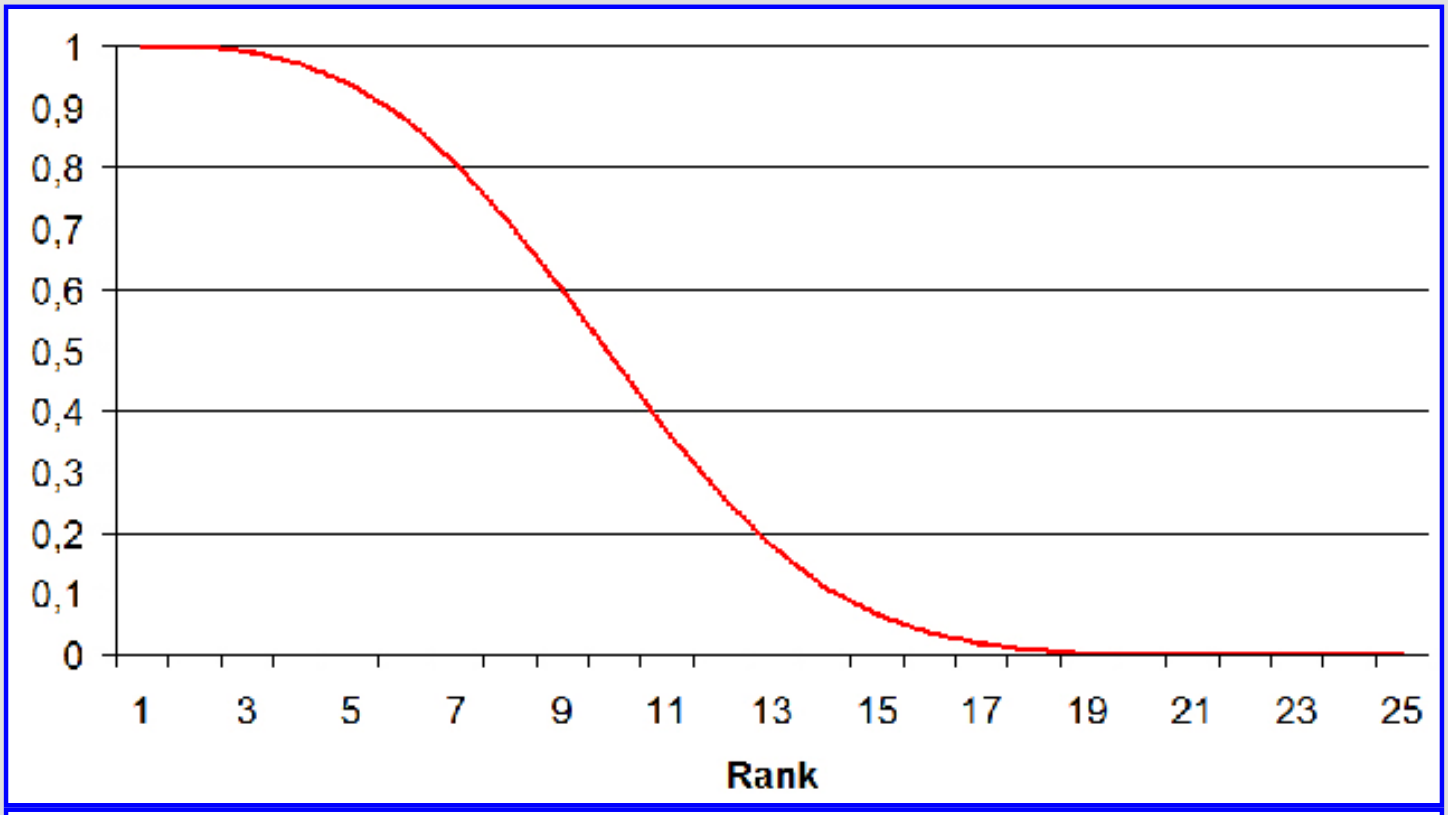


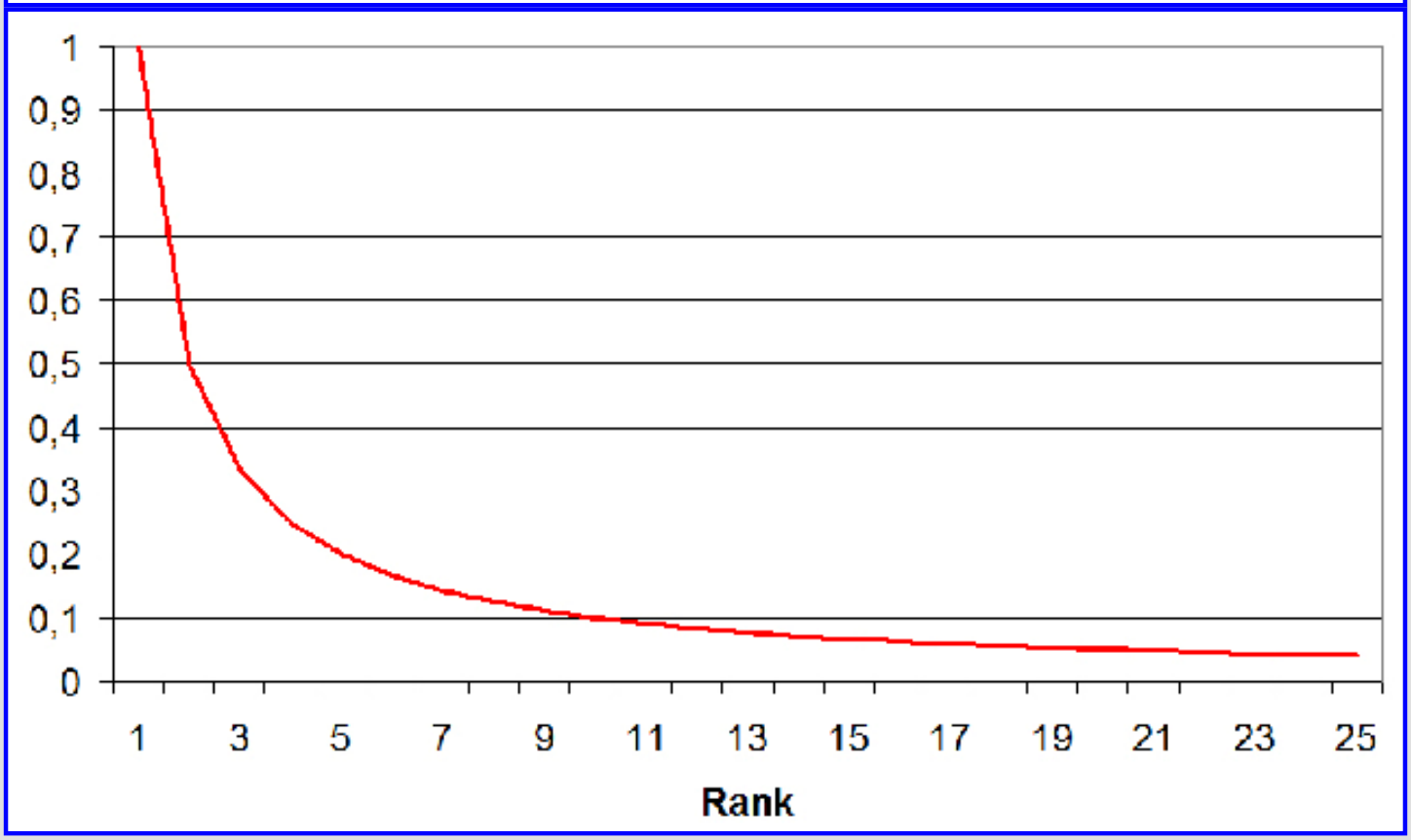
Distributions of Tags

How does a folksonomy work in practice? Let us have a look at two tag clouds (fig. 5)! The first example is a ranked list of tags referring to the Website www.visitlondon.com found in Del.icio.us. This site is tagged with *London* by more than 60 users, with *travel* by about 30 users, and with *UK* by more than 20 users. On the right-hand side we see a long tail of tags like *England*, *tourism*, *guide*, *culture* and so on. After visiting the Website www.visitlondon.com you can judge that these tags match well and describe the document's content properly. As you can see in figure 5 the tags are distributed in a power law style with an exponent of a = 1.

