Joining Knowledge Representation and Knowledge Management in Digital Libraries

Isabella Peters, Wolfgang G. Stock*, and Katrin Weller
Heinrich-Heine-University Düsseldorf, Dept. of Information Science, Universitätsstr. 1
D-40225 Düsseldorf, Germany
* Corresponding author. Mail: stock@phil-fak.uni-duesseldorf.de
Tel: 00492118112913, Fax: 00492118112917

KM and organisational repositories

Keywords

Knowledge Management, Knowledge Organisation Systems (KOS),
Folksonomy, Social Semantic Web, Emergent Semantics

Well, his position is unique. ... The conclusions of every department are passed to him, and he is the central exchange, the clearinghouse, which makes out the balance. All other men are specialists, but his specialism is omniscience. We will suppose that a minister needs information as to a point which involves the Navy, India, Canada and the bimetallic question; he could get his separate advices from various departments upon each, but only Mycroft can focus them all, and say offhand how each factor would affect the other. They began by using him as a short-cut, a convenience; now he has made himself an essential. In that great brain of his everything is pigeon-holed and can be handed out in an instant. Again and again his word has decided the national policy. Sherlock on his brother Mycroft Holmes.

Abstract

The goal of this paper is to assess the potential of different knowledge organisation systems (KOS) to support corporate knowledge management systems (KMS), namely digital libraries. We will discuss how classical KOS (nomenclatures, classification systems, thesauri, ontologies) are able to reflect explicit knowledge in sense of the Semantic Web and will introduce persons as documents and folksonomies as a means for externalising implicit knowledge in sense of the Web 2.0.

Introduction

Mycroft Holmes is one of the first knowledge managers mentioned in the literature (Conan Doyle 1908). In the service of the British government, he managed the information dissemination of all kinds of the requested knowledge – Navy, foreign affairs, financial questions, etc. Obviously, he had had a perfect overview on

- all appropriate documents,
- all relevant access points to the documents,
- the information needs and the information behaviour of his clients and their organisations.

For today’s knowledge management (KM) it is impressive to learn that Mycroft's position is 'unique' and that 'his word has decided national policy.'
What is "knowledge management" and what role do digital libraries play in knowledge management? And what does "knowledge representation" mean? Knowledge management is the science and practice of the administration of internal and external information in institutions. It includes typical tasks such as knowledge

sharing and information dissemination. According to Gust von Loh (2010), knowledge management consists of social methods (e.g., communities of practice, storytelling, and knowledge cafés) and technical tools and systems (e.g., records management, content management, and intranets). An important subtask is the creation and maintenance of a corporate memory. Technically spoken, this means the construction and operation of an information retrieval system. Both, the technical information retrieval system and the stored content form a corporate digital library. After Pejtersen, Hansen and Albrechtsen (2008: 2) 'most contemporary digital libraries are repositories of information wherein contents can be searched or browsed by the user.' To guarantee optimal access to digital libraries, methods and tools of knowledge representation become applied. The main task of knowledge representation is the creation of metadata which can be attached to documents in order to permit searching and browsing of content. To represent the documents' content, knowledge representation compiles several tools for knowledge organization, namely folksonomies, nomenclatures, classification systems, thesauri, and ontologies (Stock and Stock 2008).

Our paper describes a theoretical approach to combine knowledge management with knowledge representation in digital libraries. For knowledge has two dimensions – explicit and implicit knowledge – we have to find knowledge representation methods for both dimensions, which leads to two basic research questions:

- Research question one: Which knowledge representation methods are appropriate to map explicit knowledge?

- Research question two: Which knowledge representation methods are appropriate to map implicit knowledge?

In corporate environments, knowledge management is confronted with myriads of documents which are more or less relevant for storing and being processed.

- Research question three: Are all documents equal in regard to knowledge management? If not, how can we model the documents' space of a company?

Knowledge representation consists of several methods, elaborated Knowledge Organisation Systems (KOS) such as thesauri or ontologies, and user-driven methods as folksonomic tagging as well. The corporate KOS should represent the corporate language. If there is an opportunity for employees to tag documents, they will probably use 'their' corporate language to index these pieces of information.

- Research question four: What kinds of KOS (which can be applied in digital libraries) do exist? Are there differences in the semantic richness of the types of KOS?

- Research question five: Can we construct a corporate KOS by using the employees' tags? Is it possible to automate the process of emergent semantics? Or are we in need of intellectual endeavours?

Knowledge managers in contemporary institutions make heavy use of information technology, store digital documents in huge databases and advice users to work with powerful information retrieval systems. Brown and Duguid, in their book 'The Social Life of Information', call this IT-driven approach 'information fetishism'. Their monograph is 'concerned with the superficially plausible idea ... that information and its technologies can unproblematically replace the nuanced relations between people' (Brown and Duguid 2002: XVI). If we measure progress of knowledge management in organisations in terms of amounts of digital information (in Gigabyte, Terabyte and so on), this leads to the paradigm: the more information stored, the better is the KM system. 'Consequently, 'knowledge bases' can reach maximum capacity very quickly – and at exactly the same time maximum inutility' (Brown and Duguid 2002: XIII). What's going wrong? 'At the root of the problem lie issues of meaning, judgment, sense making, context, and interpretation – issues far beyond a simple search and embedded in social life' (Brown and Duguid 2002: XIV). It looks like that we have forgotten people and their social and cultural contexts. People in 'networks of practice' – such as the famous Mycroft Holmes – have 'practice and knowledge in common' (Brown and Duguid 2002: 141).

