Semantify del.icio.us: automatically turn your tags into senses

Maurizio Tesconi, Francesco Ronzano,
Andrea Marchetti, and Salvatore Minutoli

Institute for Informatics and Telematics
National Research Council - C.N.R.
Via G. Moruzzi, 1 - 56100 Pisa - Italy
{maurizio.tesconi, francesco.ronzano, andrea.marchetti,
salvatore.minutoli}@iit.cnr.it

Abstract. At present tagging is experimenting a great diffusion as the most adopted way to collaboratively classify resources over the Web. In this paper, after a detailed analysis of the attempts made to improve the organization and structure of tagging systems as well as the usefulness of this kind of social data, we propose and evaluate the Tag Disambiguation Algorithm, mining del.icio.us data. It allows to easily semantify the tags of the users of a tagging service: it automatically finds out for each tag the related concept of Wikipedia in order to describe Web resources through senses. On the basis of a set of evaluation tests, we analyze all the advantages of our sense-based way of tagging, proposing new methods to keep the set of users tags more consistent or to classify the tagged resources on the basis of Wikipedia categories, YAGO classes or Wordnet synsets. We discuss also how our semanitified social tagging data are strongly linked to DBPedia and the datasets of the Linked Data community.

1 Introduction

Tagging is currently one of the most widespread patterns to create, collect and share huge amounts of social data over the Web, represented by the set of tags adopted by the community of users of a tagging service to describe resources of interest. The number of Web tagging services and, in particular, the amount of social bookmarking sites, that are tagging services devoted to tag URLs, is rapidly growing [4]: among them Del.icio.us[1], mainly, but not only, used by Web experts, Bibsonomy[2], exploited by researchers to share links to papers and other relevant works and Technorati[3], widely adopted by communities of bloggers, represent relevant examples.

The possibility of freely choosing tags is probably one of the main reasons for the popularity of social tagging but it also makes difficult to produce a clean

[1] http://del.icio.us/
[2] http://www.bibsonomy.org/
[3] http://www.technorati.com/

and consistent organization and classification of the tagged resources. During the last few years, starting from a lot of different statistical analyses of tagging data collections, many distinct approaches to better create, structure and search for information querying tagging services have been proposed; the most relevant are briefly reported in Section 2. These studies mainly point out problems like synonymy, polysemy and in general all the different lexical forms or tags that can be used by each user to describe a concept; these issues represent the main causes of loss of consistency in tagging data collections as well as of decrease of precision and recall of tag based searches. The general and globally agreed way to face these issues consists of using some sort of semantic classification process to give an explicit meaning to each freely chosen tag; some of these techniques rely only upon the data retrievable from a tagging service, others also exploit external semantic resources.

In this context, we propose the *sense-based tagging*, a new approach to automatically structure the set of tags collected by each user of a tagging system. We automatically disambiguate the meaning of tags mining the information contained in Wikipedia: for this purpose we present and evaluate the Tag Disambiguation Algorithm, described in details in Section 3. In Section 4, we expose and evaluate all the advantages of the *sense-based tagging* in terms of classification of resources and creation of semantic metadata.

In conclusion, in Section 5, we expose our future works as well as the ongoing effort to implement all the approaches described creating our Web sense-based tagging service.

2 Related work: how to improve tags organization

Many studies have been carried out to describe and make new proposals to better organize and expose the information collected by tagging services. We can find a lot of statistical investigations about the structure of tags collections and the dynamics of their usage together with all the subsequent considerations about the most frequent tagging patterns: among them two interesting works based on del.icio.us bookmarks are [14] and [11]. There are also relevant analysis concerning other quantitative studies of social bookmarking systems aiming at finding how tagging data can support Web search and improve its results, providing additional information like in [6] and [16].

