Social Semantic Bookmarking

Simone Braun, Valentin Zacharias, Hans-Jörg Happel

FZI Research Center for Information Technology, Haid-und-Neu-Str. 10-14,
76131 Karlsruhe, Germany
{braun, zacharias, happel}@fzi.de

Abstract. In this paper we present the novel paradigm of Social Semantic Bookmarking. Social Semantic Bookmarking combines the positive aspects of semantic annotation with those of social bookmarking and tagging while avoiding their respective drawbacks like the cumbersome maintenance of ontologies or the lacking semantic precision of tags. Social semantic bookmarking tools allow for the annotation of internet resources based on an ontology and the integrated maintenance of the ontology by the same people that use it. We introduce Social Semantic Bookmarking and present the SOBOLEO application as an implementation of this paradigm.

Keywords: Social tagging, semantic tagging, social bookmarking, SOBOLEO

1 Introduction

A big challenge for today's internet users is the focused discovery of new information that is likely to be interesting and useful as well as the rediscovery of information that they had once found and identified as such. Social bookmarking systems (e.g. such as del.icio.us[1]) allow for the collection, management, and sharing of bookmarks, i.e., references to such information entities. The users can easily annotate these bookmarks with arbitrary tags that help in organizing, navigating and searching the bookmark collection.

These tags, however, are completely unstructured. Problems such as homonyms, synonyms, multilinguality, typos or different ways to write words, and tags on different levels of abstraction hamper search and retrieval; in particular in complex domains. Replacing tags with semantic annotations based on an ontology as a controlled vocabulary can help here.

Many systems that allow for annotating documents with respect to ontologies struggle, however, with a number of problems, too. Not only are they cumbersome to use but they also view ontology creation as a process separate from its use, performed by people different from those that use it. These systems also often assume that the ontology stays unchanged for longer periods of time and is updated only seldomly. All this leads to unsatisfied users being confronted with out-of-date, incomplete,

[1] http://delicious.com

inaccurate and incomprehensive ontologies that they cannot easily use for annotation; this problem is particular acute in fast changing domains [9].

The novel paradigm of Social Semantic Bookmarking combines the positive aspects of semantic annotation with those of social bookmarking while avoiding their respective drawbacks. Social semantic bookmarking tools allow for the annotation of internet resources with respect to an ontology and the integrated maintenance of the ontology by the same people that use it. Through the use of state-of-the-art web technologies such as bookmarklets and AJAX (e.g., for auto complete functionality), these systems make ontology-based annotation of web documents as simple as tagging. Through easy-to-use, lightweight web ontology editors that are integrated into the system, the barrier between ontology creation and use is removed; users who annotate with the help of the ontology are the same who continuously evolve this ontology. Because internet resources are annotated with concepts (and not keywords), the problems of homonyms, synonyms etc. are avoided.

We present Social Semantic Bookmarking using the example of our system SOBOLEO (Social Bookmarking and Lightweight Engineering of Ontologies) – a system combining the above mentioned features with an innovative search engine and functionality supporting the discovery of experts on specific topics based on their interaction with the system. We also shortly discuss other social semantic bookmarking systems such as Bibsonomy, int.ere.st, GroupMe!, Fuzzy, and Annotea.

2 Social Tagging vs. Semantic Annotation

2.1 Social Tagging and Its Problems

Social bookmarking systems allow their users for annotating bookmarks with several arbitrary tags they find most suitable for describing them. In this way – in contrast to the traditional folder structure like browser favorites – users can organize their bookmarks according to more than one category. This facilitates the organization, navigation, and search in the bookmark collection. The popularity of such social tagging applications have shown that this organizing principle with tags and folksonomies evolving from these is much easier accessible for users than structured and controlled vocabularies; in particular for collaborative applications.