People use – or do not use – knowledge management systems (KMS) or corporate digital libraries (DL). Ease of use, perceived usefulness (Davis 1989), trust (Kim and Han 2009) and fun (Knautz, Soubusta and Stock 2010) are main indicators of the users' acceptance of systems. DeLone and McLean (2003), in their information systems success model, add the dimensions of service quality and information quality besides the dimension of technical system's quality. Jennex and Olfman (2006) rename the information dimension by the term 'knowledge quality' and split it into the KM strategy and process, richness and linkages. Jennex, Smolnik and Croasdell (2007) put the knowledge itself into the focal point of knowledge management and knowledge management systems. 'KM and KMS success are a multidimensional concept. Each includes capturing the right knowledge, getting the right knowledge to the right user, and using this knowledge to improve organisational and/or individual performance. KM success is measured using the dimensions of impact on business processes, strategy, leadership, efficiency and effectiveness of KM processes, efficiency and effectiveness of the KM system, organisational culture, and knowledge content' (Jennex, Smolnik and Croasdell 2007). How is it possible to store the 'knowledge content', the 'right knowledge', in a knowledge management system, so that the user will improve her or his institution's performance by using this KMS? There are two sub-tasks:

- to represent the 'right' knowledge (and only the right one) in the KMS or DL,
- to search and find the 'right' knowledge (and only this) by using the KMS or DL.

From an information science point of view, the first task leads to theories, methods and tools of knowledge representation (Stock and Stock 2008), the second one to information retrieval (Stock 2007a). But what does the word 'right' mean?

Explicit and implicit knowledge (research questions one and two)

What is knowledge in KM, DL and KMS? In a first approximation, we can define 'knowledge' philosophically as a true proposition (e.g. Chisholm 1989). But there are more kinds of knowledge besides propositions, which are true. Ryle warns, 'it is a ruinous but popular mistake to suppose that intelligence operates only in the production and manipulation of propositions, i.e., that only in ratiocinating are we rational' (Ryle 1946: 8). Ryle distinguishes 'knowing that' (propositions) from 'knowing how' (when a person knows to do things of a certain sort). 'Knowing how' has two subspecies, implicit knowledge which can be reconstructed (e.g., know-how on cooking omelettes), and implicit knowledge, which is exhibited only by deeds and not by dicta and which cannot be fully reconstructed (e.g., making good jokes). Polanyi deepens Ryle's observations on knowing how. He states, 'I shall consider human knowledge by starting from the fact that we can know more than we can tell' (Polanyi 1967: 4). This so-called 'tacit knowledge' is embedded in the body of the knowing person. The structure of tacit knowing 'shows that all thought contains components of which we are subsidiary aware in the focal content of our thinking, and that all thought dwells in its subsidiaries, as if they were parts of our body' (Polanyi 1967: X). How can we pass on implicit knowledge? The first possibility is to communicate knowledge bodily. 'The performer co-ordinates his moves by dwelling in them as parts of his body, while the watcher tries to correlate these moves by seeking to dwell in them from outside. He dwells in these moves by interiorizing them' (Polanyi 1967: 30). The second way of passing on tacit knowledge is to try to understand the knower's thought. It is a kind of empathy. 'Chess players enter into a master's spirit by rehearsing the games he played, to discover what he had in mind' (Polanyi 1967: 30).

'Knowing that' and the subspecies of implicit knowledge which can be reconstructed are objects of externalisation; that is, such knowledge can be materialised in documents. The documents containing the knowledge are objects of knowledge representation techniques. If there exist 'shared conceptualizations' (Gruber 1993: 199), it is possible to arrange the knowledge in so-called Knowledge Organisation Systems (KOS). KOS are the basis for any applications of the Semantic Web. In the area of explicit knowledge we are able to work with (externalised) documents and Knowledge Organisation Systems to represent the 'right' knowledge (answer of research question one).

Implicit knowledge, which cannot be reconstructed, will be expressed – if at all – by vague descriptions like analogies or metaphors (Nonaka and Takeuchi 1995). In this case we fail to construct computer-driven KMS. 'Humans make excellent use of tacit knowledge. Anaphora, ellipses, unstated shared understanding are all used in the service of our collaborative relationships. But when human-human collaboration becomes human-computer-human collaboration, tacit knowledge becomes a problem' (Reeves and Shipman 1996: 24). Nonaka and Takeuchi (1995) point out that such tacit knowledge has to be learnt by 'socialisation'. Memmi states, 'in short, know-how and expertise are only accessible through contact with the appropriate individuals. Find the right people to talk or to work with, and you can start acquiring their knowledge. Otherwise there is simply very little you could do (watching videos of expert behaviour is a poor substitute indeed)' (Memmi 2004: 876). It is a task of KM to bring together people with similar interest and with the same or similar skills. For Schreyögg and Geiger KM will be in this case 'skill management' (Schreyögg and Geiger 2003) as well. Implicit knowledge is inseparable connected with persons. Therefore, persons are 'documents' for our KM (this unusual concept of a document, which includes persons and other material objects, is common in information science (Buckland 1997)). In Web 2.0 applications (such as wikis, blogs, social bookmarking services, file sharing services) all kinds of persons can (and do) participate. It would be a misinterpretation that people externalise their tacit knowledge in such applications, but for KM tasks we obtain hints on the topics (not the knowledge itself) of some implicit knowledge. Especially when persons index resources of Web 2.0 services, they use their own (maybe otherwise tacit) concepts and their own words to describe the content of the resource. Indexing within Web 2.0 services is called 'tagging'; all tags of a service form a 'folksonomy'. In KM, we learn about the person's explicit and some aspects of the tacit knowledge by monitoring his or her tagging behaviour. In the area of implicit knowledge we are able to work withpersons as documents and Folksonomies to represent the 'right' knowledge (answer of research question two).