The second example is quite different. It is a ranked list of tags referring to the homepage of the American Society for Information Science and Technology, also retrieved from Del.icio.us. The Website www.asis.org has been annotated about 15 to 30 times with each of the tags *associations*, *library*, *information*, *informationscience*, *IA*, *technology* and *professional*. Here you can see that the distribution does definitely not follow the distribution of a power law. There is a long tail (*research*, *usability*, *science*, *libraries* and so on) on the right-hand side, but it seems that there is a second long tail on the left-hand side. These long tail's tags describe the site's content only more or less appropriately. In this case, the users' collective intelligence seems to have difficulties giving a "true" picture of ASIST's Website.

Figure 6: Possible relevance distributions. Source: Stock (2006), p. 1127 (modified).





In a theoretical study Stock (2006) could show, that there are (at least) two different relevance distributions: the well known power law ("informetric") and a second kind of distribution, called "inverse logistic" (fig. 6). The inverse logistic distribution (upper diagram) shows a lot of relevant tags (the "long trunk") and the known "long tail". This distribution follows the formula

$f(x) = e^{-C'(x-1)^b}$

(e is the Euler number, x is the rank of the tag, C' is a constant and the exponent b is approximately 3. In our example b is 3 and C' is 0.1). In most cases the "long trunk" will be shorter than the "long tail".

Regarding the distribution of tags, particularly two types of curves have to be kept in mind: the power law distribution and the inverse logistic distribution. (Perhaps there are further types of distributions as well). This consideration becomes important as we come along to develop and create information retrieval tools for tagged documents.

Some services provide informetric analyses of tags, e.g. time series of blog postings per day for a given topic (Technorati; for an example see Stock & Weber, 2006, p. 387) or time series of tags to a given URL in Del.icio.us (Cloudalicious; Russell, 2006).

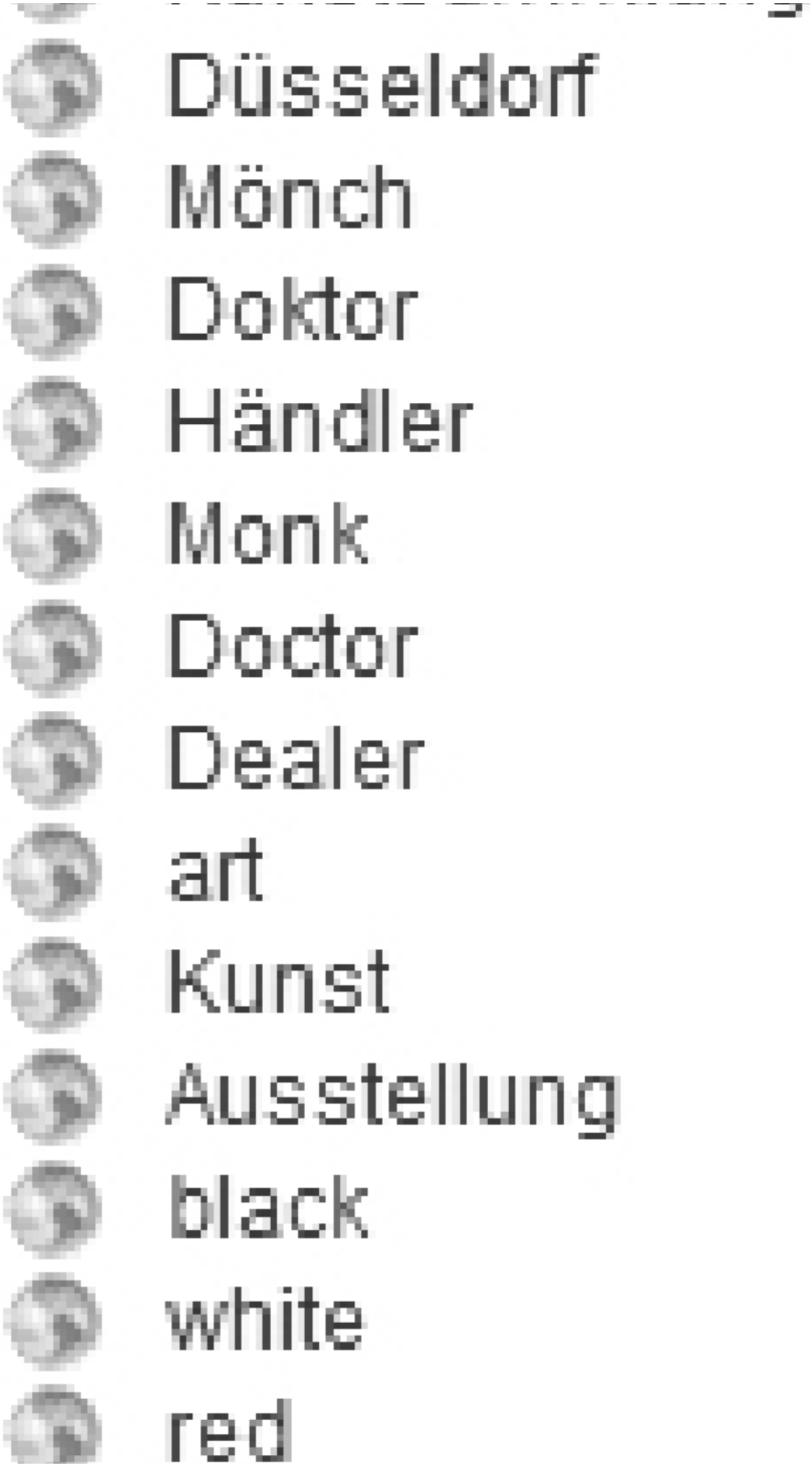
Ofness – Aboutness – Iconology

The topics given in a document form its aboutness (Bruza, Song & Wong, 2000; Hutchins, 1978; Ingwersen, 2002; Maron, 1977). Some Web 2.0 services deal with images (e.g., Flickr) or videos (e.g., YouTube). This non-textual information (as well as some fictional texts) consists not only of one single level of aboutness as, e.g., scientific articles. According to Erwin Panofsky (2006) there are three different semantic levels of interpretation for artwork (Smith, 2006, p. 9). We will explain Panofsky's theory by giving an example of a photo found in Flickr (fig. 7). The photograph shows a group of sculptures (called "Monk – Doctor – Dealer", in German "Mönch – Doktor – Händler") created by the German Artist Katharina Fritsch. The sculptures are located in Düsseldorf's museum *K21*. Additionally we present the tags which the photographer has attached to his photo.

Figure 7: Tagged photo found in Flickr. Source: Flickr.com.

Tags

- KatharinaFritsch
- K21
- Kunstsammlung
- Düsseldorf


Düsseldorf
Mönch
Doktor
Händler
Monk
Doctor
Dealer
art
Kunst
Ausstellung
black
white
red

Panofsky's basic semantic level, called "pre-iconographic", is the world of primary or natural objects. Following the pre-iconographic approach the photo shows three figures: a bright-red one with a cloven hoof, a white one with a death's-head and a black one wearing a habit. It is not necessary that the interpreter has more than everyday-life experience to interpret the photograph. The literature in information science calls the content on the pre-iconographic level the "ofness" of the document (Lancaster, 2003, p. 13; Layne, 2002; Markey, 1986; Shatfort, 1986; Turner, 1995). Tags describing the ofness are *black*, *white* **and** *red* **in our example. Level 2, the "iconographic" level, is the field of secondary objects – it is the world of "pictures, anecdotes and allegories" (Panofsky, 1975, p. 50). For interpreting this level the interpreter must have some cultural and social experience in the topical field of the artwork. This level may be called the "aboutness" of the document. Tags describing the aboutness are** *monk*, *doctor* **and** *dealer* **in the example. Level 3 relates to the "iconological" level of the "true content, forming the world of 'symbolic values'" (Panofsky, 1975, p. 50). The interpreter has to be an expert with excellent experiences in the topical field to gain the intention, e.g., an expert in history of arts. Information professionals normally do not describe level 3 with index terms and abstracts, but end users are free to do so. In our example, there is only one (very weak) level-3-tag:** *art***. More useful tags may be** *unholy trinity* **or** *masterpiece of German sculpture of the beginning 21st century* **which could be freely added. Additionally, users may tag non-content aspects, for example the name of the artist (***KatharinaFritsch***) or of the museum (e.g.,** *K21***). Such tags describe aspects of "isness" in the sense of Peter Ingwersen (2002, p. 293-294). Katharina Fritsch** *is* **the creator of the sculpture; K21** *is* **the museum which it exhibits. Folksonomies do not make a distinction between the levels of description and interpretation. Tags from all levels melt into only one semantic level.**