These applications, however, often reach their limits because of lacking semantic precision of tags. Folksonomies have only very limited structure. Their missing semantic precision hampers efficient search and retrieval support, in particular in complex domains, because of problems like the following (cf. [6,7]):

- **(Mis-)Spelling:** The most obvious problem is that tags are simply misspelled or written in different ways because of occurring plurals, abbreviations or compound words, e.g. 'spagetti' vs. 'spag**h**etti', 'noodle' vs. 'noodle**s**', or 'spaghettiCarbonara' vs. 'spaghetti_**c**arbonara'.
- **Multilingualism:** Tags only relate to one language. That means, especially in Europe with many different languages, users have to annotate a resource with many tags in different languages, e.g. with 'pasta', 'noodles', and 'Nudeln', in order

to ensure that other users will find it later on (e.g. to promote their own great spaghetti recipe).

- **Polysemy:** Tags can have several similar meanings. This leads to search results with low precision because of irrelevant resources; e.g. with the tag 'pasta' the users can think of a dish that contains pasta as its main ingredient or of the aliment itself as shaped and dried dough made from flour and water and sometimes egg.
- **Homonymy:** The problem of homonymy is comparable to the problem of polysemy. However, in this case, one tag can have several totally different meanings. This also leads to irrelevant results as all resources that relate to these different meanings are annotated with the same tag. For instance the word 'noodle' can have the meaning of an aliment but also of a swearword for a human head.
- **Synonymy:** Resources are not found because they are annotated with another tag with the same meaning, e.g. with the tag 'vermicellini' instead of 'spaghettoni'. Similar to mulitlingualism, the users have to annotate the resources with many synonymous tags in order to ensure the retrieval by other users.
- **Mismatch of abstraction level:** Also a typical search problem emerges because tags are specified on different abstraction levels, i.e. either too broad or too narrow. This problem, also known as the "basic level phenomenon" [16], can be traced back to different intentions and expertise levels of the users. For instance, one user tags a resource on the basic level with 'spaghetti', another with 'noodles' and a third differentiates 'spaghetti' from 'bigoli' (thicker spaghetti) and 'vermicelli' (thinner spaghetti). A resource annotated with 'spaghetti', however, cannot be found with the search term 'pasta'.

2.2 Semantic Annotation and Its Problems

Replacing tags with semantic annotations based on an ontology promises to solve the limits of (linguistic) tagging-based applications. Ontologies, as formalizations of a shared understanding of a community [8], contain background knowledge of a certain domain. They improve the description or the retrieval of resources (in its broadest sense) by making subject, creation, usage, relational or other context of these resources explicit.

These semantic annotation approaches also rely on the use of some standardized formal language for representing the ontology, such as RDF [14], SKOS [15], or one of the OWL languages [5]. They have a number of potential benefits:

- **Better Retrieval:** The formally represented relations between the concepts in the ontology can be used to offer superior browse or query facilities. In the case where a powerful language like OWL is used, queries may even be answered using reasoning algorithms.
- **Better Use of Annotation:** The availability of machine understandable context for the used annotation terms can be utilized to make better use of the annotation; e.g. information that some annotations represent geographic locations for which a latitude and longitude is known can be used to show the annotated document in a map or to make them available based on the users current location.
- **Better Quality Assurance:** The information contained in the ontology about concepts used for annotation can enable checks on whether an annotation is likely

to make sense; this can help to catch errors early. Also changes in the ontology can be checked whether they violate its integrity.

- **Better (Semantic Web) Integration:** The ontology that is used in the annotation is usually assumed to be also used in other systems and the common usage of the ontology can enable the integration of data created and managed in these diverse systems. Another related aspect is that semantically annotated data can become part of the Semantic Web and then Semantic Web aware agents and applications can make use of it.
- **Better Support of Vocabulary Management:** Through the use of standardized languages to represent the ontologies, these approaches can rely on a landscape of tools that is available to create, manage and evolve these ontologies.