Our basic idea is a mash-up of Web 2.0 technologies (folksonomies) and Semantic Web technologies (Knowledge Organisation Systems), or – in other words – of social knowledge representation methods and semantic knowledge representation methods for the application in digital libraries. Our ambition is to construct social semantic knowledge representation methods, which are able to represent documents (normal documents and persons as well) and therefore explicit knowledge (to the full extent) and some aspects of implicit knowledge.

The shell model of corporate documents (research question three)

We have stressed the term 'document' – even persons are now considered as documents. 'Normal' documents are formally published documents (e.g., scientific works, patents and other documents of intellectual property rights, news articles, legal texts, grey literature), informally published documents (e.g., Web pages, blogs and posts in forums, wikis on the WWW, files like photos, videos, pieces of music on file sharing services), and internal documents (e.g., reports, memos, mails, invoices, individual files). Optimally, documents in KMS are stored in a digital form and are specified by formal aspects (author, year, source, etc.) and by content aspects (applying KOS or folksonomies). This specification of the document is called a 'surrogate' and consists of metadata. As material documents (such as persons or companies) cannot become digitised, we have to work only with the surrogates (Stock 2007a: 82-94).

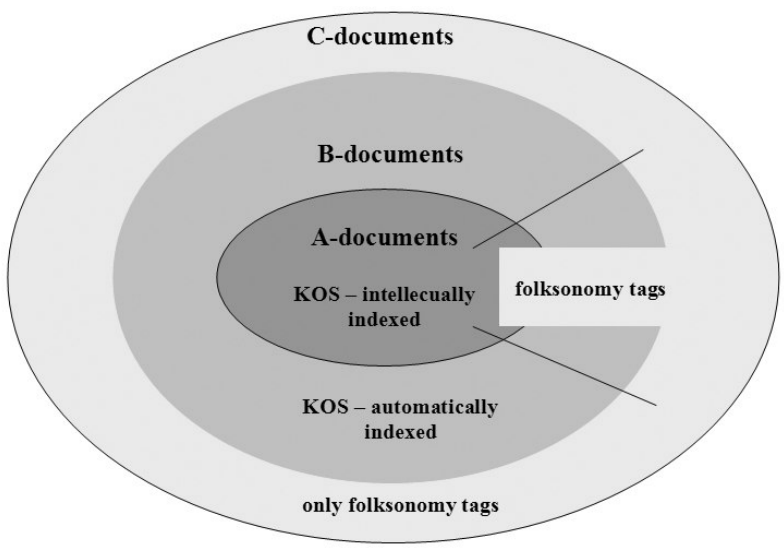


Figure 1. Shell model of corporate documents and preferred methods of indexing.

Internal documents of a company and relevant external documents are more or less important for the institution and so for KM. We will work with a layer model or shell model of document importance, which was introduced into the information science literature by Krause (2006). In a corporate environment (Peters 2006), we have some A-documents or core-documents, which are very important for the institution, say, a strategy paper by the CEO or a fundamental patent describing the company's main invention. Every A-document (normal document or person) has to be found at every time – with a maximum degree of findability. B-documents are not as important as the core documents, but have to be found in certain situations. Finally, C-documents consist of little important information, but they are worth to be stored in a KMS – perhaps some day there is a need to find the given document. With the shell model and the A/B/C documents we are able answer our research question three.

All users of the corporate KMS are allowed to tag all documents. For everyone will tag a given document with his or her specific (even implicit) concepts, the corporate folksonomy will directly reflect the language of the institution. The less important C-documents are indexed only by tags. All other documents are represented in terms of the corporate KOS. Indexing is an elaborate, time-consuming and therefore costly task. So only the A-documents are indexed intellectually by professional indexers, and the B-documents are indexed automatically. If the full texts of the documents are stored as well, there are many additional (but uncontrolled) access points to the documents (full text search).

As the shell model needs KOS for indexing documents of high quality and KOS which reflect the particular language of the company, the concept of emergent semantics within folksonomies is used to generate this specific KOS. The explication of this idea follows in paragraph 5.

	Folksonomy	Nomenclature	Classification	Thesaurus	Ontology
Equivalence	--	yes	yes	yes	yes
Hierarchy	--	--	yes	yes	yes
-- Hyponymy	--	--	--	yes	yes
-- Meronymy	--	--	--	yes	yes
Association	--	--	--	yes	--
Specific Relations	--	--	--	--	yes
Syntagmatic Relations	yes	yes	yes	yes	--

Table 1. Main semantic relations in KOS.

Knowledge Organisation Systems (research question four)

Knowledge Organisation Systems (KOS) consist of (a) concepts and (b) relations between the concepts. A relation is called 'paradigmatic', when two concepts are rigidly coupled in the KOS; a relation is called 'syntagmatic', when two terms co-occur in documents (Peters and Weller 2008a).