Summarizing the results of many relevant works regarding the possibility to improve the effectiveness of tagging, we have pointed out that all of them have identified two main causes for the poor structure and organization of tag-based classifications: the complete freedom users have when they choose tags and the lack of any semantic information to support tagging activity. In order to face these issues, different strategies have been set up. Tag recommendations systems have been proposed to keep users tagging data more consistent [19]; some of them exploit also external semantic resources like Wordnet[4] to perform

[4] http://wordnet.princeton.edu/

this task [15]. Moreover we can also find services that analyze the tags of a specific user to detect tag usage inconsistencies like slightly different keywords: relevant examples are Bookmark Cleaner[5] and Del.icio.us tag cleaner[6], all related to del.icio.us bookmarks.

The systems and the procedures that use some sort of semantic information to better organize tags and understand their meaning can be divided into *two groups*. The first one comprises *all the methods that introduce some sort of structure to the sets of tags taking into account only the information retrievable from tagging services*, that is the collections of users, tags and tagged resources. They mainly try to group together similar tags on the basis of their relations with users and resources. In this way they identify sets of strictly related tags or understand the sense of ambiguous ones: some of these procedures are described in [5] and [10]. The second group of semantic based approaches *exploits external semantic resources to structure sets of tags*. Some of them try to define the right meaning of each tag retrieving the semantic relations that occur between related tags so as to visualize tags on the basis of their sense and relevance. In order to achieve that, data extracted from different ontologies available over the Web are collected and merged; examples of this kind of methods are [13], [17] and [3]. They often suffer the poor terms coverage of Web ontologies, providing encouraging results only in particular domains. To point out the right meaning of tags other techniques exploit Wikipedia as well as its "semantic version", the DBPedia ontology, connected with a growing number of external datasets. Among them we can point out our last project, SemKey [1] that help users to disambiguate tags relying upon Wikipedia and Wordnet. Also the MOAT Project [2] provides an infrastructure to collect the concepts associated to tags, identified by means of DBPedia URIs. Similarly, the Faviki [7] Web system allows users to describe Web resources through Wikipedia meanings. All these projects don't support users enough in the management of the semantic data needed to describe their tags; often the choice of a particular tag or better a specific meaning, is not easy and requires a lot of additional user interactions. Moreover it is limited by the set of available word-meaning associations. In this context, also different ontologies have been defined to provide a reference model to describe the tagging activity in terms of the relations between tags, resources, users and tagging actions: a relevant example is the SCOT Ontology [7], that reuses and extends the previous Tag Ontology [8].

We propose and evaluate a methodology to overcome these limitations using an algorithm that automatically points out the right meaning of each tag, considering those available in Wikipedia, thus semantifying the tag collection of each user of a tagging service. We allow users to provide new tags to refer to a specific meaning thanks to the exploitation of Tagpedia, a semantic resource built ad hoc from Wikipedia to disambiguate tags. We rearrange user resources

[5] http://www.bookmarkcleaner.com/
[6] http://delicious.isnotworking.com/
[7] http://www.faviki.com/

on the basis of external classifications like Wikipedia categories or Yago and Wordnet classes.

3 From tags to Wikipedia senses: Tagpedia and the Tag Disambiguation Algorithm

Our idea of *sense-based tagging* is grounded on the possibility to define the meaning of each tag chosen by a user to describe one or more resources over the Web. In order to find out the right meaning of a tag we need a sort of **global semantic reference containing a rich collection of senses to search into**. Currently Wikipedia[8] represents the richest and constantly updated encyclopedic reference over the Web with a huge set of semantic contents, even if not explicitly exposed and easily accessible. Because of its features, Wikipedia is an ideal starting point to retrieve the information needed to define the meaning of a tag. Based on this assumption, by mining the contents of Wikipedia, we have built *Tagpedia*: this is a semantic reference for organization and classification of tags, intended as words or more in general as short textual expressions used to refer to a specific topic. Tagpedia is based on the model of term-concept networks [12]; *for each meaning of Wikipedia, Tagpedia groups together all the different words used to refer to it.* Currently Tapgedia includes more than *1,92* millions of distinct concepts and more than *4,23* millions of words used to refer to these meanings: these data have been extracted from Wikipedia pages. Tagpedia, built ad hoc to support the semantic characterization of Web contents through sense-based tagging and thus to easily disambiguate the meaning of a tag, is accessible over the web at the URL *http://www.tagpedia.org/*. It can be also queried by means of a dedicated Web API and it can be collaboratively edited. To get more information about the structure and the usage of Tagpedia see [9].

