

Folksonomies and ontologies: two new players in indexing and knowledge representation

Katrin Weller

Scientific Assistant/Researcher, Heinrich-Heine-University, Düsseldorf, Germany

Abstract

The current trends of the semantic web and web 2.0 have led to a growing interest in knowledge representation methods. On the one hand ontologies are being designed by knowledge engineers to perform advanced information integration, on the other hand users apply content-descriptive tags to large-scale document collections.

These two new systems broaden the spectrum of knowledge representation methods in different directions. Both approaches will be introduced and discussed regarding their novelty, singular characteristics, advantages and shortcomings. The possibilities of combining social tagging with other techniques will also be sketched and proposed for discussion.

Introduction

A key problem of today's information society is how to structure, find and retrieve information precisely and effectively. One approach to this problem is to use methods of knowledge representation to annotate or index documents, which helps to perform effective retrieval and aids users in deciding on a document's relevance. Classical methods comprise classification systems (taxonomies), thesauri, and controlled keywords (nomenclatures) [Aitchison et al. 2004; Cleveland and Cleveland 2001; Lancaster 2003; Stock and Stock 2008]. As controlled vocabularies they provide a unified access to documents. They are used in libraries and professional databases by, for example, commercial content providers.

Recently, two new developments can be observed which both address issues of document indexing and knowledge representation: folksonomies and ontologies. They complement traditional techniques in different ways. Folksonomies include novel social dimensions of tagging [Mathes 2004; Smith 2004]; ontologies extend the possibilities of formal vocabulary structuring [Alexiev et al. 2005; Davies et al. 2003; Staab and Studer 2004]. Both approaches have revived discussions about metadata on the web [Madhavan et al. 2006; Safari 2004]. Their use has led to an increasing awareness of knowledge representation issues in scientific areas and even within the common web-user community. This paper will offer an overview of these two new trends in document indexing. Still, folksonomies and ontologies should not be viewed as rival systems [Gruber 2005]. After introducing characteristics of each of them, options for combinations will be presented.

Folksonomies: metadata for everyone

Within current web 2.0 discussions, folksonomies are among the new developments that receive most attention. They address the well-known problem of indexing data with content-descriptive metadata. But they add the new

dimensions of user-involvement and web-collaboration. Recently, users have begun to publish their own content on the web on a large scale. They have also started to use social software to store and share documents (such as photographs, videos and bookmarks) [Hammond et al. 2005; Gordon-Murnane 2006; O'Reilly 2005], and they annotate these documents with their own keywords to make them retrievable. But keywords are not called keywords anymore (and by no means are they termed 'index-terms' or 'descriptors'); in web 2.0 they are called tags. The indexing process is called (social) tagging, and the collection of tags used within one platform is called *folksonomy*.

Folksonomies have become well-known through social software such as the photo-sharing platform Flickr¹ or the video-community YouTube². By now they are an essential part of many web 2.0 applications. Users tag documents, references, pictures, videos, blog posts, discussion, bookmarks or even groups and other users. Tags are used as a new entry point for data collections.

Two different types of folksonomies can currently be distinguished according to Vander Wal: broad folksonomies are systems in which one document can be tagged by different people, so that tags can be assigned more than once (for example, del.icio.us³). Systems where a tag can be assigned to a document only once are called narrow folksonomies (such as systems where only the 'author' of a document may tag it, such as Flickr) [Vander Wal 2005; see also Peters and Stock 2007]. In broad folksonomies tag clouds can be generated. These display the popularity of tags for one document or within the entire folksonomy (terms that are applied more often are displayed in bigger font sizes).

In terms of document indexing, folksonomies offer advantages regarding cost-efficiency and feasibility for large data collections. They are easy to handle and take into account the users' own vocabulary. Thus, social tagging has made large communities aware of the use of content-descriptive metadata. On the other hand, folksonomies do not include any vocabulary control. For example, synonyms are not bound together, homonyms are not distinguished,

¹ Flickr: <http://www.flickr.com>.

² You Tube: <http://www.youtube.com>

³ Del.icio.us: <http://del.icio.us>.

and there is a lack of recall and precision. Users of social tagging systems may not yet be familiar with these effects but, in time, they may be able to learn about them and change their tagging behaviour respectively [Guy and Tonkin 2006].

The pros and cons of folksonomies are well discussed [for example, Kroski 2006; Peters and Stock 2007]. Some key aspects are the following:

Controlled versus active vocabulary

The main property of a folksonomy is that it captures the active language-use of a community. This freedom in the choice of tags also means that folksonomies are entirely uncontrolled vocabularies, which leads to the so-called 'vocabulary problem' [Furnas et al. 1987] that is considered to be a main problem of tagging systems [for example, Golder and Huberman 2006; Furnas et al. 2006; Mathes 2004]: different people use different words to describe the same object (or document). In folksonomies, synonyms, trans-language synonyms, spelling variants and abbreviations are not bound together. Homonyms are not distinguished. Misspellings and encoding limitations are serious problems for folksonomies [Guy and Tonkin 2006]. This all leads to a lack of precision and recall when executing a search.

On the other hand, this flexibility in the choice of tags is a great advantage when it comes to timeliness and multiple perspectives. A controlled vocabulary is always bound to a certain point in time and to a certain point of view. Folksonomy users can create tags quickly in response to new developments and changes in terminologies [Kroski 2006].