Ingwersen (2002, p. 289) distinguishes between author aboutness (the content as it is), indexer aboutness (the interpretation of content with a purpose), request aboutness (content as expressed by the request) and user aboutness, i.e. "*user interpretation* **of objects". Tagging of images, videos and other non-textual information is an expression of the last kind; it refers to user-ofness (level 1), user-aboutness (level 2), user-iconology (level 3) and isness, but – and that is a problem – the levels are indistinguishably mixed in one single tag cloud. But, besides the problems Smith (2006, p. 13) concludes for tagging museum objects, "Tagging has the potential of increasing access to artwork images and records dramatically for searchers of all levels of expertise". Studies about implementing new documentation methods in museums show that most of the user-created tags were new (Trant, 2006b, p. 21) and form new possibilities to find an artwork (Trant, 2006a).**

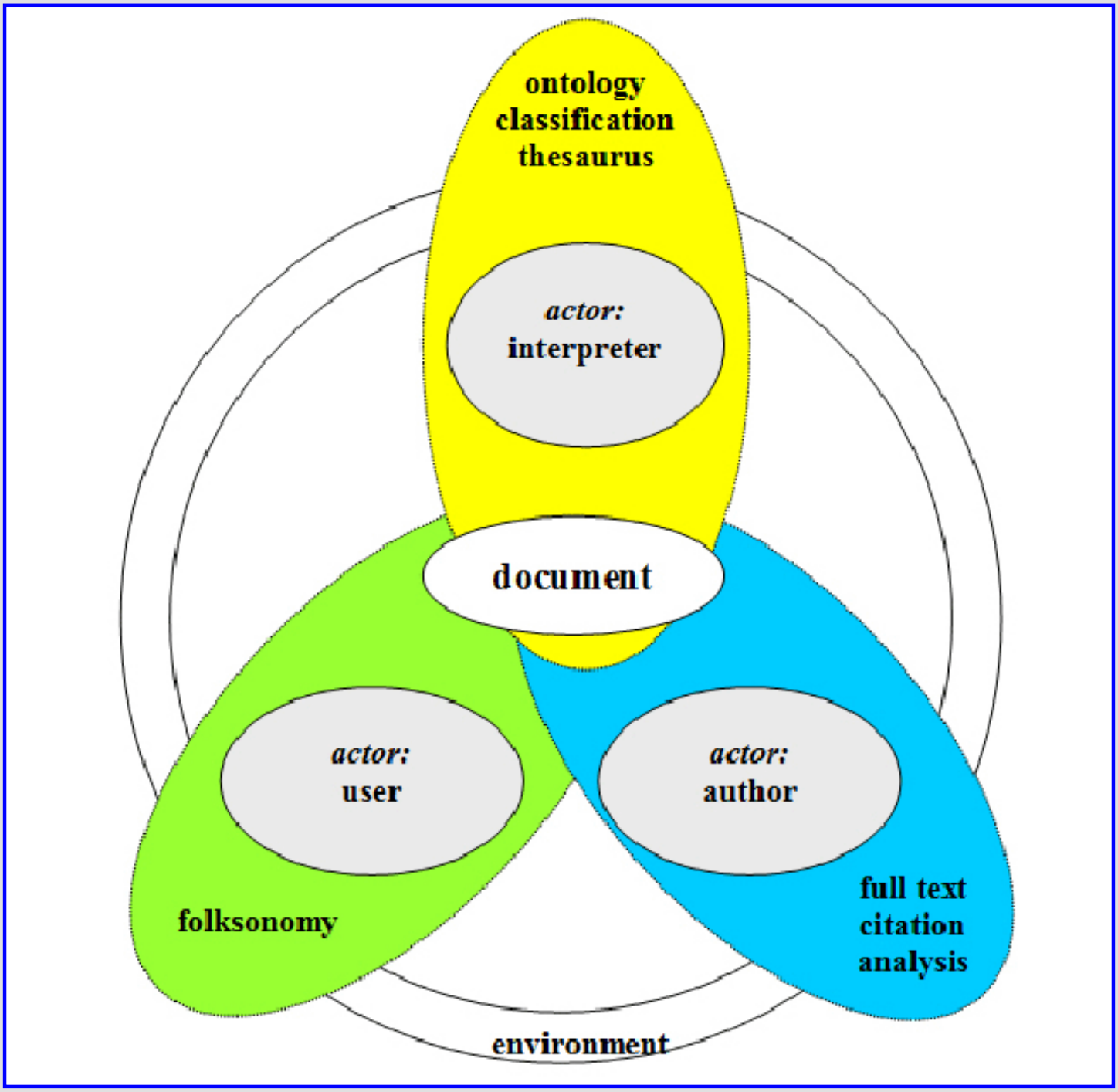
There is a huge amount of literature about indexing images in information science (Rasmussen, 1997), yet the studies are not focused on folksonomy-like tagging, but on, e.g., indexing newspaper images. Ornager (1995, p. 214) deduced the hypothesis that the following five categories represent the minimum criteria for

subject analysis and indexing of photos: "named person (who), background information about the photo (when, where), specific events (what), moods and emotions (shown or expressed), size of photo". The results found by Markkula and Sormunen (2000, p. 270-271) highlight the indexing of concrete objects: "The most often used index terms referred to specifics, i.e. to individual objects, places, events and linear time, and to the theme of the photo". Indexers very often seem to use nouns to describe those objects, but when they try to index events and actions, they use verbs as well ("Czechs *celebrate* the victory ..."; Markkula & Sornumen, 2000, p. 271). Their study shows that the output of the indexing process seemed to be quite inconsistent (Markkula & Sornumen, 2000, p. 273).

Folksonomy: The Benefits

Folksonomies are a new method of describing a document's content. Here the prosumers – or the "actors" in the sense of cognitive work analysis (Fidel, 2006) – index the documents by themselves. Tagging reflects the prosumers' conceptual model of information (Quintarelli, 2005) and tags authentically represent the language of authors and users. This sort of indexing leads to "multiple interpretations", different (and sometimes disparate) opinions and "multicultural views" of the same piece of information (Peterson, 2006). "Shared intersubjectivities" enable the users "to benefit, not just from their own discoveries, but from those of others" (Campbell, 2006, p. 10).

Figure 8: Folksonomy – ontology – text-oriented methods in different contexts.



Folksonomies complete the tools and methods of information indexing (see fig. 8). There are three groups of actors who are able to index documents: authors, professional indexers and users (Kipp, 2006b). Probably all of them may use different ways for indexing and along with it may focus on different characteristics. Text-oriented methods make use of the author's language, e.g. in forms of indexing titles, abstracts or references (Garfield, 1979). In contrast to text-orientated methods folksonomies do not only represent the producer's view, but the views of the consumers as well. Ontologies and other tools of controlled vocabularies (like thesauri or classification systems) are in need of interpreters: a) experts who create such vocabularies and b) other experts who are able to use the controlled terms in order to index the documents. Ontology-creating interpreters have to analyze „literature, needs, actors, tasks, domain, activities, etc." (Mai, 2006, p. 17) – undoubtedly a time-consuming procedure. What is more, this system is

very expensive. In comparison, folksonomies are cheap in practice because the indexing is done by volunteers in a collaborative way. "Tagging has dramatically lower costs because there is no complicated, hierarchically organized nomenclature to learn. Users simply create and apply tags on the fly" (Wu, Zubair, & Maly, 2006, p. 111). Lots of people are willing to add tags. The feedback of the community helps to reinforce the process and encourages more tagging (Fichter, 2006). For practical reasons, it seems to be impossible to index all URLs, blog posts, images, and videos on the Web intellectually by using controlled vocabularies. Textual documents can be indexed automatically like all popular search engines (Google, Yahoo!) do. But there is not yet an applicable (content-based) automatic method for indexing non-textual documents (images, videos, music, etc.) because such documents do not contain metadata.

The development and updating of controlled vocabularies can profit from folksonomies (Aurnhammer, Hanappe, & Steels, 2006; Christiaens, 2006; Gendarmi & Lanubile, 2006; Macgregor & McCulloch, 2006; Mika, 2005; Spyns, de Moor, Vandenbussche, & Meersman, 2006; Zhang, Wu, & Yu, 2006), because the tags, their frequency and their distribution are sources for new controlled terms, for modifications of terms and perhaps for deleting concepts in the sense of a "bottom-up categorization" (Vander Wal, 2004). In this way tags guarantee a fast response to changes and innovations in the knowledge domain.