In information science and practice there are three classical knowledge organisation systems, differing in their semantic expressiveness: nomenclatures, classification systems, and thesauri (see table 1). Nomenclatures work with controlled terms; they make mainly use of one semantic relation, namely the relation of equivalence (synonymy and quasi-synonymy). A typical example of a nomenclature is the Registry File of the Chemical Abstracts Service (Weisgerber 1997). In classification systems, the concepts are denominated by artificial notations, which are connected by relations of equivalence and (unspecific) hierarchy (Batley 2005). Examples are the Dewey Decimal Classification (DDC) and the International Patent Classification (IPC). In a thesaurus, the concepts are called 'descriptors'. They are related by equivalence, hierarchical relations (hyponymic 'is-a' and meronymic 'part-of' relations), and by association ('see also') (Aitchison, Gilchrist and Bawden 2000). Well-known thesauri are the Medical Subject Headings (MeSH) and the physics thesaurus of INSPEC.

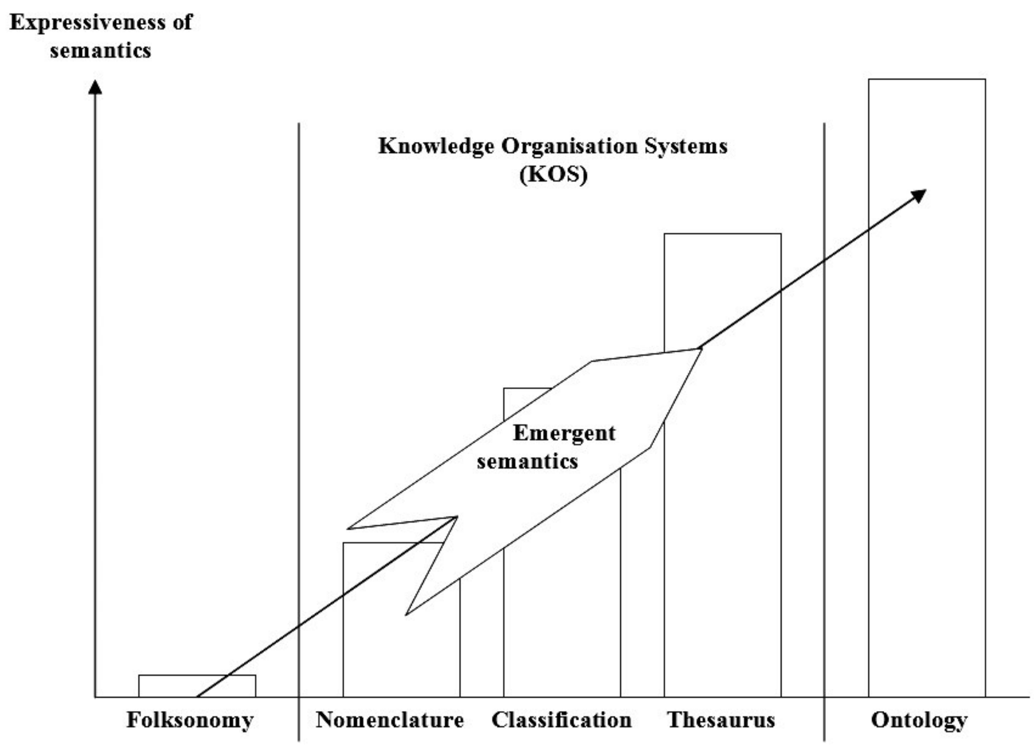
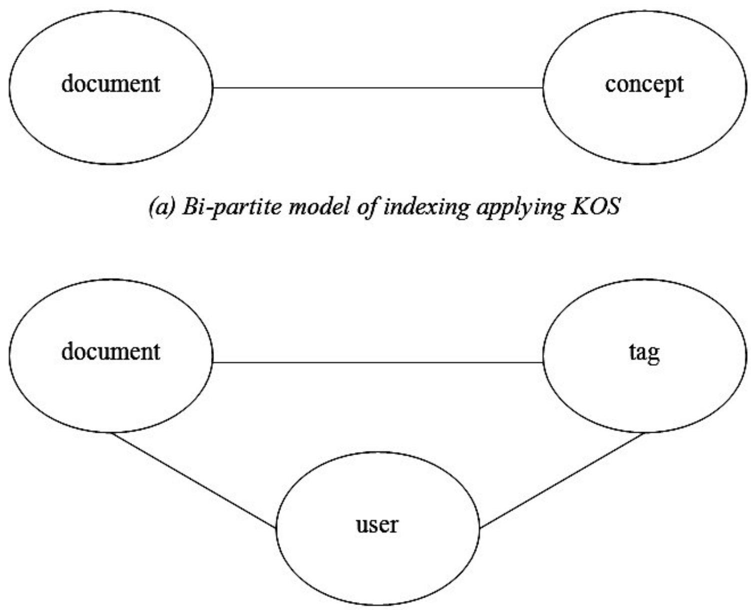


Figure 2. KOS and emergent semantics.

Documents will be indexed through the controlled terms of the KOS, and documents will be retrieved by the same controlled terms – the vocabulary problem (Furnas et al. 1987) in information retrieval is reduced. Adopted from research on Artificial Intelligence, ontologies are a new form of KOS. Ontologies make use of a standardised language (e.g., OWL, the Web Ontology Language), permit automatic reasoning (e.g., adopting description logics), make use of general and individual terms, and do not apply unspecific associative relations, but many specific relations besides hierarchy (Stock and Stock 2008: 256-257). Normally, ontologies do not form syntagmatic relations, because they are not used for indexing documents, but for representing the knowledge itself. The youngest kind of KOS is the folksonomy (Peters 2009). Here users tag documents with uncontrolled terms; there are no indexing rules at all. Folksonomies do not make use of paradigmatic relations, but only of syntagmatic relations, i.e. of tag co-occurrences.