This also leads to new options in deriving inherent knowledge from social tagging systems. For example, the Tagline Generator⁴ can visualise how popular terms within a topic-specific tag cloud change over time. The result is a timeline-based tag cloud, which can help to observe developments or opinions regarding certain events.

Social versus personal tagging

Although we speak of social tagging, the intentions of tags are not always social. Users who tag documents do not necessarily do this with the objective of helping a community to find relevant documents or photos or bookmarks. Many users simply use tags to organise their own private documents. Thus, many tags in use are rather personal than social [Guy and Tonkin 2006].

Al-Khalifa and Davis [2007] have analysed the nature of tags by grouping them into three categories: personal, subjective and factual. According to their study, the majority of assigned tags in a test collection (62 per cent) were factual in that they referred to the actual content of a document. A still large number of tags (32 per cent) were of a personal nature. In other words they were clearly intended for the creator's own use and would not be of practical use for other users, examples for this category are self-references such as a photo tagged with 'me' or 'my_dog', or instructions for personal document organisation, for instance 'toread'. Only 4 per cent of tags were classified as subjective judgements, such as 'cool' or 'interesting'.

Similar studies have been processed [see, for example, Golder and Huberman 2006; Kipp 2006] and more will surely be conducted with larger or different test sets to gain even better understanding on tagging behaviour. Maybe the tag categorisations may also be specified so that, for example, the personal tags are divided into different sub-categories.

Golder and Huberman [2006] identify seven possible functions of tags:

1. Identifying what (or who) a document is about.
2. Identifying what the document itself is.
3. Identifying who owns the document.
4. Refining documents or other tags.
5. Identifying qualities or characteristics.
6. Self reference.
7. Task organising.

One question is, whether the users of folksonomies might be willing to distinguish between different types of tag functions. This would help to distinguish individual organisational tags from objective content-descriptive ones.

Although tagging is often motivated by personal interests, a social community profits from folksonomies in different ways. The most important way this works is that they provide alternative views; they 'include everyone's vocabulary and reflect everyone's needs without cultural, social, or political bias' [Kroski 2006], so that even niche interests can be included.

Retrieval versus exploration

Considering the previous two confrontations, we end up with a discussion about whether folksonomies are more about (classical) retrieval or whether they are about new ways of discovering interesting resources. Compared to other retrieval strategies, folksonomies lack precision and recall. Peters and Stock [2007] provide some ways to resolve these retrieval problems using methods drawing on natural language processing (NLP).

Others argue that the strength of folksonomies lies in serendipity [Mathes 2004], in discovering information via different paths, and in easy-to-handle search mechanisms [Quintarelli 2005]. Folksonomies provide these different entry points to document collections as they consist of three basic types of entities which offer different information when combined: documents, tags and users. Thus, users may browse the relations between all of them, if the platform supports it. Documents are related via the tags and via the users that tagged them. Users are related if they use the same tags or if they tag the same documents [Peters and Stock 2007]. There are even ways of identifying communities of interest with the help of folksonomies [Diederich and Iofciu 2006; Wu et al. 2006].

Ontologies: semantic structures for the web

Ontologies are a key component in research efforts to establish a semantic web [Antoniou and van Harmelen 2004; Breitman et al. 2007]. They aim to handle the problem of organising information and documents in a way which is quite contrary to folksonomies. Ontologies are designed by experts and should be used for making the meaning of documents explicit and unambiguous, not only for interpersonal communication but also for human-computer and inter-computer interactions. They are formal conceptualisations of a knowledge domain [Gruber 1993], expressed in systems of concepts (classes), instances (individuals) and the relations between them.

At first glance, ontologies strongly resemble traditional

⁴ Tagline Generator: <http://chir.ag/tech/download/tagline/>.

thesauri but they also include possibilities for defining new types or relations between concepts and they allow the adding of rules and axioms [Stock and Stock 2008]. Thus, they are even more formalised than traditional knowledge representation systems and can be used for detailed depiction of knowledge domains and contexts. Ontology editors (such as, for example, Protégé⁵) help to create and maintain ontologies, and to store them in special ontology language formats (such as OWL⁶). Creating an ontology is laborious and requires careful consideration about how to represent a domain of interest adequately [Michlmayr 2005]. Ontology engineers or information architects as well as domain experts are needed to formalise precise definitions within the ontology [Paulsen et al. 2007].

What is an ontology? The spectrum of knowledge representation systems

The term 'ontology' is not always used in a consistent way and so definitions of ontology vary. This is to some extent because of the heterogeneity of the (scientific) community dealing with ontologies today. Areas involved in this research include computer science, information science, philosophy, computer linguistics, artificial intelligence, life sciences and bioinformatics, which all have different backgrounds and different requirements for ontologies. Furthermore, ontologies originally arose as a method of information sharing and exchange; their possible role in a semantic web and advanced information integration developed some years later [Gruber 2005]. Yet, they fit very well into the set of knowledge representation methods applicable for document indexing. The main problem, then, is how to set the scope of the definition; how to mark down what exactly an ontology is.