Are folksonomies and ontologies opposing ideologies? For Tom Gruber (2006, 994) this "is nonsense, and it is time to embrace a unified view". How can folksonomies help to develop ontologies or the semantic Web in practice? So far applied experimental methods work with co-occurrences (Schmitz, 2006) or with the vector space model (dimensions: documents, vectors: tags; tag-similarity: cosine) (Heymann & Garcia-Molina, 2006). By means of the developed similarity values it is possible to create a similarity graph. The position of the tag within the graph allows conclusions concerning its hierarchical location in a "latent hierarchical taxonomy" (Heymann & Garcia-Molina, 2006, p. 4).

If a journalist or a scientist wants to publish an article, he has to pass gate keepers who control the content's quality.Are there such gate keepers or editors on the Web as well?Clay Shirky (2005) states, "the Web has an editor, it's everybody". The quality control happens after publication.The more people tag or comment a document, the more relevance does this document seem to have for this people.

Of course, users can search by entering tags, but additionally they can browse the system following the related tags. For Adam Mathes (2004) the primary strength of folksonomies is serendipity. "The long tail paradigm is also about *discovery of information*, not just about finding it", Quintarelli (2005) adds. Searching with tags is much easier for non-information professionals than searching with elaborated retrieval tools like for example the International Patent Classification.

Some tags may be neologisms. Mathes (2004) discusses the words "sometaithurts" (for „so meta it hurts") and "flicktion" on Flickr. "Although small, there is a quick formation of new terms to describe what is going on, and others adopting that term and the activity is describes" (Mathes, 2004). Such an unanticipated and unexpected use of tags reflects a „communication and ad-hoc group formation facilitated through metadata" (Mathes, 2004).

Considering all documents, tags and users in a network (fig. 2), you can identify communities which share the same topic or matter (Wu, Zubair, & Maly, 2006). Applicable methods are social network analysis, informetrics and singular value decomposition. The authors conclude, “collaborative tagging systems have the potential of becoming a technological infrastructure for harvesting social knowledge” (Wu, Zubair, & Maly, 2006, p. 114). Based on shared tags or shared tagged documents it is possible to use folksonomies for implicit collaborative recommender systems (Diederich & Iofciu, 2006).

As a further benefit, although not scientifically confirmed, we can notice that people do recognize folksonomies as a type of information indexing. Maybe there is a hope that folksonomies will make people more sensitive to other indexing methods.

In conclusion, tagging shows a lot of benefits. Folksonomies

- represent an authentic use of language,
- allow multiple interpretations,
- are cheap methods of indexing,
- are the only way to index mass information on the Web,
- are sources for the development of ontologies, thesauri or classification systems,
- give the quality “control“ to the masses,
- allow searching and – perhaps even better – browsing,
- recognize neologisms,
- can help to identify communities,
- are sources for collaborative recommender systems,
- make people sensitive to information indexing.

Folksonomy: The Problems

There are many problems resulting from the absence of controlled vocabulary. “Lack of precision is a problem, though a function of user behavior, not the tags themselves“, Shirky (2004) states. In folksonomies we find different word forms, nouns in singular (“library“ in fig. 5), nouns in plural (“libraries“) and abbreviations (“IA“ or “IT“).Because Del.icio.us only allows entering one-word tags,users create phrases by leaving out blanks between single words (”informationscience“) or by combining words with an underscore (“information_science“).There is no control of synonymy and homonymy, there are many formats for dates and a lot of typing and orthographic errors. Actually, folksonomies are an unstructured list of keywords (Reamy, 2006). “Currently most users don’t give much thought to the way they tag resources, and bad or ‘sloppy’ tags are ten-a-penny in folksonomies“ (Guy & Tonkin, 2006). According to Guy and Tonkin (2006) about 40% of Flickr tags and 28% of Del.icio.us tags “were either misspelt, from a language not available via the software used, encoded in a manner that was not understood by the dictionary software, or compound words consisting of more than two words or a mixture of languages“. It seems that there are certain patterns of inconsistencies (e.g., the relationships between tags often follow

the synonymy relation) "that systems designers can anticipate" (Kipp & Campbell, 2006).

The prosumers, who tag documents, act in different contexts, have different tasks and different motivations. One user tags a document from her or his work-related view; another tags it by keeping the aspect of vacation in mind. A common basic level of indexing is missing. "There is ... systematic variation across individuals in what constitutes a basic level" (Golder & Huberman, 2006, p. 200). Expertise, linguistic power of expression, cognitive talents, actual motivations, and so on are factors which determine the basic level of indexing. Moreover, we have to consider (particular for Del.icio.us, probably not for Technorati, Flickr or YouTube) that "a significant amount of tagging, if not all, is done for personal use rather than public benefit" (Golder & Huberman, 2006, p. 207).

Web 2.0 services such as Technorati, Flickr and YouTube are used almost all over the world. Many users in non-English-speaking countries tag documents using their own language ("London", "Londres", "Londra" in fig. 5). This merging of languages (Gordon-Murnane, 2006) leads to the problems of trans-language synonymy (i.e., translation) and homonymy (in English a "gift" is a present, the meaning of the German word "Gift" is "poison" – which is certainly not a good gift).

Folksonomies are not making use of paradigmatic relations (relations of concepts in controlled vocabularies, e.g. hyponymy and meronymy). All tags of a specific document form syntagmatic relations (Stock, 2007, p. 451). Think about an image of a cow in Texel (The Netherlands) and the tags "cow", "Texel", "Northern Holland" and "Netherlands". All four tags are syntagmatically related, but the last three tags are meronyms ("Texel" is part of "Northern Holland" and "Northern Holland" is part of "Netherlands"). Indeed, this is a paradigmatic relation, but it is "hidden" and not applicable via retrieval using folksonomies.

Professionally generated metadata are segmented into different fields, e.g. the document type and the notations of classification systems. Here indexing has taken into account the aboutness and formal aspects as well. In folksonomies a strict boundary between different metadata is lacking. Of course, we find tags identifying what a document is about. But, and this is the problem, we find tags identifying the owner of the document or a formal description ("book" in fig. 1) as well (Golder & Huberman, 2006, p. 203). Besides, some tags do not describe the document, but give a judgment ("stupid"). User-specific tags describe or evaluate a document only from the user's very own perspective, so that some tags "are virtually meaningless to anybody except their creators" (Pluzhenskaia, 2006, p. 23). Some other tags can be called "performative". Often an (planned or done) activity is tagged, for example "toread" on Del.icio.us (Kipp, 2006a). Additionally there are syncategorematic tags (such terms can only be understood in the specific context). A good example for this type of tags is the tag "me" on Flickr, which describes a photo of the document's author. It may be possible that some of the keywords are spam-tags. It "involves an unethical user who propagates ... tags in order to corrupt a system" (Kroski, 2005).

Tagging images and videos leads to the problem that the different levels of ofness, aboutness, iconology and isness melt into one single level. Furthermore it is known, that indexing information objects like these

is often inconsistent.

To sum up the underpinnings of folksonomies, we see the following problems:

- **absence of controlled vocabulary,**
- **different basic levels,**
- **language merging,**
- **hidden paradigmatic relations,**
- **tags which do not only identify aboutness,**
- **spam-tags, user-specific tags, and other misleading keywords,**
- **conflation of ofness, aboutness, iconology and isness.**

Resolving (Some of) the Problems of Folksonomies by Means of NLP and Retrieval Tools

There are two different approaches aiming to solve the mentioned problems. Both approaches complement each other. First one can focus on the actors and try to educate users to improve "tag literacy" (Guy & Tonkin, 2006). An important condition for this way of resolving problems is to establish user researches concerning folksonomies (see for example Bar Ilan et al., 2006, Lin et al., 2006 and Winget, 2006), concerning the "deep nature" of tags (Veres, 2006a), discussing aspects of the folksonomy interoperability (Veres, 2006b) and the "semiotic dynamics" of folksonomies in terms of tag co-occurrences (Cattuto, Loreto, & Pietronero, 2007; Kipp & Campbell, 2006).

For training the user's selection of "good" tags it may be useful that the system would suggest some tags (MacLaurin, 2005; Xu, Fu, Mao, & Su, 2006). Tag-suggestions can operate on a syntactical level (e.g., a user attaches "graph" and the system suggests "graphics") or even on a relational level (e.g., a user attaches "graphics" and the system suggests"image", because both words do often co-occur in documents' tag clouds).

The second approach considers tags as elements of natural language and treats them by means of automatic methods of natural language processing (NLP) (Stock, 2007, chapter 13 – 18). It is known, that approximately 90% of all tags are nouns (Guy & Tonkin, 2006). But it is an open question, whether images and videos (or even further information objects) will be additionally tagged by verbs in future.