In terms of semantic relations, folksonomies are the weakest form of KOS; ontologies are the most expressive one. The upgrading from one KOS to a richer one (in fig. 2 on the way from left to right) is called 'semantic enrichment' (Angeletou et al. 2007), 'ontology maturing' (Braun et al. 2007) or 'emergent semantics' (Zhang, Wu and Yu 2006). This is our answer of research question four.

The indexing procedure with the aid of concepts taken from a KOS always consists of two elements, namely the document and the concepts (fig. 3a). The concepts are linked to the document through the work of one single professional indexer. Ontologies are exceptions from this proposition because they only link concepts via different relations and do not attach concepts to documents. Indexing with folksonomic tags forms a tri-partite graph between documents, tags, and – what is a new element – users (fig. 3b) (Marlow et al. 2006; Yeung, Gibbins an



(a) Bi-partite model of indexing applying KOS

(b) Tri-partite model of indexing applying folksonomy tags

Figure 3. Indexing models.

This structural change is important because we are now able to gain information about users and we are able to analyse the users' tagging behaviour (which documents did the user tag? which tags did the user apply?). Because of the tri-partite structure of folksonomies we can search and browse for users via tags (e.g., who used this tag at least once for indexing?) and documents (e.g., who saved and indexed this document?) and we are able to recommend users with specific knowledge on requested topics (e.g., who indexed a lot of documents with a specific tag?). On the other hand, users are now able to link themselves to documents via tags and thus allocate documents in specific contexts (Boeije et al. 2009). Of course, in a corporate setting we can additionally create surrogate documents about employees (e.g., yellow pages), which can be indexed by means of KOS concepts and by folksonomy tags through the users of the yellow pages. It is also possible to add the tags (which the user indexed within the KMS) to his profile in the yellow pages to gain a personal description of his interests and activities. In this way, the explicit information about a person is combined with the implicit information about his daily work and interests. As such it is easier for other users to find out of the yellow pages what competencies the person has and if he or she is matching their needs.

Folksonomies show only little semantic structures, which has important implications for the development of KMS using tags as organisation tool. The main problems results from the variability of the natural language and the absence of a controlled vocabulary. In folksonomies we find for example different word forms, nouns in singular, nouns in plural and abbreviations ('www' in fig. 5) (Peters and Stock 2007) – all mixed up. Therefore, in professional environments and for A- and B-documents it is advisable to work with folksonomies only in combination with other KOS-based indexing methods (Peters 2006). The processes of manipulating and re-engineering folksonomy tags and combining them with KOS in order to make folksonomies more productive and effective is described in the literature as 'tag gardening' (Peters and Weller 2008b).

From tags to controlled terms of KOS. Or: From the Social to the Semantic Web (research question five)

Although folksonomies are uncontrolled collections of tags, they typically include syntagmatic relations between them. If these semantic relations can be made explicit during the procedures of tag gardening, we obtain gradually enriched semantics (the more fine-grained the used relations the richer the semantics). Relations between tags can be found out with co-occurrence analyses which may work on two levels. On the one hand it is possible to find co-occurrences on the document level, which means that two tags are somehow related when they are both indexed for the same document. On the other hand tag co-occurrences appear on the folksonomy level, which means that several documents of a folksonomy may be indexed with the same tag and therefore tag co-occurrences of the document level can be cumulated across all documents of the folksonomy. We state that it is not sufficient to calculate co-occurrences to obtain semantic relations. Co-occurrences can only work as indicators that some kind of relations may exist between two concepts.

Yet, to determine which kind of relation is at hand, we have to add intellectual analyses supported by partly automatic suggestions. Figure 5 shows an example of how semantic interrelations may be hidden in folksonomy tags. A resource 'What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software' is saved in the social bookmarking system BibSonomy (www.bibsonomy.org) and was indexed with different tags, e.g. 'web2.0', 'SocialSoftware' and 'tagging'. On the level of this single document, we can detect specific relations such as the part-of relation between 'Web2.0' and 'SocialSoftware'. In order to install semantics, we would now have to render this relation into a paradigmatic one, valid within the whole tagging system or KMS. This semantic enrichment of tags would enable the system's users to broaden a search for all documents about 'Web2.0' in a way, that also documents tagged with 'SocialSoftware' would be included.

Yet, intellectual detection of semantic relations is laborious. What is more, since tagging systems are dynamic and steadily increasing activities for the detection of tag relations have to be executed on a regular basis to guarantee a current reflection of the tagging community's needs and changes. Thus, we are in need of some frequency computations leading to most promising areas for deeper analysis. Candidate tags have to be extracted from the tagging system as starting points for which we propose power tags.

Figure 4. Hidden semantic relations in BibSonomy.