Currently the term is often assigned to systems that cannot keep up with the vision of a highly structured, complex knowledge representation system. Many so-called ontologies are merely hierarchical constructs – and thus do not exceed traditional thesauri. (Indeed, they sometimes

even lag behind a thesaurus' complexity.)⁷ Ontology editors and ontology languages can be easily used for designing simple as well as complex systems, which also makes it difficult to recognise graduations. Sometimes established thesauri or taxonomies are referred to as ontologies, for example the UMLS thesaurus⁸ or even the Yahoo categorisation⁹ [Lassila and McGuinness 2001]. Yet, in some cases, thesauri are also being upgraded and enhanced with additional semantic relations.

In this context, we are principally dealing with two distinct notions of the term ontology¹⁰:

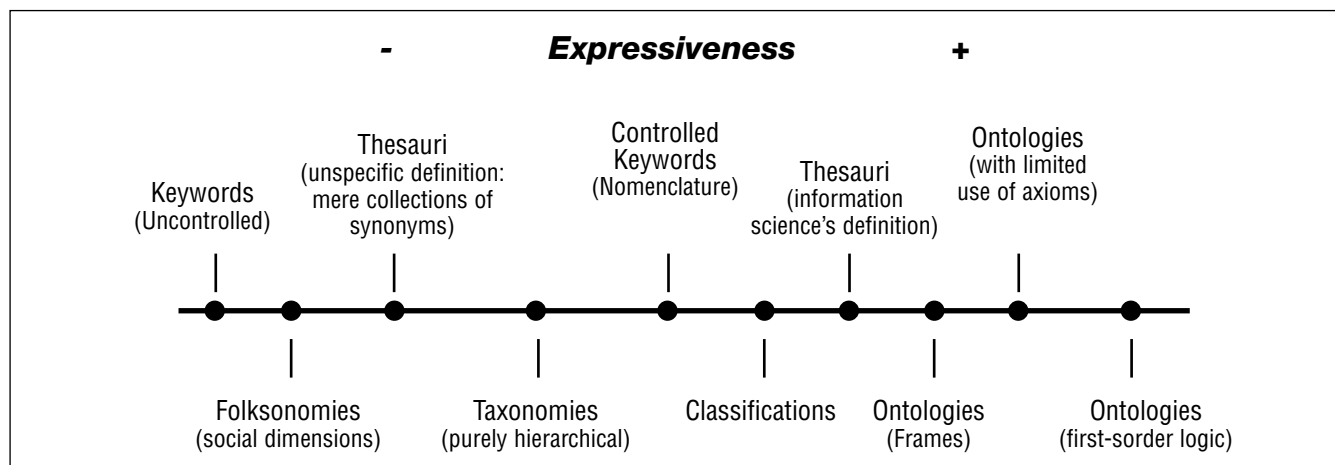
- 1) Ontology as a new general concept, subsuming all kinds of existing knowledge representation systems – or at least all that use fixed concept-relation-structures.
- 2) Ontology as a new type of knowledge representation system, expanding the possibilities of traditional methods.

Although the first option is more frequently used, we prefer the second one for use in information science and related disciplines as it allows distinguishing semantically richer systems from thesauri, classifications and folksonomies, although we may find that distinguishing on this level will not always be necessary. The main difference lies in the presence or absence of vocabulary control, especially when it comes to comparing folksonomies with other approaches (as in the last section of this paper). The level of formalisation does not apply to some basic discussions, so that ontologies, thesauri and classifications may actually be summed up and regarded as one counterpart to uncontrolled keyword systems.

Further specifications of ontology types are needed, perhaps regarding the ontology language in use (for example, different levels of OWL) or the specificity of the depicted domain [Gomez-Perez et al. 2004]. Additional delimitations should be discussed, regarding systems such as topic maps or relational databases.

Figure 1 shows one possibility for sorting popular knowledge representation systems according to their

Figure 1:
Approach to distinguish different vocabularies according to their expressiveness.
 Modified from a figure by Lassila and McGuinness (2001), see also Gomez-Perez et al. (2004), p. 28.



⁵ Protégé: <http://protege.stanford.edu/>.

⁶ OWL, Web Ontology Language: <http://www.w3.org/TR/owl-features/>.

⁷ For example, one of the most popular ontologies in the life sciences field, the Gene Ontology, merely utilises hierarchical relations (is_a and part_of).

⁸ UMLS: <http://umlsinfo.nlm.nih.gov/>.

⁹ Yahoo! Directory: <http://dir.yahoo.com/>.

¹⁰ Other contexts (e.g. philosophy, computer science) bring forward further different definitions, which can not be discussed here.

degree of expressiveness. In the following, we particularly want to focus on supported concept interrelations as the criteria for comparison and for identifying the level of complexity.

Explicit interconnections

The possibility to use self-defined knowledge relations is one major characteristic of ontologies. While folksonomies do not include any type of explicitly stated (paradigmatic) relations (and thus no control over, for example, synonyms or homonyms), for some other systems guidelines regarding the use of relationships or even regulating norms are available (amongst others [DIN 1463/1 1987]). The relations currently considered to be generally applicable and implemented in practice are the following [Bean and Green 2001; Bertram 2005, pp 34; Stock and Stock 2008; Weller and Peters 2007]:

Relations of equivalence handle synonyms and quasi-synonyms (for instance, words that have (almost) the same meaning and are therefore exchangeable in a given context), and are thus particularly important for indexing and documentation contexts, to enable a consistent use of a vocabulary and enhance recall in information retrieval.