In this section we propose and evaluate a *Tag Disambiguation Algorithm* (TDA), implemented relying upon Tagpedia. Our implementation collects the tags of a user from a tagging service (del.icio.us, in our case) and for each of them finds out the relative sense by linking it to the corresponding page of Wikipedia.

In particular the TDA identifies for each tag t a list of candidate senses, referred to also as concepts or meanings in the rest of this paper and assigns them a number, called **sense-rank** SR: **the higher the rank of a meaning, the better that meaning defines the sense intended by the user for that tag**. In the remaining part of this Section we explain in more details the TDA and provide some evaluation of its disambiguation effectiveness.

In order to calculate the sense-rank of a meaning of the tag t of the user U we consider the text of Wikipedia describing that particular meaning and we base our algorithm on the following assumptions:

- the more the meaning described by that text is similar to the one intended by the user U, the higher is, in that text, the number of occurrences of tags that are in some way connected to the tag t.

[8] http://en.wikipedia.org/wiki/

– the meaning intended by the user U for the specific tag t doesn't change while tagging [5]. This is particularly true if we think that normally the interests of a user are generally focused on defined domains, so the meaning of a tag is usually unique.

We want to apply the TDA in order to disambiguate the tag t of the user U, ranking all its meanings to choose the one with the highest sense-rank. In particular, **starting from the texts of Wikipedia, we base our analysis on the calculation of the number of occurrences of the following two groups of relevant tags**: those used by the user U along with the tag t to describe Web resources of interest and the del.icio.us popular tags used to characterize the same set of Web resources, along with the popularity value of each of the selected popular tags.

Generalizing we can state that the user U has tagged m Web resources, R_1, R_2,..., R_m with the tag t . There are n meanings M_1, M_2,..., M_n of the tag t in Wikipedia; we want to calculate the sense-rank $SR_t(M_x)$ of each of them.

In order to do that we exploit **two different groups of parameters**: some of them are retrieved from del.icio.us and other ones are retrieved from Tagpedia and Wikipedia. In particular we consider:

- from **del.icio.us**:

• all the y tags CT_1, CT_2, ..., CT_y of the user U that co-occur with the tag t, considering the Web resources R_1, R_2,..., R_m, along with their respective frequency of co-occurrence $CF(CT_1)$, $CF(CT_2)$,..., $CF(CT_y)$;

• all the z popular tags PT_1, PT_2, ..., PT_z used in del.icio.us to describe the Web resources R_1, R_2,..., R_m along with their total popularity frequency $PF(PT_1)$, $PF(PT_2)$, ..., $PF(PT_z)$; the popularity frequency of a tag for a particular resource is the number of times that word has been used to tag the same resource.

- from **Tagpedia**:

• the n meanings M_1, M_2, ..., M_n of the tag t in Wikipedia; for instance the tag *owl* can be used to refer to the Web Ontology Language but also to a nocturnal bird or to an Australian rugby union club and so on;

• the n texts $T(M_1), T(M_2), ..., T(M_n)$ of the articles of Wikipedia describing each of the meanings of the tag t and for each of those texts, the number of occurrences of a particular tag W, referred to as $OCC(W, T(M_n))$.

All these values are used to calculate $SR_t(M_x)$, that is the sense-rank of the meaning M_x of the tag t for the user U. In particular, the value of the sense-rank is the weighted sum of two contributions: the first one, $SRU_t(M_x)$, is related to the tags of the user U that co-occur with the tag t; the second one, $SRP_t(M_x)$, deals with the popular tags used in del.icio.us to refer to the resources tagged by the user U with t. The sense-rank of the meaning M_x of the tag t for the user U is equal to:

$$SR_t(M_x) =$$
$$WU\times [SRU_t(M_x) \;/\; \max(SRU_t(M_v))] + WP\times[SRP_t(M_x)/ \max(SRP_t(M_u))]$$

where v and u range from 1 to n and $WU + WP = 1$.