The determination of power tags depends on the distribution of tags regarding the frequency of their assignment to a digital resource. The basic assumption is that different distributions of tags may appear in folksonomies: a) an inverse Power Law distribution, a Lotka-like curve (Huang 2006; Munk and Mork 2007), b) an inverse logistic distribution (Stock 2006, Stock 2007b), and c) other distributions. These distributions can be detected on two levels, the document level (which tags are used for indexing and how often is a particular tag assigned to the document?) and the database level (the overall summarisation of tags per document for the whole data basis). A Lotka-like power law (Egghe and Rousseau 1990) has the form

where *C* is a constant, *x* is the rank of the tag relative to the resource, 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. The discussions about 'collective intelligence' are mainly based on this observation: the first *n* tags of the left hand side of the power law reflect the collective intelligence in giving meaning to the annotated documents (Weiss 2005).
The inverse logistic distribution shows a lot of relevant tags at the beginning of the curve (the 'long trunk') and the known 'long tail'. This distribution follows the formula

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

where e is the Euler number, x is the rank of the tag, C' is a constant and the exponent b is approximately 3. In comparison with the power law the inverse logistic distribution reflects the collective intelligence differently. The curve shows a long trunk on the left and a long tail of tags on the right. Since all tags in the long trunk have been applied with similar (high) frequency, all left-hand tags up to the turning point of the curve should be considered as a reflection of collective intelligence.
For the determination of power tags we have to keep in mind both known tag distributions. If the document-specific distribution of tags follows the inverse power law, the first n tags are considered as 'power tags'. The value of n is dependent on the exponent a. If the tag distribution forms an inverse logistic distribution, all tags on the left-hand side of the curve (up to the turning point) are marked as 'power tags'.

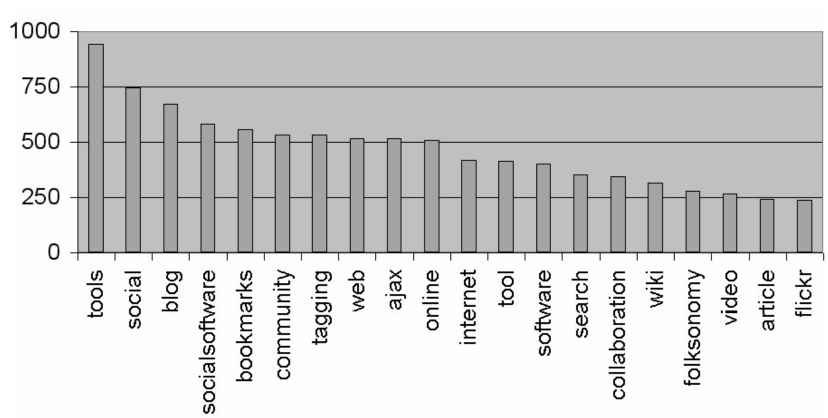


Figure 5. Tag co-occurrences with tag 'web2.0'.

Activities which turn tags into controlled terms need some candidate tags as starting points, for which we propose power tags. Since we prune both curves at a particular point and neglect the rest of the curve and the long tail-tags, respectively, we call this approach for determining tag gardening candidates 'exclusive'. The concrete processing of power tags works as follows: the first step is to determine power tags on the resource level. This calculation is carried out for each and every resource of the collaborative information service. According to the above explanations, two different tag distributions may appear which each identify different numbers of power tags (we call them power tags I). Since these power tags I are important tags in giving meaning to the resource, they have to be processed in the next step. Now, the n numbers of power tags I should each be investigated regarding their relationships to other tags of the whole database – in other words, a calculation of co-occurrence is carried out for the power tags I. This calculation produces again specific tag distributions, where we can determine power tags as well (we call them power tags II). These new power tags II are now the candidate tags for the emergence of semantics since their connection to the power tags I seem to be very fruitful. In this connection the existence of semantic relations between power tags I and power tags II will be most obvious. Let us explain the procedure with an example from the social bookmarking service BibSonomy (see figures 5 and 6):

The tags for the resource displayed in figure 5 form a power law distribution as the tag 'web2.0' is indexed 13 times, 'wismasys0809" (a course-specific tag of our institute) is indexed 4 times, 'imported' and 'web' are each indexed 3 times, '2.0', 'basics', 'folksonomy' and 'technology' are each indexed two times and the rest is used only once (mouse over tags displays indexing frequency of each tag, not displayed in picture). Since 'wismasys0809' is a course-specific tag we exclude it from the further processing and just consider 'web2.0' and 'imported' as power tags I. For those two tags we examine co-occurrences with all other tags of the database. An exemplary search in BibSonomy for the tag 'web2.0' results in the co-occurring tags displayed in figure 6 and says that the tag 'web2.0' appears together with the tag 'tools' in 941 documents.

Those co-occurring tags follow an inverse-logistic distribution. The first *n* tags (say *n = 10*) are now considered as power tags II and are the basis for the intellectual tag gardening activities and the detection of tag relations. Figure 6 shows that the tag 'web2.0' is frequently combined with the tags 'tools', 'social', 'blog' and 'socialsoftware'. Accordingly, the basic assumption for emergent semantics in folksonomies and tag interrelations respectively is that tags which often co-occur have to be of similar meaning or have to be linked in a meaningful relationship. Thus, to name an example of these power tags II, the tags 'tagging' and 'web2.0' form part-of relation where 'tagging' is a part of 'web2.0'. In order to achieve richer semantics, we would thus establish these newly detected hierarchical connections as paradigmatic relations within the folksonomy (which in return grows to become a more formal KOS). The goal is here to enhance precision and recall of search results and the expressiveness of the folksonomy by adding semantics, to enable query expansion during retrieval via semantic relations, to enhance indexing functionalities, to improve navigation within the folksonomy and to build a basis for semantic-oriented visualisations of the folksonomy.

descriptor set			relation
	web2.0		
UF	socialsoftware	used for	synonymy
BT	web	broader term	hierarchy
NTP	blog	narrower term partitive	meronymy
NTP	bookmarks		meronymy
NTP	tagging		meronymy
NTP	community		meronymy
NTP	ajax		meronymy
RT	online	related term	association

Table 2. Descriptor set extracted from tags.