For simplicity reasons we assume that most of the tags are nouns indeed.Therefore, we are able to neglect all other kinds of words within tag-NLP. A possibility is to implement *n*-grams. However, we prefer a word-based approach because it offers opportunities for better processing tasks (the three lower levels in fig. 9).

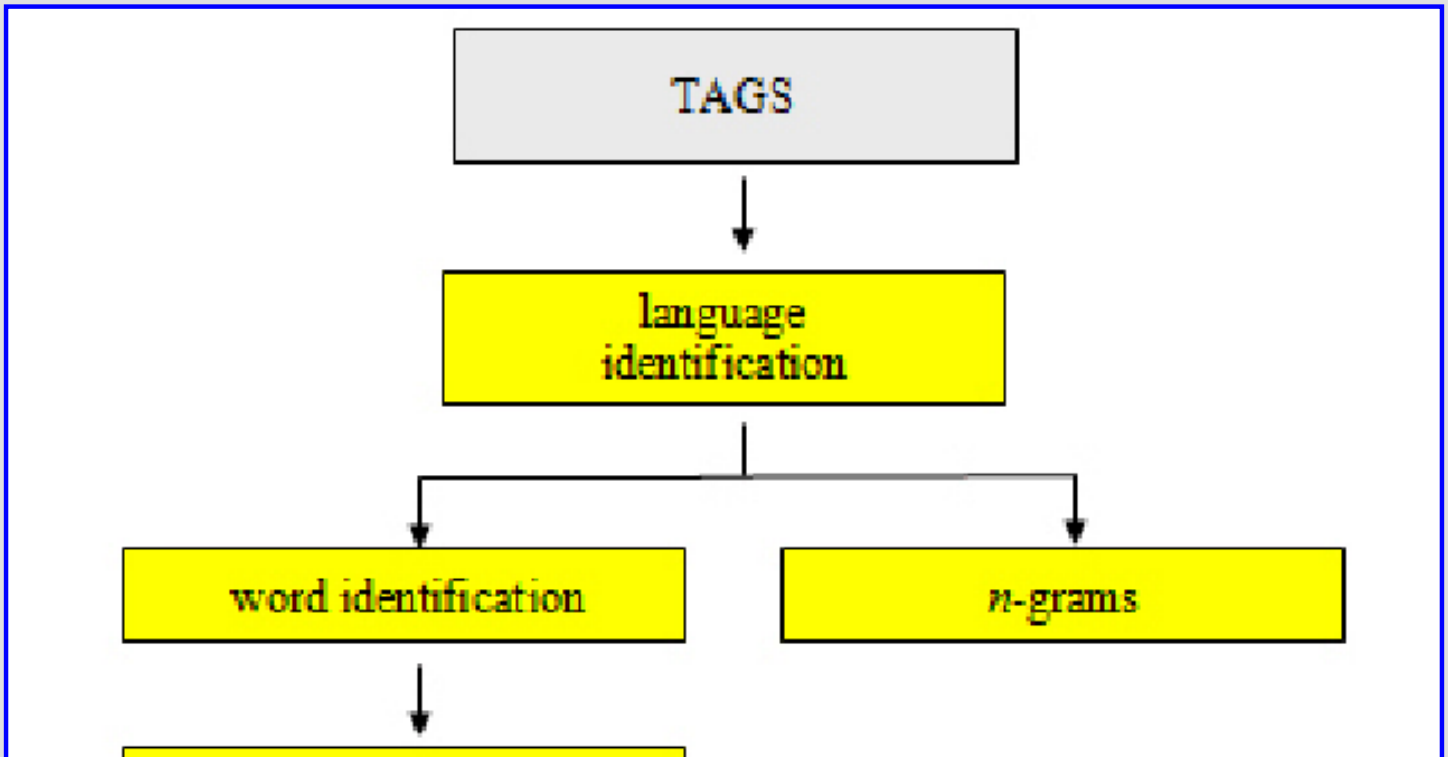
Not all documents are tagged, but at least for textual documents (e.g. blogs) we can create a substitution. Therefore we calculate important terms by TF*IDF (Brooks & Montanez, 2006a; Brooks & Montanez, 2006b). The first *n* (e.g. 3) words ranked according to the TF*IDF values are defined as tags.

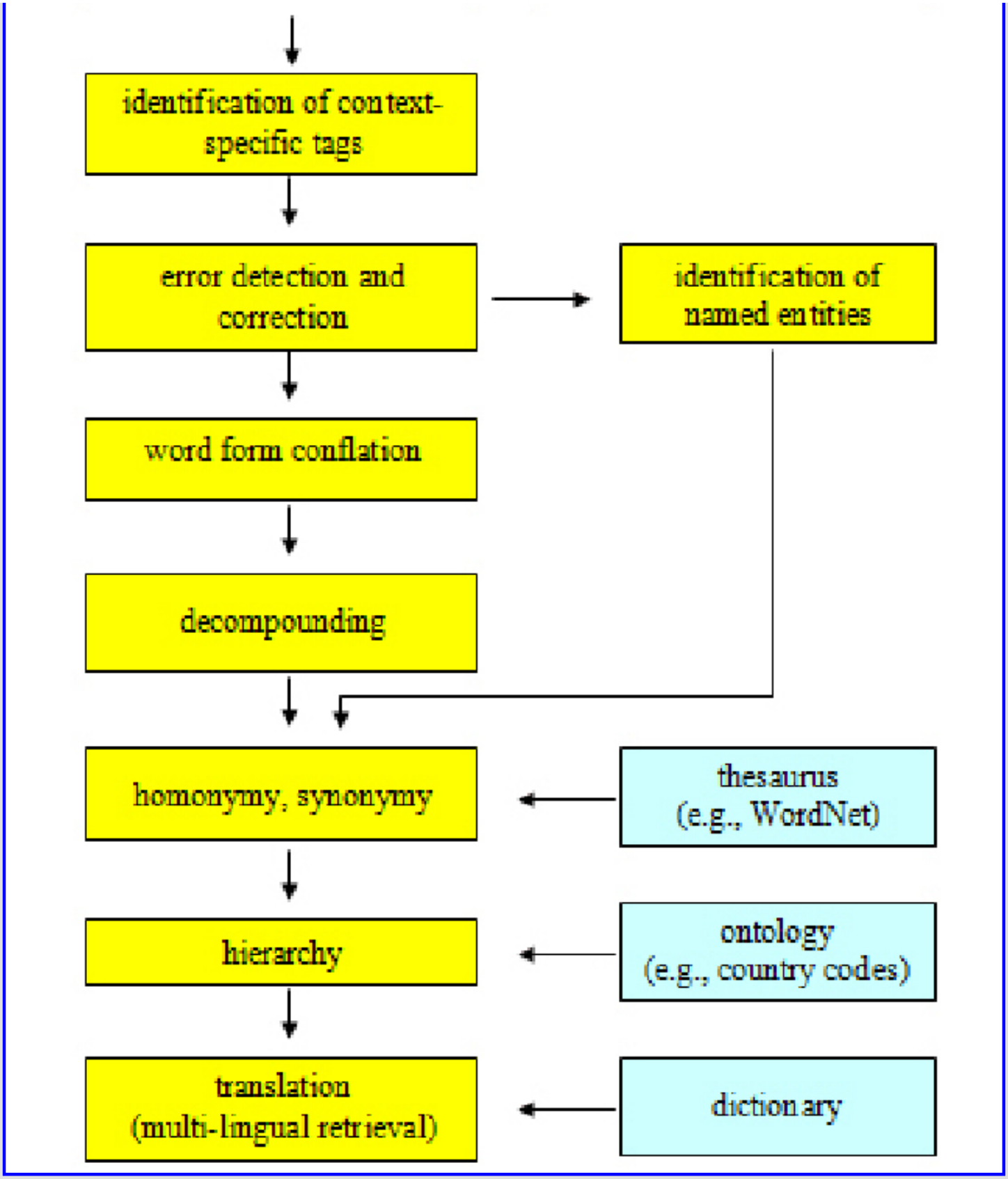
After language identification and parsing the tags we have to identify the context-specific syncategorematic tags ("me"). Our suggestion is to translate "me" into the respective user name. If this particular user himself searches for "me" he receives the matching results. A non-logged-in user receives no results after starting a search with "me". Only a search for the user name shows results.