We divide the values of $SRU_t(M_x)$ and $SRP_t(M_x)$ respectively by the max value of SRU_t and SRP_t for all the n meanings of the tag t: in this way we normalize all the sense-ranks of the tag t to the interval $[0, 1]$, making their values comparable in order to choose the highest one. We have that:

$$SRU_t(M_x) = [CF(CT_1) \times OCC(CT_1, T(M_x)) + CF(CT_2)\times$$
$$\times OCC(CT_2, T(M_x)) + .. + CF(CT_y) \times OCC(CT_y, T(M_x))] * UNN/UTOT$$

$$SRP_t(M_x) = [PF(PT_1) \times OCC(PT_1, T(M_x)) + PF(PT_2)\times$$
$$\times OCC(PT_2, T(M_x)) + .. + PF(PT_y) \times OCC(PT_y, T(M_x))] * PNN/PTOT$$

where:

- UNN is the total number of values of $OCC(CT_o, T(M_x))$ that are not equal to 0
- $UTOT$ is the total number of tags of the user U co-occurring with the tag to disambiguate t in at least a description of a Web resource.
- PNN is the total number of values of $OCC(PT_o, T(M_x))$ that are not equal to 0
- $PTOT$ is the total number of del.icio.us popular tags related to the Web resources tagged by the user U with the tag t.

The values of the weights WU and WP in our TDA evaluation phase are both set to 0.5, but we are experimenting the possibility to adapt them to the quality and the origin of the set of disambiguation data, that is the set of tags that are related in some way to the one to disambiguate. For instance, if a user has chosen a tag to characterize a small number of resources or if the tag to disambiguate has a small number of co-occurring tags chosen by the user, when we calculate the sense-rank of the meanings of that tag, we can decrease the importance of the user tags contribution (decreasing the value of WU). When the sense-rank of each meaning of a polysemous tag has been calculated, the TDA chooses the sense with the highest rank as the correct one.

3.1 TDA evaluation

Now we provide some evaluation of our Tag Disambiguation Algorithm. As the starting point, querying del.icio.us, we have chosen to consider the tagging profile of 9 del.icio.us users (U_1, U_2,..., U_9) as it was on July 2008, ranging from very active taggers to people that only occasionally save some bookmark: globally we have collected 3520 tags used to bookmark 3926 URLs with a total average number of tags per bookmark equal to $3, 38$. In the first columns of Table 1 are synthesized the distinctive features of the tagging profile of each user.

We have queried Tagpedia and in case of polysemous tags we have applied the TDA to choose the best meaning to be associated to each of them. We aim at evaluating respectively the coverage of the tags collection of Tagpedia in terms of the number of tags that can be associated to at least one sense and the effectiveness of the Tag Disambiguation Algorithm in terms of the correctness of the sense associated to each of the tags.

From the the last two columns of Table 1 we can notice that thanks to the support of Tagpedia and the execution of the TDA we can point out the

User	Bookmarked URL	Tags	Average number of tags per bookmark	Disambiguated tags	Distinct concepts
U_1	136	166	2,72	148 (89,16%)	141
U_2	275	355	3,48	321 (90,42%)	302
U_3	279	396	3,89	345 (87,12%)	323
U_4	541	511	4,02	449 (87,87%)	403
U_5	754	1149	6,54	1026 (89,30%)	885
U_6	428	175	1,38	165 (94,29%)	160
U_7	69	142	3,62	133 (94,66%)	129
U_8	453	76	2,08	71 (93,42%)	71
U_9	991	550	2,67	501 (91,09%)	470
Average	—	—	**3,38**	**3159 (89,74%)**	**2884**

Table 1. Tagging profile of del.icio.us test users and tags disambiguation results

meaning of 89, 74% of the tags of the 9 del.icio.us users. Considering the 3159 disambiguated tags, 2884 different senses have been identified.

The 11,71% of the considered tags have not been associated to a concept, because in Tagpedia there are no words for them.