Many systems that allow for annotating with respect to ontologies, however, have not found widespread adoption yet and struggle with a number of problems, too. To a large extend because the annotation process, i.e. the usage of the ontology, and the creation of the ontology are two separate processes, performed by a different set of people. Annotation is done by the users of a semantic application and the ontologies are created by dedicated knowledge engineering specialists. However, separating the use and the creation of the ontology and involving knowledge engineering specialists is causing a number of problems:

- **High Cost:** Knowledge engineers are highly paid specialists, and their effort comprises not only the actual implementation of the domain ontology, but also learning about and understanding the domain of interest. While in many Web 2.0 scenarios a large amount of work is done for free by users interested in the result, this is unlikely to work when knowledge engineers with little innate interest in the domain in question are involved.
- **Domain Errors**: Knowledge engineers are specialists for the domain of knowledge formalization – not for the domain that is being formalized. For this reason they will not have an understanding of the domain comparable to that of domain experts, this limited understanding may cause errors in the resulting ontology [2].
- **Heavyweight Process and Upfront Investment**: Because annotation cannot start without an available ontology, there needs to be an upfront investment to finance the development of this ontology, which includes a systematic requirements elicitation phase. During the usage phase of the ontology, there also needs to be a accompanying process to collect newly emerging requirements, bugs and other change requests and to implement them into a newer version of the ontology.
- **High Time Lag**: There will always be some time lag between the emergence of a new concept and the time when it is included in the ontology and can eventually be used. This time lag is relatively large, when the users of the ontology cannot make the change themselves but must rely on knowledge engineers understanding the requirement, implementing it and finally rolling out the new version of the ontology. In fast moving domains this time lag can quickly get so big that the ontology as a whole becomes unusable [7].
- **Low Appropriateness and Understandability:** An ontology is appropriate for a task if it enables the users to reach their goals more quickly. However, having different people using and developing the ontology makes reaching appropriateness of the ontology much harder. A particular challenge is to ensure

that the ontology is at the right level of abstraction to be understood by the domain experts.

3 Social Semantic Bookmarking

In the previous sections we have seen that (linguistic) social tagging approaches, while popular, struggle with problems such as polysemy, multilingualism or abstraction level mismatches. At the other end many current semantic annotation approaches struggle with the problem of timely updates and appropriateness of the underlying ontology as well as affordable creation. Social Semantic Bookmarking now combines the benefits of tagging with semantic annotation in order to address their respective weaknesses.

Social semantic bookmarking systems allow for the annotation of resources (e.g. web pages, documents) with concepts whose definition and description also evolves collaboratively within the same system. Similar to tagging approaches, they allow for creating new concepts whenever a need arises. Unlike these approaches, concepts can have powerful descriptions and can be interlinked; for example allowing the system to understand that 'swimming bath' and 'swimming pool' are synonyms for the same concept. These powerful concept descriptions are similar to those used in traditional semantic annotation, but social semantic bookmarking allows for adding and changing concepts permanently and easily at the time the concepts are used.

The SOBOLEO[2] system [17] is a particular social semantic bookmarking system that will be used to further illustrate this approach in this section. SOBOLEO is based on AJAX technology and works in most current browsers – thus does not require any local installation. It consists of four application parts: an editor for the modification of the shared ontology, a tool for the annotation of internet resources, a semantic search engine for the annotated internet resources, and an ontology browser for navigating the ontology and the bookmark collection.

SOBOLEOs functionality and the concept of Social Semantic Bookmarking will be further described with an example of a user who annotates an internet resource with a new concept 'sphaghetti', then adds some information about this new concept. A different user will then search for the annotated resource at a different level of abstraction and find it using the semantic search feature.

3.1 Annotation

The annotation process starts when a user finds an interesting resource that she wants to add to the shared repository. In this example a user discovers a tasty pasta recipe. In order to annotate the document the user clicks on a bookmarklet in her browser which opens up the small dialog window (see Fig. 1). The user can annotate the web document using any of the concepts already known to the system and is supported by auto completion in doing that. Here the user also adds a new concept named

[2] http://www.soboleo.com

'Spaghetti' – adding a concept is seamlessly done by simply entering a term that is not yet known to the system.

Once the user clicks save, the system stores the URL of the document with all assigned concepts; any new concepts are also added to the shared ontology of the repository. The SOBOLEO system crawls the content of the annotated web page that is added to a full text index associated with a repository.