Hierarchical relations are the core relations for defining the structure of a knowledge domain. They comprise meronymy (mereology, part-of relation, part-whole relation, paronymy) [for example, Pribbenow 2002] and hyponymy (kind-of-relation, taxonomic relation, taxonomy) [Cruse 2002], so 'lips' is a meronym of 'mouth' and 'duck' is a hyponym of 'bird'.

Associative relations are unspecified but indicated connections between concepts that can have any kind of relation value (except synonyms and hierarchical relations).

Let us see how the most important classical methods of knowledge representation make use of knowledge relationships [Stock and Stock 2008; Weller and Peters 2007]:

Controlled keyword indexing (nomenclature, *Schlagwortmethode*) focuses on capturing synonyms for controlling the vocabulary. Hierarchical constructions are

not indicated in this method (otherwise the emerging set of keywords would form a thesaurus); additional unspecified references may be included.

Classifications consist of non-verbal notations which represent concepts and relations between them, to represent knowledge in a uniform and language-independent way. Classifications are structured hierarchically, without further distinguishing different types of hierarchies.

Thesauri pay much attention to the relation of equivalence, which results in a collection of synonyms as non-descriptors added to the controlled set of descriptors. Furthermore, thesauri split up the hierarchical relation into hyponymy and meronymy and make use of (entirely undifferentiated) associative relations [Aitchison et al. 2004].

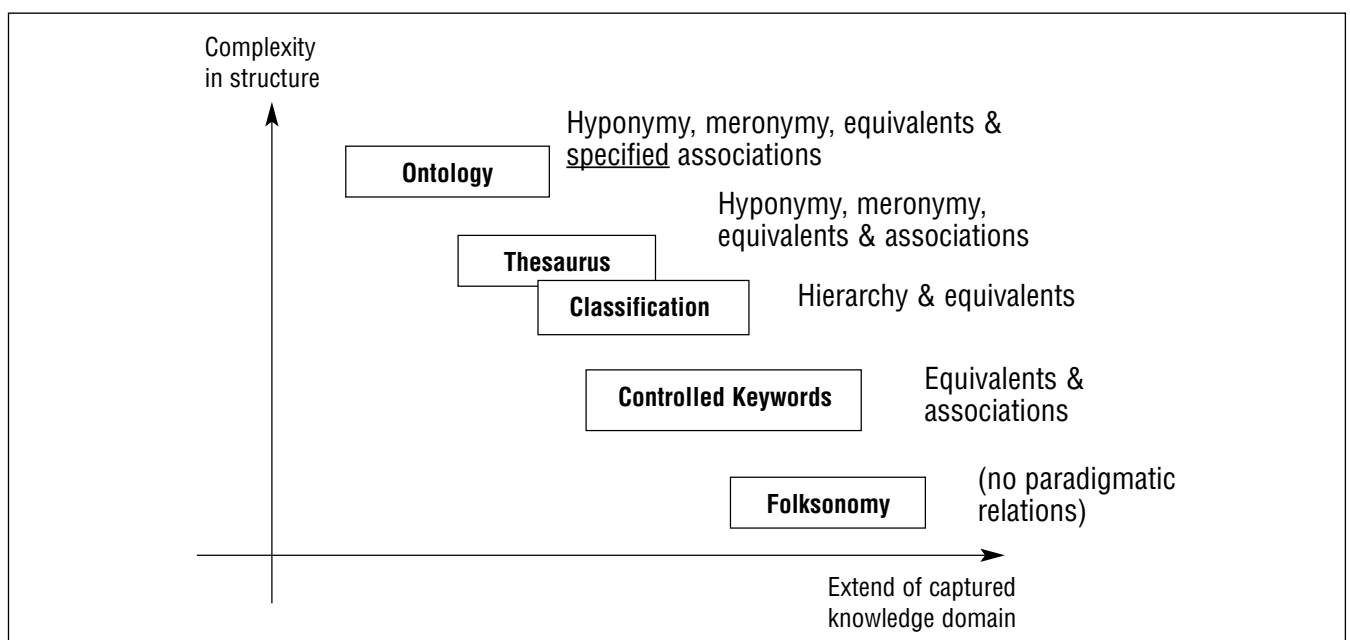
If we take the number and specificity of used relations as characteristics for a scale of knowledge representation systems, the new types of folksonomies and ontologies will occupy both endings of it (see Figure 2). As discussed above, folksonomies have no vocabulary control and thus also do not make use of any paradigmatic relations. Ontologies allow the free modelling of self-defined relations, in other words they split up the associative relation and make the underlying associations explicit. This accuracy will be needed to realise semantic annotations and information integration. On the other hand we believe that fine-grained relation-modeling will only be possible for limited domains of interest.

Elaborated ontology languages such as OWL support detailed knowledge representation and as many kinds of relation types as needed. Relational structures may sometimes be called *properties* (or *slots*), depending on the editor or representation language in use.