In conclusion, we have evaluated the TDA effectiveness through a process of human review of results: among 2589 polysemous tags the TDA has chosen the right meaning of the 89, 15% of them. Globally, the 2891 (91, 52%) of the 3159 disambiguated tags has been correctly associated to the right meaning.

4 Advantages of sense-based tagging

In this section we explore the advantages of the sense-based tagging considering new ways to group similar tags, to classify resources and to produce semantic tagging metadata.

4.1 Sense-based grouping of tags

A first advantage of the adoption of sense-based tagging is represented by the possibility to group together user tags that refer to the same concept. In fact, considering the set of 9 del.icio.us users analyzed in the previous tables, we can notice that only 2884 distinct concepts, out of 3159 disambiguated tags, have been pointed out. This means than *on average the 9% of the tags chosen by a user to describe a resource has a sense referred to by other tags already present in the tagging profile of the user*. In the following table is shown some example of groups of tags having the same meaning, defined as the outcome of our disambiguation procedure based on the TDA; all the tags grouped together are words that refer to the same sense in Tagpedia and thus the same associated page of Wikipedia.

Wikipedia concept name (page title)	**Brief Wikipedia concept description**	**Group of tags referring that concept**
Carpool	Carpooling is the shared use of a car by...	rideshare, ridesharing, carpooling, carpool
Humour	Humour or humor is the tendency to provoke laughter and provide amusement..	funny, humor, humour
Film	Film is a term that encompasses individual motion pictures, the field of film as an art form...	film, films, movies, movie

Table 2. Grouping of tags having the same meaning

4.2 Exploiting Wikipedia categories, YAGO classes and Wordnet Synsets to classify resources

Once characterized all the bookmarked URLs of a user through the concepts of Wikipedia associated exploiting our Tag Disambiguation Algorithm, we have tried to create different alternative views of user resources exploiting three classification systems: the Wikipedia Categories, the YAGO classes and the Wordnet synsets. The mapping of Wikipedia concepts and thus of the corresponding Wikipedia pages to Wordnet synsets and YAGO classes has been derived from those available in the DBpedia datasets[9].

The Wikipedia categories is a collaboratively built categorization of Wikipedia articles: Wikipedia users can place one article in one or more categories or also create new categories and connect them to the other categories through subsumption relations. Almost all the articles of Wikipedia have been placed in at least one of the more than 312 thousands categories of Wikipedia. Because of its collaborative definition, the Wikipedia categorization system is untidy and includes many subsumption cycles or other kinds of inconsistencies.

YAGO (Yet Another General Ontology)[10] is a large semantic knowledge base, that is automatically extracted from Wikipedia and uses Wordnet to organize information. It has been developed by the Max Plank Institute for Computer science. We have considered the mapping of Wikipedia pages to the corresponding YAGO classes: 1,412 millions of pages of Wikipedia has been mapped to at least one class in YAGO. Thus YAGO covers about 74% of the pages of Wikipedia. YAGO classes are arranged through the YAGO Class Hierarchy. The number of Wikipedia pages that have been mapped to one or more synsets of Wordnet is considerably lower.

Our first analysis aims at defining, considering the previously defined group of 9 del.icio.us users, how many senses of their profiles are covered by each classification system and, as a consequence, we can have an idea of the adequacy of the particular classification system to provide a new structure to organize and present the tagged resources to the users. The results are presented in Table 3.

[9] http://wiki.dbpedia.org/Downloads31/
[10] http://www.mpi-inf.mpg.de/∼suchanek/downloads/yago/

Distinct concepts	Concepts coverage of:		
	Wikipedia categories	YAGO classes	Wordnet synsets
2884	2749 (95%)	1667 (58%)	507 (18%)

Table 3. Total concepts coverage of Wikipedia categories, YAGO classes and Wordnet synsets

We can notice that the synsets of Wordnet cover an irrelevant portion of the senses of the 9 test users, while the YAGO classes includes the 58% of the concepts. The Wikipedia category system manages to classify almost all the senses of each user.

Each sense exploited by a single user describes more than one Web resource. Table 4 is similar to Table 3; it is intended to test the coverage of the three considered classification systems in terms no more of senses, but of bookmarked Web resources. We define how many tagged Web resources can be associated to at least a Wikipedia Category, a YAGO class or a Wordnet synset.