The algorithm follows typical NLP-tasks (error detection, word form conflation, identification of named entities, phrase recognition, decompounding; for the last aspect see Tonkin, 2006) displayed in figure 9. During the detection of homonyms and synonyms one has to make use of thesauri, e.g. WordNet (Miller, 1998). On the other hand, it may be possible to solve some problems with homonyms by using statistical information of co-occurrences (Butterfield et al., 2006 [0032]) and give the evaluation of the search results to the user. Considering a tag-cluster which consists of *Java*, *Pearl* and *Programming* and another tag-cluster which entails *Java*, *Sumatra* and *Indonesia*: the user who had searched for *Java* would be confronted with both clusters and may decide what meaning of *Java* was really intended. Merging ontologies, e.g. a geographic classification system, with a folksonomy allows for using the tags in their hierarchical relations as well (Gruber, 2005). In the sense of multi-lingual information retrieval it is possible to use dictionaries or multi-lingual thesauri for providing trans-lingual access to documents.

Let us look at fig. 6 again! There are two possibilities to emphasize "power tags": a) in case of an informetric distribution we pick the first *n* tags (e.g. the first three tags) and b) in case of an inverse-logistic distribution we pick all tags of the left-hand long trunk. It would be desirable to install an additional retrieval option which exclusively retrieves documents with query-matching power tags. Accordingly, the upper example in fig. 5 would only be retrieved by searching with the terms "London", "travel", and "UK". All other terms and tags, respectively, would be ignored.

Figure 9: Steps of natural language processing concerning tags.





Relevance Ranking of Tagged Documents

A still unsolved problem is the ranking of tagged documents returned as search results for a given query. By using NLP the search engine can create a set of documents which satisfy the query. This set is still unsorted. The next step is to rank the documents of the results set. A Yahoo! (Flickr) patent application

gives a list of five general ranking criteria of "interestingness" in a narrow folksonomy (Butterfield et al., 2006 [0008]): (a) the number of tags attached to a document, (b) the number of users, who tagged the document, (c) the number of users, who retrieved the document, (d) time (the older the document is the less relevance it has) and (e) relevance of metadata. Additionally, the patent application mentions two further ranking criteria for the "personalized interestingness rank" (Butterfield et al., 2006 [0010-0011]): (f) user preferences (e.g., designated favorites) and (g) the residence of the user.

Hence, there are three sets of factors for ranking tagged documents by relevance:

- tags,
- collaboration,
- actor-specific ranking criteria.

Each factor or sub-factor can be weighted or not.

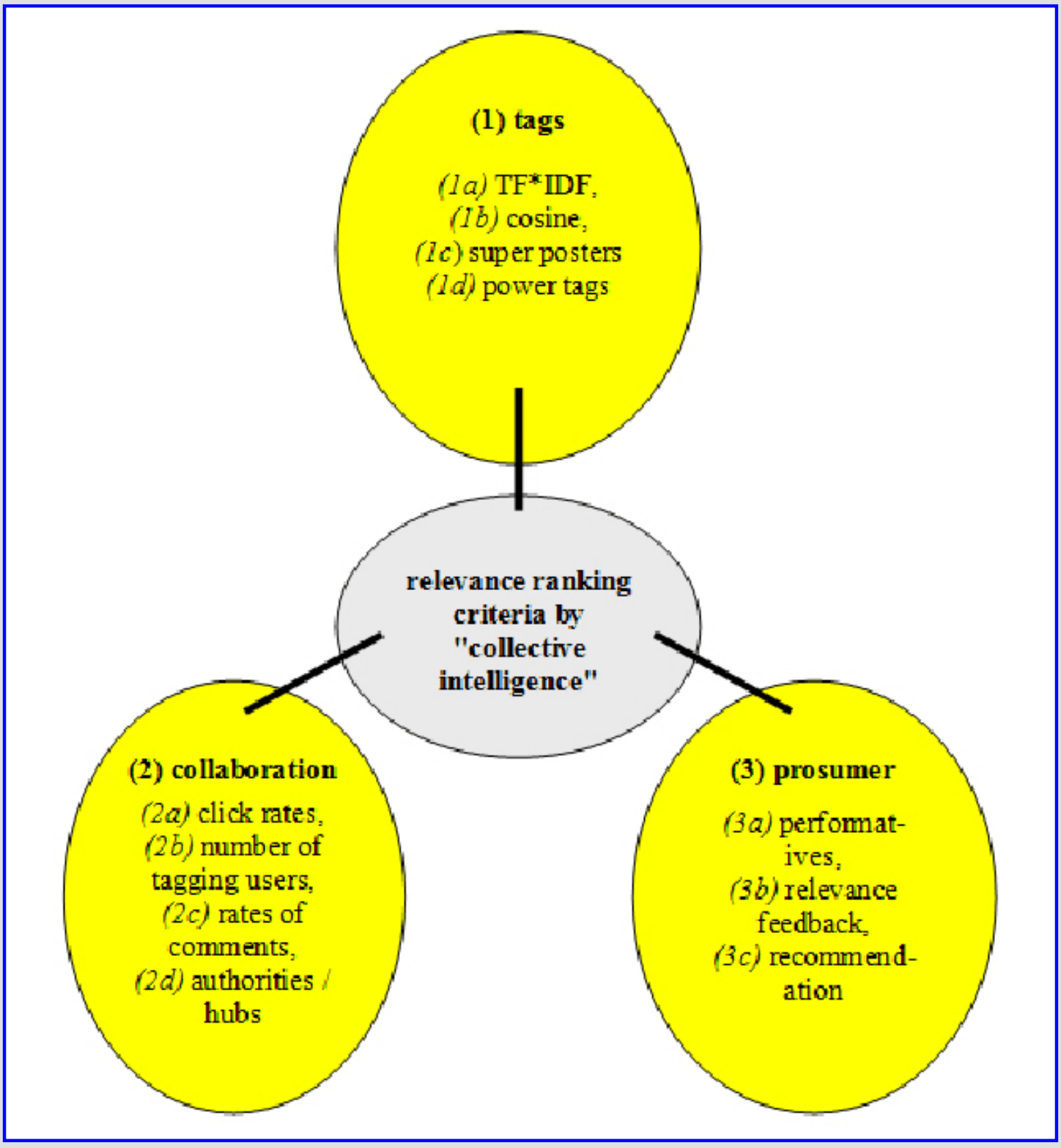
The tag factor (cloud (1) in figure 10) will work adequately in a vector space model (Stock, 2007, chapter 20), where the dimensions are the different tags in the database, the value of the dimension is determined by TF*IDF (1a), the documents are represented by vectors and finally the similarity between a query and a document is calculated by the cosine (1b). TF within broad folksonomies depends on the number of actors tagging the document with the given tag. TF within narrow folksonomies is calculated by the number of searches which found the document by the given tag.

Approaches with an adapted PageRank-algorithm according to the motto "The basic notion is that a resource which is tagged with important tags by important users becomes important itself" (Hotho, Jäschke, Schmitz, & Stumme, 2006a, p. 417) is problematic and leads to dissatisfying results. The same survey implements a modified algorithm called "FolkRank" to "focus the ranking around the topics defined in the preference vector" (p. 419). The FolkRank tracks the idea of "super-posters" or "super-authors", who publish a huge amount of content and may be experts in a particular field (1c). Accordingly, matching search results published by super-posters are ranked higher than others (Hotho, Jäschke, Schmitz, & Stumme, 2006a; 2006b). Depending on the type of tag distribution some tags (those of the long trunk or the few first tags in a power-law distribution) are marked as "power tags" (1d) – which has effects on the ranking. Power tags enable the user to use them in a special search field (and restrict the search to gain results with higher precision), and the system can weight power tags higher than others.

Ranking regarding click rates (2a) takes the concrete collaboration behavior into account (Culliss, 1997). For Jung, Herlocker and Webster (2007) click rates are aspects of implicit relevance feedback in Web search. How often a document was opened is a user-oriented ranking criterion in sense of folksonomies and social web. Perhaps the number of different users who have indexed the document is a useful ranking criterion (2b) as well. Highly participated discussions based on particular documents give the impression of being of high interest in the community. The number of comments to a given document (2c) determines on which place of the ranking list it will be placed. Especially blogs and Wikis offer potentials for detecting

hubs and authorities with known algorithms (Kleinberg, 1999; Brin & Page, 1998) because of their intense connections via hyperlinks (2d). Authorities are situated in the centre of interest of given topics and therefore may be ranked higher.

Figure 10: Criteria for relevance ranking of tagged documents.