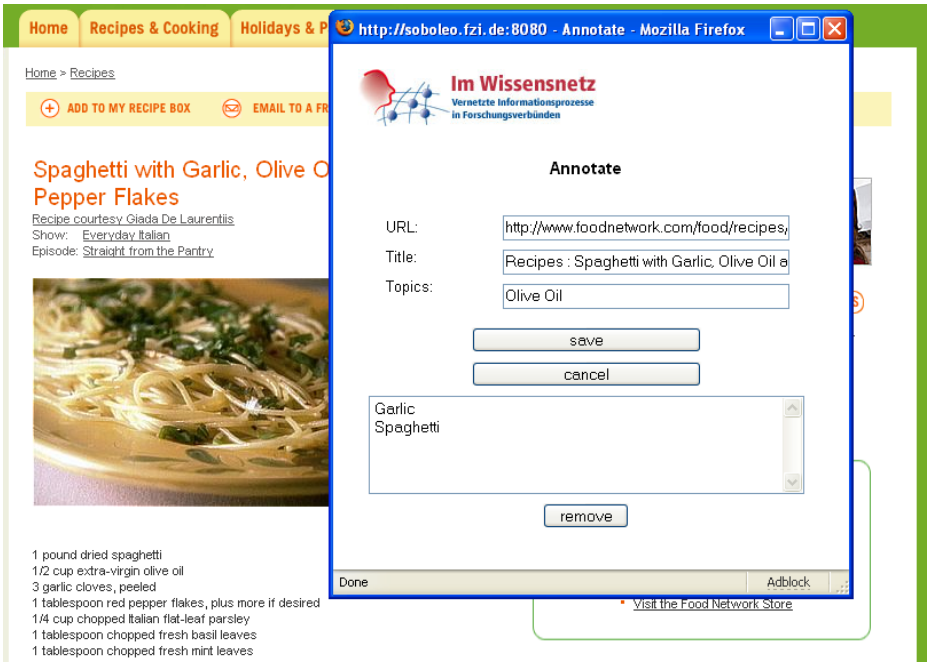

Fig. 1. Annotating a web page.

3.2 Ontology Editing

Each user of SOBOLEO belongs to a user group that has a shared repository containing the annotations and the ontology. Such a user group consists of people working on the same topic, such as a department in a large company or a special interest group spanning continents.

The ontology in the shared repository is represented using a subset of the SKOS standard; it allows for concepts with a preferred label, a description and any number of alternative labels. It also allows for broader, narrower, and related relations between concepts. The ontology in this shared repository is edited using the AJAX editor (see Fig. 2). The editor is a collaborative realtime AJAX editor; i.e., it can be used by multiple persons simultaneously in their respective browsers with edits showing up for the others in realtime.

In the example the user opens the editor to add more information about the new 'Spaghetti' concept. First the user uses the mouse to drag the 'Spaghetti' concept onto the 'Pasta' concept, quickly establishing the relation that 'Spaghetti' is a narrower concept than 'Pasta'. She also adds a short description to 'Sphaghetti' and 'Spaghetto' as synonym.