Most ontologies make use of hyponymy and meronymy, as they usually constitute the basic structure of an ontology. Hierarchical relationships are often labeled *is_a* (for hyponymy) and *part_of* (for meronymy), but may also be named differently, for example *subclass_of*. Specified associative relations may be very narrowly defined, like *has_ingredient* and reversely *is_ingredient_of* in a food ontology, or *develops_from* for gene

Figure 2:

Approach to classify popular knowledge representation systems according to complexity in relational constructions.



developments in the Gene Ontology [Gene Ontology Consortium 2000]. One challenge of ontology engineering is, to find the adequate level of specificity in modelling the interrelations of a domain. So far there are no common guidelines on how to model knowledge relations in ontologies effectively, although discussions on this have begun [Hovy 2002; Schulz et al. 2006].

Ontologies as knowledge bases

Ontologies are often considered to be a shared conceptualisation of a domain of interest or a specified knowledge base for a community [Gruber 1993 and 2005] – rather than a tool for enhancing retrieval in document collections.

Ontologies may contain explicit information about objects and contexts. This is realised by the use of attributes or restrictions added to the knowledge relations and by including instances as representations of real-life entities. With the Protégé OWL editor for example, relations can be defined as transitive, functional and symmetric. Inverse relations are bound together (*producer_of* and *is_produced_by*). Cardinalities may be applied to relations, for instance by stating that an article must have at least one author or a dog must have exactly four legs.¹¹

Instances in an ontology represent the lowest hierarchical level available. They usually do not represent abstract concepts but rather concrete entities, for example people (like Angela Merkel) or institutions (like Heinrich Heine University). The lowest level may differ according to the planned application area; in an automobile ontology the lowest level may be single car models (Ford Fiesta) or even single concrete cars, like that very Ford Fiesta someone owns.

In a way, the inclusion of instances in an ontology means that the process of indexing is, to a certain extent, incorporated in the ontology engineering process. Sometimes, instances are not viewed as part of an ontology but as an appendix that, together with the ontology, constitutes a knowledge base. Accordingly, some ontologies are not (merely) designed for the detailed description of external knowledge sources, but should rather collect complex information interrelations (and may therefore in some cases even resemble databases rather than vocabularies). Depending on the applications, retrieving this information might then require additional knowledge in query languages, which is not given for typical web-users [Christiaens 2006]. With reasoning mechanisms, information that has not been inserted to the ontology directly may be derived and the consistency may be checked [for example, Gomez-Perez et al. 2004].

Approaches for using and combining folksonomies and ontologies

A lot of new ideas are coming up regarding how to use folksonomies as additions to existing systems, and how to combine different approaches. Tom Gruber asks for 'tagging across various and varied applications' where we can 'reason about the tag data without any one application owning the "tag space" or folksonomy' [Gruber 2005]. The platform Group Me! makes it possible to manage resources

from Flickr and URLs retrieved with Google in one place, as well as tagging and grouping these different documents [Abel et al. 2007]. Other approaches will try to manage the clustering of search results into meaningful units [Gruber 2005]; so far some clustering mechanisms rely on word frequencies and co-occurrences [for example, Grahl et al. 2007; Kolbitsch 2007].

Applications that really make use of complex semantic ontologies are still rare, and this is partly due to the high effort needed to develop and maintain sophisticated ontologies. For this reason, collaborative environments as provided in social tagging may inspire new ways of efficient ontology engineering.

The main goal is to combine the popularity, convenience and flexibility of folksonomies with the semantics and high-quality structures of ontologies. Ideally both approaches inspire each other within a 'continuous feedback loop' [Christiaens 2006].

Ontologies supporting folksonomies

As discussed above, one major problem of folksonomies is the lack of control in the used vocabulary. Ways of introducing some form of semantic control in tagging systems are currently being explored [by, for instance, Michlmayer 2005].

The least promising approach is to provide web users with an existing ontology and tell them to annotate documents on this basis. This is unlikely to be very successful as it takes away the freedom and convenience from social tagging and demands training. But ontologies may be applied behind the scenes of social tagging, for example by providing recommendations for related tags which a user may add to his already entered keywords. Tag recommendations do not have to be based on an underlying ontology, they may also be calculated from users' tagging behaviour (co-occurrences)¹². But if an ontology is used, the nature of the suggested tags can be made explicit, which helps the user judge their appropriateness. For example, if a user types the tag 'Tottenham', an ontology-based system might suggest to also use the *upper-term* 'London'; another user choosing the tag 'folksonomy' might be provided with the information that a folksonomy *is_used_by* 'social software' and can then decide whether or not to add this tag.

Another option is to use ontologies (or other structured knowledge representations) for query expansion mechanisms within social tagging platforms. Search queries over folksonomy tags may be (automatically) enhanced with semantically related terms derived, for example, from an ontology. For example, WordFlickr¹³ expands query terms with the help of relational structures in WordNet¹⁴ to perform enhanced queries in the Flickr database [Kolbitsch 2007]. Users submitting a query to WordFlickr may choose which types of relations (synonyms, hypernyms, hyponyms, holonyms or meronyms) should be used for expanding the query. Thus, if a user searches for 'shoes' the query may be expanded with the hyponyms 'slippers' and 'trainers' to retrieve pictures tagged with these subtypes of shoes from the Flickr collection.

Folksonomies supporting ontologies

As ontologies are supposed to be both complex and highly domain-specific, their production and implementation is currently costly and laborious. Thus, ontology engineering

¹¹ In OWL dialect OWL-DL (description logics).