URLs	URLs described by disambiguated tags	URLs coverage of:		
		Wikipedia categories	YAGO classes	Wordnet synsets
3926	3864	3852 (100%)	3284 (85%)	1749 (45%)

Table 4. URL coverage of Wikipedia categories, YAGO classes and Wordnet synsets

Our results in this case are encouraging. Because each Web resource is on average tagged through two or more senses, we can see that, considering Wikipedia Categories, practically all of them are placed at least in a category; thus, exploiting Wikipedia Categories and the sense-based tagging, we can provide a new classification of all the bookmarks of our users. Also considering YAGO classes, we can map on average 85% of users resources on one or more of them. The coverage of the resources considering Wordnet classes is still too low to provide a valuable new classification of user resources through Wordnet hierarchy of classes.

In order to evaluate category-based classifications of user resources we have made some initial test related to YAGO classes. Considering the YAGO classes hierarchy, we have defined how many classes are necessary to classify the resources of each user, considering that 85% of these resources are present at least in a class. Moreover, we have calculated how the number of classes decreases and thus how the class-based classification of user resources gets more coarse-grained when we move to higher levels of the hierarchy. We have considered seven different levels of ancestors of the direct YAGO classes that are those containing at least one tagged resource and we have defined, considering the classes of each level, how many classes are involved in the classification of user resources. The result of our analysis, considering the average number of classes needed to classify user resources for each level of ancestors, are shown in Table 5.

User concepts belonging to YAGO classes	**Direct YAGO classes**	**Levels of ancestors**						
		I	II	III	IV	V	VI	VII
185	**162**	**132**	**102**	**69**	**42**	**29**	**19**	**12**

Table 5. Average number of classes needed to represent user resources considering seven levels of ancestors

We can notice that, going back up to the seventh level of ancestors of the YAGO classes containing at least one tagged resource, the average number of different classes needed to classify the resources in at least one of them decreases from 162 to 12 different ones, thus allowing us to organize the users resources in a number of sets that can vary from 162 to 12.

In conclusion, in this Section we have shown, thanks to some initial analysis, that once classified users resources through senses and thus through the concepts of Wikipedia, the Wikipedia Categories and the YAGO classes can provide new classifications of these resources based on global shared classificatory schemas.

4.3 Linking del.icio.us to Linked Data datasets

The *sense-based tagging* is also a way to connect the social data collaboratively created through tagging and the Semantic Web. The Linked Data community actually represents one of the most relevant attempts to collect, interlink and semantically expose over the Web the information contained in many different datasets, through the adoption of the RDF and RDF triples, putting in practice the vision of the Semantic Web. It tries to define a common set of rules and best practices to publish and browse semantic-aware information. DBPedia[11][18] is a sort of alignment ontology created by mining Wikipedia, representing in some way the glue of the Linked Data community: by exploiting DBPedia each concept of Wikipedia can be univocally referenced through a specific URI that represents also the way to retrieve over the Web the RDF triples describing that concept. The Tag Disambiguation Algorithm manages to automatically convert each tag into the intended concept of Wikipedia and thus into the related URI of DBPedia. In this way we are able to generate for each user of a *sense-based tagging service*, a set of RDF triples describing his tagging profile; they include, for instance, a triple for each sense associated to a tagged resource. Identifying each user through the URL of his FOAF profile[12] and pointing out each referred sense

[11] http://dbpedia.org/About
[12] http://www.foaf-project.org/

through a DBPedia URI, each sense-based tagging service, can make accessible over the Web throug user-dedicated URLs the RDF triple-based descriptions of its users, following the publishing rules of the Linked Data Community. In this way we can automatically link the social data of a tagging service to the datasets of the Linked Data Community, thus providing a huge amount of continuously growing collaboratively created descriptions of Web resources.