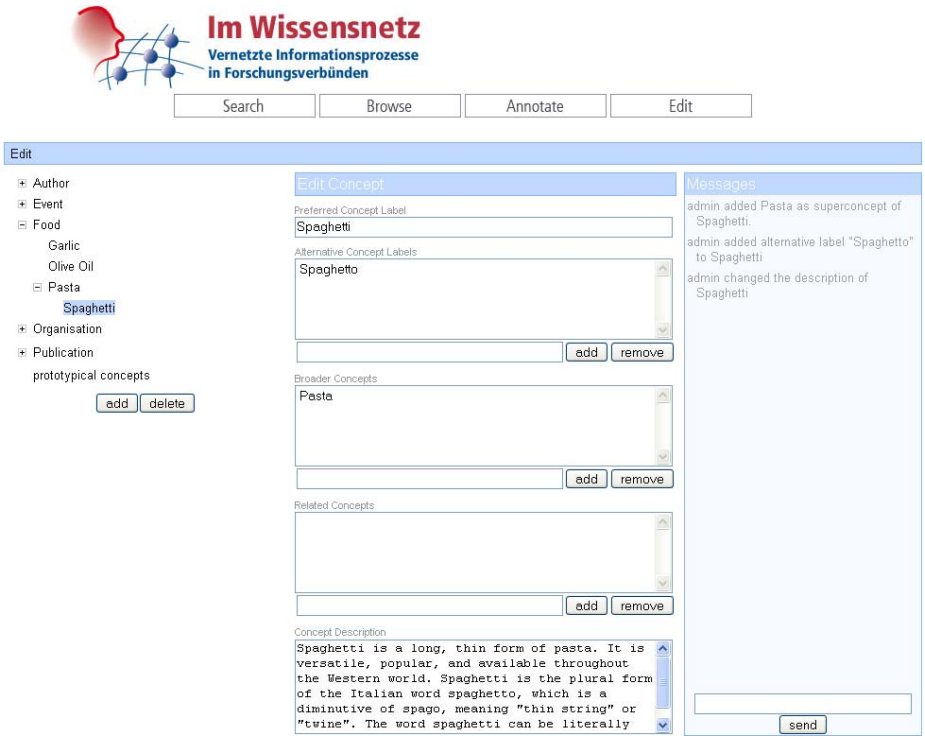

Fig. 2. Collaborative realtime onotology editor.

3.3 Browsing the Repository

Browsing the repository is the most common approach to retrieving information from a shared repository. With a browsing interface users can navigate to the concepts they are interested in and see the resources annotated with these. The browser interface also gives the chance to change any of the annotations. In SOBOLEO and social semantic bookmarking the user can also see the ontology and use the taxonomic structure for navigation. Fig. 3 shows the browsing interface for the new 'Spaghetti' concept. The interface shows the concept name, its labels and its description. Also shown are the most recently annotated documents (with links to change the annotation) and the relations to other concepts allowing for navigating there.

Fig. 3. Browsing interface for navigating to concepts and annotated resources.

3.4 Semantic Search

In addition to the browse interface the ontology is also used to enable semantic search. The semantic search in SOBOLEO combines semantic search utilizing the concept labels and their broader-narrower relations with a full text search over all annotated resources. The semantic search engine also offers query refinement and relaxation functionality.

In the example, a different user is interested in finding a recipe including noodles, garlic and basil and enters these words as search term. The semantic search recognizes that 'noodles' is a synonym for pasta and that spaghetti is a special kind of pasta. The search engine further finds that garlic refers to another concept and then that the annotation described earlier combines not only spaghetti and pasta as annotation but also includes basil in the sites content – hence this page is returned as a first result. The result is shown in Fig. 4. Please note that neither a full text engine (because 'noodles' is not written on the page), nor a social tagging system (because neither noodles nor basil is a tag), nor a pure semantic search engine (because basil is not annotated) could make a comparable ranking of the result.

Fig. 4. Result of the semantic search.

4 Related Work

There are a number of other approaches often presented as (social) semantic tagging. These include Bibsonomy, Int.ere.st, GroupMe!, Fuzzy, and Annotea, which we will describe and compare in the following:

- **BibSonomy:** Bibsonomy [10] is a system for the management of bookmarks of internet resources and publication entries. Bibsonomy offers functionality similar to that of well-known social bookmarking services but specifically tailored towards academics – e.g., it offers sophisticated support for uploading and exporting bibliographic information. At its core, Bibsonomy has a functionality very similar to social bookmarking services, but additionally offers users the possibility to create broader/narrower relations between tags. However, tag relationships are only local, i.e., each user can (and has to) maintain its own relationships and cannot profit from others' contributions in that respect.
- **Int.ere.st:** Int.ere.st [11] is a system concentrating on the transferability of tags and tagged resources between systems. Int.ere.st is created by the Digital Enterprise Research Institute, Galway and the Biomedical Knowledge Engineering of Seoul National University, Korea. Its functionality centers on making uploading and exporting tagging data simple and to allow for creating relations between tags (potentially coming from different systems).
- **GroupMe!:** GroupMe [1] attempts to bridge the gap between the Semantic Web and Web2.0 with an RDF based social bookmarking application. GroupMe! is developed by the Semantic Web Group at the University of Hannover in Germany. The main unique functionality of GroupMe! is the extension of the tagging idea with the concept of 'groups': all annotated resources can be organized into groups and these form another level of information that can be used for browsing and search.
- **Fuzzy:** Fuzzy [13] is a system for managing bookmarks of internet resources and ISBN numbers. Fuzzy is developed within the PhD project of Roy Lachica at the University of Oslo. It is based on Topic Maps technology and besides parent/child and horizontal tag relations the users can choose of 22 specific predefined