¹² Similar to recommendations in Google Suggest: <http://www.google.com/webhp?complete=1&hl=en>.

¹³ WordFlickr: <http://www.kolbitsch.org/research/wordflickr/>.

¹⁴ WordNet: <http://wordnet.princeton.edu/>.

may profit from the growing communities that are tagging web documents and are becoming aware of the use of metadata.

Analyses of how people tag documents on the web might lead to a better understanding of how humans organise and process information [Lodwick 2005]. This may be of general value for providing guidelines for effective ontology construction. Folksonomies provide insight into the vocabulary used to annotate documents for private organisational purposes and can capture up-to-date language use. A comparison of social tags and terms from a controlled vocabulary for a given domain can be performed. This helps to update existing systems and to evaluate the timeliness, perceivability and suitability of a knowledge representation system designed by professionals [Christiaens 2006; Macgregor and McCulloch 2006; Mika 2005; Zhang, Wu, & Yu, 2006]. Term frequencies and distributions can be used as suggestions for new controlled terms [Peters and Stock 2007]. This method may be used for ontologies as well as for other knowledge representation systems. What is needed is a substantial amount of user-created tags which match the domain of the system to be evaluated; ideally, users tag a document collection that has already been annotated with the controlled vocabulary. Furthermore, a new ontology might even be created on the basis of tag evaluations. In these cases the users doing the tagging do not have to be aware of these processes.

Even more valuable for ontology engineering will be entirely collaborative approaches in which users may actively contribute to the construction of ontologies from the very beginning [Weller 2006]. This means, that a community would actually and actively perform certain steps in an ontology development processes. The collaborative approach of social software mechanisms is adapted. Different levels of collaboration are possible. On a basic level, a community may work with an already existing ontology and simply suggest new concepts or instances that are missing in the ontology. Users may support ontology engineering via feedback and annotation systems and may contribute their individual knowledge to broaden and update the ontology. Much more sophisticated is the task of letting a community design a whole ontology from scratch. For this purpose, a specific platform will be needed that particularly supports the planning and conceptualisation phases within the ontology development. Collaborative ontology engineering is an issue of growing interest. Different ontology editors are under development that will support user collaboration in different ways [for example, Bao and Honavar 2004; Paulsen et al. 2007; Zacharias and Braun 2007].

Folksonomies in professional environments

'Classical' professional (or scientific) databases – such as those provided by such hosts as STN International, Ovid, Thomson Scientific or Questel – are indexed exclusively by professionals.

Folksonomies started as a simple tool for web users to use to organise files on the web. Since then, folksonomies have been used within intranets or other corporate information systems [Fichter 2006; Peters 2006], as well as in professional databases [Stock 2007] and libraries [Kroski 2006; Spiteri 2007].

The library of the University of Pennsylvania has established its own tagging system, PennTags¹⁵. Within this system, users can collect and tag general web bookmarks as well as journal articles and records from the library's online catalogue and videos from its video catalogue [Kroski 2006].

Among professional databases, Elsevier's Engineering Village¹⁶ has become a pioneer in using social tagging. Documents from engineering research databases (including Compendex and Inspec) and patent databases can now be tagged by Engineering Village users.

One way of embedding social tagging into existing systems is to allow user-created tags alongside professional indexing. Thus, a document in a professional database may be professionally indexed (for example, with a thesaurus) first, and users will then be able to add tags for personal organisation. Of course, the origin of each keyword should then be outlined to allow other users to judge its quality. In this way, index terms representing different points of view are created, providing alternative starting points for researches.

Yet, this approach of indexing documents twice might not be appropriate for each provider. Other systems may help to reduce the effort for professional indexers and thus be more cost effective. For this purpose, the layer model as proposed by Krause [1996] and modified for corporate blogs by Peters [2006] should be mentioned. In a first step, documents are classified according to their importance. Next, methods of knowledge representation are chosen according to the importance rating. Only the most important resources are indexed professionally. The less important sources are tagged by users, in a cost-saving way.

More examination of the efficient crossovers between knowledge representation systems [Stock and Stock 2008] and their applicability for retrieval purposes will surely be a focus of future research.

Conclusion

Generally, folksonomies and ontologies can be seen as the two ends of a scale of documentation languages ranging from unstructured to highly formalised systems. Yet they are not to be seen as rivals but rather as elements in a toolbox which can be used together and may provide inspirations for future applications. The key challenge now is to find the right approaches, or combinations of approaches, to support concrete applications. Solutions have to be well balanced in effort and complexity and they must take specific requirements into account. If this is achieved, content providers and online suppliers, as well as private users and corporate information managers, will profit from new indexing methods which are socially or formally enhanced.

Acknowledgements

The author would like to thank her colleagues within the Ontoverse research project and in the Department of Information Science at Heinrich-Heine-University, Düsseldorf. Financial support was provided by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), Bonn, of the Federal Republic of Germany.

¹⁵ PennTag: <http://tags.library.upenn.edu/>.

¹⁶ Engineering Village: <http://www.engineeringvillage.org>.