Documents which are tagged with “to do” or “to read” seem to be very important for the users and should

influence the ranking (3a). A further possibility is to use the user's interest in participating actively in the ranking procedure. We can imagine two ways of user feedback: (3b) a relevance feedback after displaying and judging search results (e.g. Rocchio, 1971) or (3c) a formal recommending system via "stars assessment" or a question dialogue like 'Did this document help you?'. The users have to judge the documents and to give values for the document's quality, which is a "democratic way" of information review. Both values are taken into account implicitly, while ranking search results, and explicitly, as the judgements are displayed as hints to the community, like e.g. 'Other users with these search terms considered following documents as very helpful' or just  **. Such relevance feedback tools have not yet been put into action in known Web 2.0 services.**

A drop-down menu is the most user-oriented way of ranking search results. Here the user has to choose in which order the results are displayed. Known are rankings regarding date of publication, tags, etc., but what about popularity (authorities), topics, super-posters, in which post, bookmark, photo did the search term appear first? Or should we use postal codes (learned from user profiles) to connect communities in the real world?

Conclusion and Further Research

Our paper is a first step to NLP and relevance ranking of tagged documents. We could show that there are many benefits of folksonomies. But, there are many shortcomings as well which we have to face and to solve properly. Our suggestions of resolving the problems of information retrieval with folksonomies are of a provisional nature.

Although there is no 'one-fits-all'-solution, we summarize our considerations concerning natural language processing and relevance ranking of tagged documents with the following important questions:

- **What steps of NLP are usable in practice?**
- **Does the differentiation between broad and narrow folksonomies make sense in the practice of information retrieval?**
- **In absence of tags: What are the most commonly used words in the document?**
- **Which role do power tags play?**
- **Is it really true, that most of the tags are nouns? Or do verbs play an important role in social tagging as well?**
- **Which author did publish the most documents? Is this fact really relevant for retrieval algorithms?**
- **How often has a document been opened? Is that a ranking criterion as well?**
- **Which document has the most comments? (Ranking criterion?)**
- **What do other users want to do with this document?**
- **What is the nature of a tag?**
- **Which documents are authorities, which are hubs?**
- **Can I recommend this document? How can I rate it? Is recommending useful in retrieval systems of Web 2.0 services?**

- **Which are the most wanted ranking criteria?**

References

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting navigation for collaborative tagging with emergent semantics. *Lecture Notes in Computer Science,* **4273, 58-71.**

Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y, & Shachak, A. (2006). Structured vs. unstructured tagging - A case study. *In Proceedings of the 15th International WWW Conference.* **Collaborative Web Tagging Workshop.**

Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems,* **30, 107-117.**

Brooks, C.H.; Montanez, N. (2006a). An analysis of the effectiveness of tagging in blogs. *Computation Approaches to Analyzing Weblogs. Papers from the 2006 AAAI Spring Symposium* **(pp. 9-15). Nicolov, N., Salvetti, F., Liberman, M., & Martin, J.H. (eds.), Menlo Park, Cal: AAAI Press. (Technical Report SS-06-03. American Association for Artificial Intelligence).**

Brooks, C.H., Montanez, N. (2006b). Improved annotation of the blogosphere via autotagging and hierarchical clustering. *In Proceedings of the 15th International World Wide Web Conference* **(pp. 625-632). New York: ACM.**

Butterfield, D.S., Costello, E., Fake, C., Henderson-Begg, C.J., & Mourachow, S. (2006). Interestingness ranking of media objects. *Patent Application No. US 2006/0242139 A1.* **Filed: Feb. 8, 2006. Publ.: Oct. 26, 2006.**

Butterfield, D.S., Costello, E., Fake, C., Henderson-Begg, C.J., Mourachow, S., & Schachter, J.E. (2006). Media object metadata association and ranking. span class = 'referencesource'>Patent Application No. US 2006/0242178 A1. Filed: Feb. 8, 2006. Publ.: Oct. 26, 2006.

Bruza, P.D., Song, D.W., & Wong, K.F. (2000). Aboutness from a commonsense perspective. *Journal of the American Society for Information Science,* **51, 1090-1105.**

Campbell, D.G. (2006). A phenomenological framework for the relationship between the semantic Web and user-centered tagging systems. *In 17th ASIS&T SIG/CR Classification Research Workshop.*

Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *In Proceedings of the National Academy of Sciences of the United States of America,* **104(5), 1461-1464.**

Christiaens, S. (2006). Metadata mechanisms: From ontology to folksonomy ... and back. *Lecture Notes in Computer Science,* **4277, 199-207.**

Culliss, G.A. (1997). Method for organizing information. *Patent-No. US 6.006.222.* **Granted: Dec. 21, 1999. Applied: Jan. 8, 1997.**

Diederich, J., Iofciu, T. (2006). Finding communities of practice from user profiles based on folksonomies. *In Proceedings of the 1st International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice* **(TEL-CoPs'06).**

Dye, J. (2006). Folksonomy: A game of high-tech (and high-stakes) tag. *E-Content,* **29(3), 38-43.**

Egghe, L., Rousseau, R. (1990). Introduction to Informetrics. *Elsevier.* **Amsterdam**

Fichter, D. (2006). Intranet applications for tagging and folksonomies. *Online,* **30(3), 43-45.**

Fidel, R. (2006). An ecological approach to the design of information systems. *Bulletin of the American Society for Information Science and Technology,* **33(1), 6-8.**

Furnas, G.W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., & Naaman, M. (2006). Why do tagging systems work? *CHI '06 Extended Abstracts on Human Factors in Computing Systems* **(pp. 36-39). New York: ACM.**

Garfield, E. (1979). Citation Indexing. *Wiley* **New York**

Gendarmi, D., Lanubile, F. (2006). Community-driven ontology evolution based on folksonomies. *Lecture Notes in Computer Science,* **4277, 181-188.**

Golder, S.A., Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science,* **32(2), 198-208.**

Gordon-Murnane, L. (2006). Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher - The Magazine for Database Professionals,* **14(6), 26-38.**

Gruber, T. (2005). Ontology of folksonomy: A mash-up of apples and oranges. *1st On-Line Conference on Metadata and Semantics Research (MTSR '05).*

Gruber, T. (2006). Where the social Web meets the semantic Web. *Lecture Notes in Computer Science,* **4273, 994.**

Guy, M., Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine,* **12(1).**

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools. A general review. Part 1. *D-Lib Magazine,* **12(1).**

Heymann, P., Garcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging system. *InfoLab Technical Report;* **2006-10. Stanford: Stanford University / Computer Science Dept.**

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006a). Information retrieval in folksonomies: Search and ranking. *Lecture Notes in Computer Science,* **4011, 411-426.**

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006b). Trend detection in folksonomies. *Lecture Notes in Computer Science,* **4306, 56-70.**

Huang, H. (2006). Tag distribution analysis using the power law to evaluate social tagging systems: A case study in the Flickr database. *17th ASIS&T SIG/CR Classification Research Workshop.* **Abstracts of Posters (pp. 14-15).**

Hutchins, W.J. (1978). The concept of 'aboutness' in subject indexing. *Aslib Proceedings,* **30(5), 172-181.**

Ingwersen, P. (2002). Cognitive perspectives of document representation. *CoLIS 4: 4th International Conference on Conceptions of Library and Information Science [pp. 285-300].* **Greenwood Village: Libraries Unlimited.**

Jung, S., Herlocker, J.L., & Webster, J. (2007). Click data as implicit relevance feedback in web search. *Information Processing & Management,* **43, 791-807.**

Kipp, M.E.I. (2006a). @toread and cool: Tagging for time, task and emotion. *17th ASIS&T SIG/CR Classification Research Workshop.* **Abstracts of Posters (pp. 16-17).**

Kipp, M.E.I. (2006b). Complementary or discrete contexts in online indexing: A comparison of user, creator and intermediary keywords. *Canadian Journal of Information and Library Science,* **30(3), in prep.**

Kipp, M.E.I. & Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Proceedings of the 69th Annual Meeting of the American Society for Information Science & Technology* **(Vol. 43), Austin, TX (CD-ROM).**

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM,* **46(5), 604-632.**

Kroski, E. (2005). The Hive Mind: Folksonomies and user-based tagging. Blog post; 2005-12-07http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging

Lancaster, F.W. (2003). Indexing and Abstracting in Theory and Practice. *University of Illinois.* **3rd ed. Champaign**

Layne, S. (2002). Subject access to art images. *Introduction to Art Image Access* **Baca, M. (ed.): [pp. 1-19]. Los Angeles: Getty Research Institute.**

Lin, X., Beaudoin, J.E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. *17th ASIS&T SIG/CR* **Classification Research Workshop.**

Maron, M.E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science,* **28, 38-43.**

Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social bookmarking tools (II). A case study - Connotea. *D-Lib Magazine,* **11(4).**

Macgregor, G., McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review,* **55(5), 291-300.**

MacLaurin, M.B. (2005). Selection-based item tagging. *Patent Application no. US 2007/0028171 A1.* **Filed: Jul. 29, 2005. Publ.: Feb. 1, 2007.**

Mai, J.M. (2006). Contextual analysis for the design of controlled vocabularies. *Bulletin of the American Society for Information Science and Technology,* **33(1), 17-19.**

Markey, K. (1986). Subject Access to Visual Resource Collections. *Greenwood* **Westport**

Markkula, M., Sormunen, E. (2000). End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval,* **1, 259-285.**

Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. *University of Illinois Urbana-Campaign / Graduate School of Library and Information Science.* **Urbana, Ill.**

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. *Proceedings of the 17th Conference on Hypertext and Hypermedia* **(pp. 31-40). New York: ACM.**

McFedries, P. (2006). Folk wisdom. *IEEE Spectrum,* **43(Febr.), 80.**

Merholz, P. (2004). Metadata for the masses. Blog post, 2004-10-19http://www.adaptivepath.com/publications/essays/archives/000361.php

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. *Lecture Notes in Computer Science,* **3729, 522-536.**

Millen, D.R., Feinberg, J., & Kerr, B. (2006). DOGEAR. Social bookmarking in the enterprise. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* **(pp. 111-120). New York: ACM.**

Miller, G.A. (1998). Nouns in WordNet. *WordNet. An Electronic Lexical Database* **Fellbaum, C. (ed.), (pp. 23-46). Cambridge, Mass., London: MIT Press.**

Murison, J. (2005). Messageboard topic tagging: User tagging of collectively owned community content. *Proceedings of the 2005 Conference on Designing for User eXperience* **(art. no 5). New York: American Institute of Graphic Art.**

O'Reilly, T. (2005). What is Web 2.0. Design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Ornager, S. (1995). The newspaper image database. Empirical supported analysis of users' typology and word association clusters. *Proceedings of the 18th Annual International ACM SIGIR Conference on*

Research and Development in Information Retrieval **(pp. 212-218). New York: ACM.**

Ott, E.S. (2006). Picture tagging. *Patent Application US 2007/0079321 A1.* **Filed: Feb. 17, 2006. Publ.: Apr. 5, 2007.**

Panofsky, E. (1975) Sinn und Deutung in der bildenden Kunst [Sense and Interpretation in Fine Arts]. *DuMont* **Köln**

Panofsky, E. (2006). Ikonographie und Ikonologie [Iconography and Iconology]. *DuMont* **Köln**

Peters, I. (2006a). Inhaltserschließung von Blogs und Podcasts im betrieblichen Wissensmanagement [Indexing of blogs and podcasts for the corporate knowledge management]. *Content. 28. Online-Tagung der DGI, 58. Jahrestagung der DGI. Proceedings* **Ockenfeld, M. (ed.): [pp. 143-151]. Frankfurt: DGI.**

Peters, I. (2006b). Against folksonomies: Indexing blogs and podcasts for corporate knowledge management. *Preparing for Information 2.0. Online Information 2006. Proceedings* **Jezzard, H. (ed.): (pp. 93-97). London: Learned Information Europe.**

Peters, I. (2007). Web 2.0 und Fachinformation. Folksonomies im Web 2.0 [Web 2.0 and professional information. Folksonomies in Web 2.0 services]. *Password,* **no. 5, 10-11.**

Peters, I., Stock, W.G. (2006). Corporate Blogs im Wissensmanagement [Corporate blogs used in knowledge management]. *Wissensmanagement,* **no. 6, 40-41.**

Peterson, E. (2006). Beneath the metadata. Some philosophical problems with folksonomies. *D-Lib Magazine,* **12(11).**

Pluzhenskaia, M. (2006). Folksonomies or fauxsonomies: How social is social bookmarking? *17th ASIS&T SIG/CR Classification Research Workshop.* **Abstracts of Posters (pp. 23-24).**

Quintarelli, E. (2005). Folksonomies: Power to the people. *ISKO Italy UniMIB meeting,* **Milan, June 24, 2005.**

Rasmussen, E.M. (1997) Indexing images. *Annual Review of Information Science and Technology,* **32, 169-196.**

Reamy, T. (2006). Folksonomies and complexity theory: Evolved information structures. *Preparing for Information 2.0. Online Information 2006. Proceedings* **Jezzard, H. (ed.): (pp. 111-113). London: Learned Information Europe.**

Rocchio, J.J. (1971). Relevance feedback in information retrieval. *The SMART Retrieval System - Experiments in Automatic Document Processing* **Salton, G. (ed.): (pp. 313-323). Englewood Ciffs, N.J.: Prentice Hall**

Russell, T. (2006). Cloudalicious: Folksonomy over time. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* **(p. 364). New York: ACM.**

Schmitz, P. (2006). Inducing ontology from Flickr tags. *Proceedings of the 15th International WWW Conference.* **Collaborative Web Tagging Workshop.**

Shatfort, S. (1986). Analyzing the subject of a picture. A theoretical approach. *Cataloguing and Classification Quarterly,* **6(3), 39-62.**

Shirky, C. (2004). Folksonomy. Blog post; 2004-08-24http://many.corante.com/archives/2004/08/25/folksonomy.php

Shirky, C. (2005). Ontology is overrated: Categories, Links, and Tags. www.shirky.com/writings/ontology_overrated.html

Smith, G. (2004). Folksonomy: Social classification. Blog post; 2004-08-03http://atomiq.org/archives/2004/08/folksonomy_social_classification.html

Smith, M.K. (2006). Viewer tagging in art museum: Comparisons to concepts and vocabularies of art museum visitors. *17th ASIS&T SIG/CR* **Classification Research Workshop.**

Spiteri, L.F. (2006). The use of folksonomies in public library catalogues. *The Serials Librarian,* **51(2), 75-89.**

Spyns, P., de Moor, A., Vandenbussche, J., & Meersman, R. (2006). From folksonomies to ontologies: How the twain meet. *Lecture Notes in Computer Science,* **4275, 738-755.**

Stock, W.G. (2006). On relevance distributions. *Journal of the American Society for Information Science and Technology,* **57(8), 1126-1129.**

Stock, W.G. (2007). Information Retrieval. Suchen und Finden von Informationen [Information Retrieval. Searching and Finding Information]. *Oldenbourg* **München, Wien**

Stock, W.G. & Weber, S. (2006). Facets of informetrics. *Information - Wissenschaft und Praxis,* **57(8), 385-389.**

Toffler, A. (1980). The Third Wave. *Morrow* **New York**

Tonkin, E. (2006). Searching the long tail: Hidden structure in social tagging. *17th ASIS&T SIG/CR* **Classification Research Workshop.**

Trant, J. (2006a). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia,* **12(1), 83-105.**

Trant, J. (2006b). Social classification and folksonomy in art museums: Early data from the steve.museum tagger prototype. *17th ASIS&T SIG/CR* **Classification Research Workshop.**

Vander Wal, T. (2004). Feed on this. Blog post; 2004-10-03http://www.vanderwal.net/random/category.php?cat=153

Vander Wal, T. (2005). Explaining and showing broad and narrow folksonomies. Blog post; 2005-02-21http://www.vanderwal.net/random/category.php?cat=153

Veres, C. (2006a). The language of folksonomies: What tags reveal about user classification. *Lecture Notes in Computer Science,* **3999, 58-69.**

Veres, C. (2006b). Concept modeling by the masses: Folksonomy structure and interoperability. *Lecture Notes in Computer Science,* **4215, 325-338.**

Weiss, A. (2005). The power of collective intelligence. *netWorker,* **9(3), 16-23.**

Winget, M. (2006). User-defined classification on the online photo sharing site Flickr ... Or, how I learned to stop worrying and love the million typing monkeys. *17th ASIS&T SIG/CR* **Classification Research Workshop.**

Wu, H., Zubair, M., & Maly, K. (2006). Harvesting social knowledge from folksonomies. *Proceedings of the 17th Conference on Hypertext and Hypermedia* **(pp. 111-114). New York: ACM.**

Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). Towards the semantic Web: Collaborative tag suggestions. *Proceedings of the 15th International WWW Conference.* **Collaborative Web Tagging Workshop.**

Zhang, L., Wu, X., & Yu, Y. (2006). Emergent semantics from folksonomies: A quantitative study. *Lecture Notes in Computer Science,* **4090, 168-186.**