association types to link tags. Another main concept is voting for gardening and maintenance: the users can vote on bookmarks, tags a bookmark is annotated with, relations between tags, and users.

- **Annotea:** Annotea [12] is a metadata standard for semantic web annotations, it is implemented in a number of tagging tools and server applications. Annotea and its implementations have been developed by the W3C. Annotea differs from other approaches to social tagging in its emphasis on standards on decentrality, that it has sharing of bookmarks among services build in from ground up.

Table 1. Comparison of social semantic bookmarking tools.

	Public	**Full Text Indexing**	**Import/ Export Formats**	**Synonyms**	**Other Relations**	**Shared Relation Editing**	**Open Source**
Bibsonomy	Yes	No	XML, RSS, BURST, SWRC, Bibtex	No	Broader/ Narrower	No	No
Int.ere.st	No	No	SCOT, SIOC, FOAF	Yes	Identical	No	No
GroupMe!	Yes	No	RSS, DC, FOAF	No	Group	Yes	No
Fuzzzy	Yes	Yes	XTM, RSS	Yes	Broader/ Narrower, Specific association types	Yes	No
Annotea	No	No	Annotea	Yes	Broader/ Narrower	No	yes
SOBOLEO	No	yes	SKOS, RSS	Yes	Broader/ Narrower, Related	Yes	No

4.6 Comparison

To give a comprehensive overview of the respective strength and weaknesses of the approaches shortly introduced above, Tab. 1 details the main discriminating features among the applications, including SOBOLEO. The features used for the comparison are the following:

- **Public**: Whether the application has a public installation that can be used by any user.
- **Full Text Indexing**: Whether the application stores the text content of the annotated resources and uses it to facilitate search.
- **Import/Export Formats:** All tools discussed have some means to import or export the bookmarks, this row details which formats are used.
- **Synonyms:** Whether the application supports a notion of two natural language terms representing the same thing.
- **Other Relations**: The relations between tags/concepts that are supported by the applications, other than synonyms.
- **Shared Relation Editing**: Whether relations between tags exist only for one user or whether they are shared, i.e. in some systems the relation between tags is only visible to one user. Other users would need to create the same relation again.
- **Open Source**: Whether the source code of the applications is available as open source.

As a general conclusion, there is a big interest to extend social bookmarking in the direction of more semantics and in particular to tackle the problem how tagging data can be exchanged between systems, however, at the same time the table shows that there still is considerable disagreement about what are the most important features and – even more crucially – what are suitable formats to exchange the tagging data. Without an agreement in this domain, the promise of exchanging tagging data can obviously not be achieved. It is also interesting to see that the majority of the approaches still restricts the editing of relations between tags to only the private space and/or do not allow for a real community driven evolution of the semantic model.

5 Conclusion

Social Semantic Bookmarking allows a group of users to collaboratively create and evolve an index of resources together with the powerful semantic vocabulary used to organize it. Social Semantic Bookmarking promises better retrieval, better use of annotation, better integration of the repository with semantic web infrastructure etc. while avoiding the problems commonly associated with semantic annotation approaches – such as a high initial cost to build ontologies.

Parts of the vision of Social Semantic Bookmarking are already realized and used today, and evaluation studies like [3] confirm that users appreciate the new paradigm. In three user studies with 4, 24, and 33 participants we found that users liked the ease of use of the ontology editing (in comparison to other, more heavy-weight applications) and particular enjoyed the simple way of annotating resources with concepts or tags. Some users had initial problems, due to their very basic knowledge about ontologies, but all were able to obtain the necessary skills within a very short time.