References

- Abel F, Frank M, Henze N, Krause D, Plappert D, Siehndel P (2007): Group Me! Where Semantic Web meets Web 2.0. In: Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea, Nov. 2007.
- Alexiev V, Breu M et al. (2005): Information Integration with Ontologies. Experiences from an Industrial Showcase, Chichester: Wiley & Sons.
- Al-Khalifa HS, Davis HC (2007): Towards Better Understanding of Folksonomic Patterns. In: Conference on Hypertext and Hypermedia, Proceedings of the 18th Conference on Hypertext and Hypermedia, Manchester, UK: ACM Press, pp.163–166.
- Aitchison J, Bawden D, Gilchrist A (2004): Thesaurus Construction and Use, 4th ed., Aslib: London.
- Antoniou G, van Harmelen F (2004): A Semantic Web Primer, Cambridge, Mass.: MIT Press.
- Bao J, Honavar V (2004): Collaborative Ontology Building with Wiki@nt. A Multiagent Based Ontology Building Environment. In: Proceedings of the 3rd International Workshop on Evaluation of Ontology-based Tools (EON), Hiroshima 2004, S. 1–10.
- Bean CA, Green R (ed.) (2001): Relationships in the Organization of Knowledge. Dordrecht: Kluwer.
- Bertram J (2005): Einführung in die inhaltliche Erschließung. Grundlagen, Methoden, Instrumente. Würzburg: Ergon.
- Breitman K, Casanova MA, Truszkowski W (2007): Semantic Web. Concepts, Technologies and Applications. London: Springer.
- Christiaens S (2006). Metadata Mechanisms: From Ontology to Folksonomy ... And Back. In: Lecture Notes in Computer Science, 4277, pp 199–207.
- Cleveland DB, Cleveland A (2001): Introduction to Indexing and Abstracting, Englewood, Colorado: Greenwood Press.
- Cruse DA (2002): Hyponymy and its Varieties. In: Green, R, Bean, C, & Myaeng, SH (ed.): The Semantics of Relationships. Dordrecht: Kluwer, pp 3–21.
- Davies J, Fensel D, van Harmelen F (AQ what is this? Hrsg., 2003): Towards the Semantic Web. Ontology-Driven Knowledge Management. Chichester: Wiley & Sons.
- 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).
- DIN 1463/1 (1987): Erstellung und Weiterentwicklung von Thesauri. Einsprachige Thesauri. Berlin: Beuth.
- Fichter D (2006): Intranet Applications for Tagging and Folksonomies. In: Online 30 (3), pp 43–45.
- Furnas GW, Landauer TK, Gomez LM, Dumais ST (1987): The Vocabulary Problem in Human-System Communication. An Analysis and a Solution. In: Communications of the ACM, 30, pp 964–971.
- Furnas GW, Fake C, von Ahn L, Schachter J, Golder S, Fox K, Davis M, Marlow C, Naaman M (2006): Why Do Tagging Systems Work? In CHI 06 Extended Abstracts on Human Factors in Computing Systems, S. 36–39. New York: ACM.
- Gene Ontology Consortium (2000): Gene Ontology. Tool for the Unification of Biology. In: Nature Genetics, 25, pp 25–29.
- Golder SA, Huberman BA (2006): The Structure of Collaborative Tagging Systems. Journal of Information Science, 2006 32(2), pp. 198–208.
- Gomez-Perez A, Fernandez-Lopez M, Corcho O (2004): Ontological Engineering. London: Springer.
- Gordon-Murnane L (2006): Social Bookmarking, Folksonomies, and Web 2.0 Tools. In: Searcher. The Magazine for Database Professionals, 14(6), 26–38.
- Grahl M, Hotho A, Stumme G (2007): Conceptual Clustering of Social Bookmarking Sites. In: Proceedings of I-KNOW 07, Graz, Austria, September 5–7, 2007, pp 356–364.
- Gruber T (1993): A Translation Approach to Portable Ontology Specification. In: Knowledge Acquisition 5(2), S. 199–220.
- Gruber T (2005): Ontology of Folksonomy. A Mash-Up of Apples and Oranges, In: First On-Line Conference on Metadata and Semantics Research (MTSR 05), online available: <http://tomgruber.org/writing/ontology-of-folksonomy.htm>.
- Guy M, Tonkin E (2006): Folksonomies. Tidying up Tags? In: D-Lib Magazine 12(1). Available at: <http://www.dlib.org/dlib/january06/guy/01guy.html>.
- Hammond T, Hannay T, Lund B, Scott J (2005): Social Bookmarking Tools (I): A General Review. In: D-Lib Magazine, Jg. 11, H. 4.
- Hovy E (2002): Comparing Sets of Semantic Relations in Ontologies. In: Green, R, Bean, CA, Myaeng, SH, (eds): The Semantics of Relationships. Dordrecht: Kluwer, pp 91–110.
- Kipp ME (2006): Exploring the context of user, creator and intermediate tagging. In: IA Summit 2006, Canada.
- Kolbitsch J (2007): WordFlickr. A Solution to the Vocabulary Problem in Social Tagging Systems. In: Proceedings of I-MEDIA '07 and I-SEMANTICS '07, Graz, Austria, Sept 2007.
- Krause J (1996): Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung. Schalenmodell. Arbeitsbericht Informationszentrum Sozialwissenschaft 6, Bonn.
- Kroski E (2006): The Hive Mind. Folksonomies and User-Based Tagging. Retrieved September 17, 2007 from: <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging>.
- Lancaster FW (2003): Indexing and Abstracting in Theory and Practice, 3rd ed., University of Illinois, Champaigne.
- Lassila O, McGuinness D (2001): The Role of Frame-Based Representation on the Semantic Web. Technical Report KSL-01-02. Knowledge Systems Laboratory, Stanford University, Stanford, California.
- Lodwick J (2005): Tagwebs, Flickr, and the Human Brain. Retrieved 17 September 2007 from <http://www.blumpy.org/tagwebs/>.
- Macgregor, G, McCulloch, E (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. In: Library Review, 55(5), 291–300.