The given assumptions result in a set of descriptors with which we are now able to build a thesaurus or other kinds of KOS. This thesaurus can then be used as basis for the indexing of documents, e.g. via recommendations of tags or descriptors during indexing, or the retrieval of documents, when it is used for query expansion. According to above example, a set of descriptors may look like table 2. Here we are able to answer research question five. Yes, it is possible to construct a corporate KOS by using the employees' tags. But: no, it is not possible to automate this process of emergent semantics to the full extend. We are always in need of some intellectual endeavours.

So, if a user is indexing a resource with the tag 'web2.0' the KMS may suggest 'tagging' as additional or substitutional tag and aids the user in finding appropriate index tags. The same holds for retrieval with tags, when the KMS recommends the tag 'blog' instead of 'web2.0' for query expansion because the user gains too many document when searching with the first term.

Conclusion

Ease of use, perceived usefulness, trust, fun, service quality and knowledge quality are the factors which determine the acceptance of users concerning a system, e.g. a knowledge management system (KMS) or a digital library (DL). We discussed how to improve the knowledge quality of KMS. In sense of information science this leads to two main aspects which both have to be considered: 1) representing knowledge in KMS via theories, methods and tools of knowledge representation, and 2) finding knowledge via theories, models and tools of information retrieval. Because knowledge representation and information retrieval have a reciprocal relationship; considerations on one of them directly affect considerations on the other. So we claimed that both sides of the same coin have always to be considered in combination.

The distinction between explicit and implicit knowledge is important for knowledge representation and information retrieval as explicit knowledge can directly be found in documents and thus can be represented by formal Knowledge Organisation Systems (KOS), in an elaborated form sometimes called Semantic Web. Implicit knowledge belongs to people. In this case, people have to be treated as documents: they need to be connected to exchange their implicit knowledge. In KM this should be fulfilled by means of KMS, but in KMS usually people are not considered as documents. In Web 2.0 environments, people are placed in the center of attention – a way of thinking which could be transferred to KMS. In Web 2.0 people turn implicit knowledge into explicit knowledge when they add comments to documents or index documents with tags. Comments or tags are each manifestations of the person’s thoughts and knowledge.

Therefore, we proposed a mash-up of Web 2.0 and Semantic Web techniques to simultaneously represent explicit knowledge via KOS and implicit knowledge via folksonomies. A simple way of combining Web 2.0 and Semantic Web is the application of a shell model which allows different indexing methods and qualities for different sorts of documents. Because the shell model needs KOS for indexing documents of high quality and KOS which reflect the particular language of the company, the approach of emergent semantics within folksonomies is used to generate this specific KOS. We discussed how to enrich folksonomies with semantic relations in order to achieve semantically richer results in indexing and along with it more precise results in information retrieval. Thus, we proposed co-occurrence analyses based on power tags I and power tags II to gain candidate tag pairs which are fruitful for detection of semantic relations and for the creation of descriptor sets. In this way we turned simple tags as manifestations of implicit knowledge into more formal descriptors which then change into explicit knowledge of the KMS.The limitation of this research is that it only discusses a theoretical approach. We did not yet perform practical implementations to evaluate the formulated methods which should be conducted in future research.

References

Aitchison, J., Gilchrist, A., and Bawden, D. 2000. *Thesaurus Construction and Use* (4th Ed.). London, UK, New York, NY: Europa Publ.

Angeletou, S., Sabou, M., Specia, L. and Motta, E. 2007. Bridging the Gap Between Folksonomies and the Semantic Web. An Experience Report, pp. 30-32. In *Bridging the Gap between Semantic Web and Web 2.0, European Semantic Web Conference.*

Batley, S. 2005. Classification in Theory and Practice, Oxford: Chandos.

Boeije, R., de Vries, P., Kolfschoten, G.L., and Veen, W. 2009. Knowledge Workers and the Realm of Social Tagging. In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS-42)*. IEEE Computer Society.

Brown, J.S., and Duguid, P. 2002. *The Social Life of Information* (2nd Ed.). Boston, MA: Harvard Business School Press.

Braun, S., Schmidt, A., Walter, A., and Zacharias, V. 2007. The Ontology Maturing Approach for Collaborative and Work Integrated Ontology Development: Evaluation Results and Future Directions, pp. 5-18.

In Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution. ESOE, Busan. Buckland, M.K. 1997. What Is a Document? Journal of the American Society for Information Science 48: 804-809.

Chisholm, R.M. 1989. *Theory of Knowledge* (3rd Ed.). Englewood Cliffs: Prentice-Hall International.

Davis, F.D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13(3): 319-340.

Conan Doyle, A. 1908. The Adventure of the Bruce-Partington Plans. *The Strand Magazine* 35(215).

DeLone, W.H., and McLean, E.R. 2003. The DeLone and McLean Model of Information Systems Success: A Ten-year Update. *Journal of Management Information Systems* 19(4): 9-30.

Egghe, L., and Rousseau, R. 1990. *Introduction to Informetrics*. Amsterdam: Elsevier.

Furnas, G.W., Landauer, T.K., Gomez, L.M., and Dumais, S.T. 1987. The Vocabulary Problem in Human-System Communication. *Communications of the ACM* 30(11): 964-971.