Social Semantic Bookmarking applications promise a huge potential for future development as part of the developments towards a Web 3.0 as a user-centered semantic web. However, to realize this potential we also need a better understanding of the emergence and evolution of ontologies as part of everyday collaborative

activities and appropriate models and support mechanisms. Promising research approaches include the ontology maturing process [4], which is further explored as part of the Integrating Project MATURE[3].

References

1. Abel, F., Henze, F.M., Krause, D., Plappert & D. Siehndel, P.: Group Me! Where Semantic Web meets Web 2.0. In: Proc. of the 6th Int. Semantic Web Conf. (2007)
2. Barker, K., Chaudhri, V.K., Char, S.Y., Clark, P., Fan J., Israel, D., Mishra, S., Porter, B.W., Romero, P., Tecuci, D. & Yeh P.: A question-answering system for AP Chemistry: Assessing KR&R technologies. In Proc. of the 9th Int. Conf. on Principles of Knowledge Representation and Reasoning (2004) 488-497
3. Braun, S., Schmidt, A., Walter, A., Nagypal, G., Zacharias, V.: The Ontology Maturing Approach to Collaborative and Work-Integrated Ontology Development: Evaluation Results and Future Directions. In: Proc. of the ESOE-Workshop at ISWC'07, CEUR Workshop Proc. Vol. 292 (2007) 5-18
4. Braun, S., Schmidt, A., Walter, A., Nagypal, G., Zacharias, V.: Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. In: Proc.of the Workshop on Social and Collaborative Construction of Structured Knowledge at WWW'07, CEUR Workshop Pro. Vol. 273 (2007)
5. Dean, M., Schreiber, G.: OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004 (2004)
6. Golder, S., Huberman, B.A.: The Structure of Collaborative Tagging Systems. J. of Inf. Sc. Vol. 32(2) (2006) 198-208
7. Guy, M., Tonkin, E.: Folksonomies: Tidying Up Tags? D-Lib Magazine Vol. 12 Nr. 1 (2006)
8. Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. Int. J. of Human-Computer Studies Vol. 43 (1995) 907–928
9. Hepp, M.: Possible Ontologies: How reality constraints building relevant ontologies. IEEE Internet Computing Vol. 11(1) (2007) 90-96
10. Hotho, A., Jäschke, R., Schmitz, C.; Stumme, G.: BibSonomy: A Social Bookmark and Publication Sharing System. In: CS-TIW'06. Aalborg: Aalborg University Press (2006)
11. Kim, H.L., Yang, S.-K., Song, S.-J., Breslin, J.G. & Kim, H.-G.: Tag Mediated Society with SCOT Ontology. In: Proc. of the 5th Semantic Web Challenge at ISWC'07 (2007)
12. Koivunen, M.-R.: Semantic Authoring By Tagging with Annotea Social Bookmarks and Topics. In: Proc. of the Semantic Authoring and Annotation Workshop at ISWC'06 (2006)
13. Lachica, R., Karabeg, D.: Metadata creation in socio-semantic tagging systems: Towards holistic knowledge creation and interchange. In: Scaling Topic Maps. Topic Maps Research and Applications 2007, Springer, (2007)
14. Manola, F., Miller, E.: RDF Primer. W3C Recommendation 10 February 2004 (2004)
15. Miles, A., Bechhofer, S.: SKOS Simple Knowledge Organization System Reference. W3C Working Draft 25 January 2008 (2008)
16. Tanaka, J.W.; Taylor, M.: Object categories and expertise: Is the basic level in the eye of the beholder? Cognitve Psychology Vol. 23(3) (1991) 457-482.
17. Zacharias, V., Braun, S. (2007). SOBOLEO – Social Bookmarking and Lightweight Ontology Engineering. In: Proc. of the Workshop on Social and Collaborative Construction of Structured Knowledge at WWW'07, CEUR Workshop Pro. Vol. 273 (2007)

[3] http://mature-ip.eu