- Madhavan J, Halevy A, Cohen S, Dong X, Jeffery SR, Ko D, Yu C (2006): Structured Data Meets the Web: A Few Observations. In: IEEE Data Engineering Bulletin 29(4): 19–26.
- Mathes A (2004): Folksonomies. Cooperative Classification and Communication Through Shared Metadata. Retrieved 7 May 2007, from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.
- Michlmayr E (2005): A Case Study on Emergent Semantics in Communities. In: Workshop on Semantic Network Analysis, International Semantic Web Conference (ISWC2005), Galway, Ireland, November.
- Mika P (2005): Ontologies are us. A Unified Model of Social Networks and Semantics. In: Proceedings of the Fourth International Semantic Web Conferences (ISWC2005), pp 522–536.
- O'Reilly T (2005): What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. Retrieved 17 September 2007 from: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- Paulsen I, Mainz D, Weller K, Mainz I, Kohl J, von Haeseler A (2007): Ontoverse. Collaborative Knowledge Management in the Life Science Network. In: Proceedings of the German eScience Conference 2007, Max Planck Digital Library, ID 316588.0.
- Peters I (2006): Against Folksonomies. Indexing Blogs and Podcasts for Corporate Knowledge Management. In: Online Information 2006, Conference Proceedings, London: Learned Information Europe Ltd, 2006, 93–97.
- Peters I, Stock WG (2007): Folksonomy and Information Retrieval. In: Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology (Vol. 45) (CD-ROM).
- Pribbenow S (2002): From Classical Mereology to Complex Part-Whole-Relations. In: Green, R, Bean, CA, Myaeng, SH, (eds): The Semantics of Relationships. Dordrecht: Kluwer, pp 35–50.
- Quintarelli E (2005): Folksonomies: Power to the People. Paper presented at the ISKO Italy UniMIB Meeting, Milan, 24 June 2005. Online available: <http://www.iskoi.org/doc/folksonomies.htm>.
- Safari M (2004): Metadata and the Web. In: Webology, 1(2), Article 7. Available at: <http://www.webology.ir/2004/v1n2/a7.html>.
- Schulz S, Kumar A, Bittner T (2006): Biomedical Ontologies: What Part-of Is and Isn't. In: Journal of Biomedical Informatics, 39, pp 350–361.
- Smith G (2004): Folksonomy. Social Classification (Blog Post, 2004-08-03). Retrieved 17 September 2007 from: http://atomiq.org/archives/2004/08/folksonomy_social_classification.html.
- Spiteri LF (2007): Structure and Form of Folksonomy Tags: The Road to the Public Library Catalogue. In: Webology, 4(2), Article 41. Available at: <http://www.webology.ir/2007/v4n2/a41.html>.
- Staab S, Studer R (Hrsg, 2004): Handbook on Ontologies, Berlin, Heidelberg, New York: Springer.
- Stock WG (2007): Folksonomies and Science Communication. A Mash-up of Professional Science Databases and Web 2.0 Services. In: Information Services and Use 27(3).
- Stock WG, Stock M (2008): Wissensrepräsentation. Informationen Auswerten und Bereitstellen. Oldenbourg: München, Wien.
- Vander Wal, T (2005). Explaining and Showing Broad and Narrow Folksonomies (Blog Post 2005-02-21). Retrieved 17 September 2007 from: <http://www.vanderwal.net/random/category.php?cat=153>.
- Weller K (2006): Kooperativer Ontologieaufbau. In: Ockenfeld M (ed.): Content, 28. Online-Tagung der DGI, 58. Jahrestagung der DGI, Proceedings, Frankfurt (Main): DGI, 2006, pp 227–234.
- Weller K, Peters I (2007): Reconsidering Relationships for Knowledge Representation. In: Proceedings of I-Know '07, Graz, 5–7 September, 2007, 501–504.
- Wu M, Zubair M, Maly K (2006): Harvesting Social Knowledge from Folksonomies. In: Proceedings of the 17th Conference on Hypertext and Hypermedia, New York: ACM, pp 111–114.
- Zacharias V, Braun S (2007): SOBOLEO. Social Bookmarking and Lightweight Ontology Engineering. In: Workshop on Social and Collaborative Construction of Structured Knowledge (CKC), 16th International World Wide Web Conference (WWW 2007), Banff, Alberta, Canada, May.
- Zhang L, Wu X, Yu Y (2006): Emergent Semantics from Folksonomies: A Quantitative Study. Lecture Notes in Computer Science, 4090, 168–186.

Contact

Katrin Weller
weller@uni-duesseldorf.de