Gruber, T.R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5(2): 199-220.

Gust von Loh, S. 2010. *Evidenzbasiertes Wissensmanagement* [Evidence-Based Knowledge Management]. Wiesbaden: Gabler.

Huang, H. 2006. Tag Distribution Analysis Using the Power Law to Evaluate Social Tagging Systems: A Case Study in the Flickr Database, pp. 14-15. In *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop*.

Jennex, M.E., and Olfman, L. 2006. A Model of Knowledge Management Success. *International Journal of Knowledge Management* 2(3): 51-68.

Jennex, M.E., Smolnik, S. and Croasdell, D. 2007. Towards Defining Knowledge Management Success.

In *Proceedings of the 40th Hawaii International Conference on Systems Sciences (HICSS-40)*. IEEE Computer Society.

Kim, B., and Han, I. 2009. The Role of Trust Belief and its Antecedents in a Community-Driven Knowledge Environment. *Journal of the American Society for Information Science and Technology* 60(5): 1012-1026.

Knautz, K., Soubusta, S., and Stock, W.G. 2010. Tag Clusters as Information Retrieval Interfaces.

In *Proceedings of the 43th Annual Hawaii International Conference on System Sciences (HICSS-43)*. IEEE Computer Society Press.

Krause, J. 2006. Shell Model, Semantic Web and Web Information Retrieval. In *Information und Sprache*, pp. 95-106, edited by Harms, I., Luckhardt, H.D., and Giessen, H.W. Munich: Saur.

Marlow, C., Naaman,, M., Boyd, D., and Davis, M. 2006. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read. In *Proceedings of the 17th Conference on Hypertext and Hypermedia*, pp. 31-40. New York: ACM.

Memmi, D. 2004. Towards Tacit Information Retrieval. In *RIAO 2004 Conference Proceedings*, pp. 874-884. Paris: Le Centre de Haute Études Internationales d'Informatique Documentaire.

Munk, T., and Mork, K. 2007. Folksonomy, the Power Law and the Significance of the Least Effort.

Knowledge Organisation 34(1): 16–33.
Nonaka, I., and Takeuchi, H. 1995. The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation., New York, NY, Oxford, UK: Oxford University Press.

Pejtersen, A.M., Hansen, P., and Albrechtsen, H. 2008. A Work-Centred Approach to Evaluation of a Digital Library: Annotation System for Knowledge Sharing in Collaborative Work. *World Digital Libraries* 1(1): 1-17.

Peters, I. 2006. Against Folksonomies. Indexing Blogs and Podcasts for Corporate Knowledge Management.

Preparing for Information 2.0. Proceedings of Online Information Conference London, pp. 93-97. London: Learned Information Europe.

Peters, I. 2009. Folksonomies: Indexing and Retrieval in Web 2.0 (Knowledge and Information. Studies in Information Science, Vol. 1). Munich, Germany: DeGruyter Saur.

Peters, I., and Stock, W.G. 2007. Folksonomies and Information Retrieval. In *Proceedings of the 70th ASIS&T Annual Meeting* (CD-ROM), pp. 1510-1542.

Peters, I., and Weller, K. 2008a. Paradigmatic and Syntagmatic Relations in Knowledge Organisation Systems. *Information – Wissenschaft und Praxis* 59(1): 100-107.

Peters, I., and Weller, K. 2008b. Tag Gardening for Folksonomy Enrichment and Maintenance. *Webology* 5(3).

Polanyi, M. 1967. *The Tacit Dimension*. Garden City, NY: Doubleday (Anchor Books).

Reeves, B.N., and Shipman, F. 1996. Tacit Knowledge. Icebergs in Collaborative Design. *SIGOIS Bulletin* 17(3): 24-33.

Ryle, G. 1946. Knowing How and Knowing That. *Proceedings of the Aristotelian Society* 46: 1-16.

Schreyögg, G., and Geiger, D. 2003. *Kann die Wissensspirale Grundlage des Wissensmanagement sein?* (Can the Knowledge Spiral be a Foundation of Knowledge Management?)'. Diskussionsbeiträge des Instituts für Management. Neue Folge 20/03. Berlin: Freie Universität.

Stock, W.G. 2006. On Relevance Distributions. Journal of the American Society for Information Science and Technology 57(8): 1126–1129.

Stock, W.G. 2007a. *Information Retrieval*. Munich, Germany: Oldenbourg.

Stock, W.G. 2007b. Folksonomies and Science Communication: A Mash-Up of Professional Science Databases and Web 2.0 Services. *Information Services & Use* 27: 97–103.

Stock, W.G., and Stock, M. 2008. *Wissensrepräsentation* (Knowledge Representation). Munich, Germany: Oldenbourg.

Weisgerber, D.W. 1997. Chemical Abstracts Service Chemical Registry System: History, Scope, and Impacts. *Journal of the American Society for Information Science* 48: 349-360.

Weiss, A. 2005. The Power of Collective Intelligence. *netWorker* 9(3): 16-23.

Yeung, C.M.A., Gibbins, N., and Shadbolt, N. 2006. Mutual Contextualization in Tripartite Graphs of Folksonomies. *Lecture Notes in Computer Science* 4090: 168-186.

Zhang, L., Wu, X., and Yu, Y. 2006. Emergent Semantics from Folksonomies: A Quantitative Study. *Lecture Notes in Computer Science* 4090: 168-186.