5 Conclusions and future works

In this paper, we have described and evaluated a new way to automatically semantify the tags of the users of a Web tagging service, thanks to the Tag Disambiguation Algorithm. Considering del.icio.us and the tags of 9 users, we have managed to correctly find out the meanings of the $91,52\%$ of these tags, linking them to the right concepts of Wikipedia. Consequently we have described an evaluated the possibility to clean users tags grouping them by sense, but also to classify the tagged resources on the basis of Wikipedia categories, YAGO classes and Wordnet synsets. Wikipedia categories and YAGO classes, because of their wide coverage of the concepts of Wikipedia, can support the definition of new way to classify users resources. Characterizing Web resources through Wikipedia concepts we can also connect the social data produced by tagging systems to the datasets of the Linked Data community.

In our future works we would like to improve and better tune our Tag Disambiguation Algorithm. We want also better investigate and test the adequacy of Wikipedia categories and Yago class hierarchy to provide new views of the user resources of a tagging system. Moreover, we are developing a Web system that enable users to automatically semantify their tagging profile, retrieving it from other tagging services, thus making them easily shift to the adoption of the sense-based tagging.

In conclusion, we believe that the sense-base tagging thanks to the automated semantification of tags supported by the Tag Disambiguation Algorithm and because of its strict connection with the Linked Data community, can represent a valuable way to improve the quality and the organization of social tagging allowing to produce a considerable set of interlinked semantic data over the Web.

This work is funded by the European Community, in the 7th Framework Project KYOTO[13].

References

1. F. Ronzano, M. Rosella, S. Minutoli, A. Marchetti, M. Testoni. Semkey: A semantic collaborative tagging system. In Proc. of The workshop Tagging and Metadata for Social Information Organization of the World Wide Web Conference 07, May 8-12, 2007, Banff, Canada.

[13] http://www.kyoto-project.org/

2. A. Passant, P. Laublet. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In Proc. of The Linked Data on the Web Workshop of the World Wide Web Conference 08, April 19-25, 2008, Beijing, Cina.
3. T. Roth-Berghofer, B. Adrian, L. Sauermann. Contag: A semantic tag recommendation system, 2007.
4. L. Baker. 125 social bookmarking sites : Importance of user generated tags, votes and links. Blog article, December 2007.
5. N. Gibbins, C. M. Au Yeung, N. Shadbolt. Understanding the semantics of ambiguous tags in folksonomies. At The International Workshop on Emergent Semantics and Ontology Evolution at ISWC/ASWC 2007, 12 November 2007, Busan, South Korea.
6. Shuyi Zheng-Hongyuan Zha, C. Lee Giles Ding Zhou, Jiang Bian. Exploring social annotations for information retrieval. In Proc. of The World Wide Web Conference 08, April 19-25, 2008, Beijing, Cina.
7. Hak Lae Kim, J. G. Breslin, S. Scerri, S. Decker, Hong Gee Kim, Sung Kwon Yang SCOT Ontology Specification. DERI Galway at the National University of Ireland, Galway, Ireland, August 2008.
8. R. Newman Tag ontology design. Blog Article, 2005.
9. M. Tesconi F. Ronzano, A. Marchetti, S. Minutoli. Tagpedia: a semantic reference to describe and search for web resources. In Proc. of The workshop Social Web and Knowledge Management of the World Wide Web Conference 08, April 19-25, Beijing, Cina.
10. P. Keller G. Begelman, F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. In Proc. of The World Wide Web conference 06, April 23-26, 2006, Edinburgh, Scotland.
11. Scott A. Golder, Bernardo A. Huberman. The structure of collaborative tagging systems. Technical report, Information Dynamics Lab, HP Labs, 2005.
12. A. Gregorowicz, M. A. Kramer. Mining a large-scale term-concept network from wikipedia. Mitre Technical Report, October 2006.
13. M. Espinoza J. Gracia, R. Trillo, E. Mena. Querying the web: a multiontology disambiguation method. In Proc. of The 6th international conference on Web engineering, July 10-14, 2006, Menlo Park, California.
14. M. E. I. Kipp, D. Grant Campbell. Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In Proc. of The Annual General Meeting of the American Society for Information Science and Technology, 2006, Austin, Texas (USA).
15. F. Lanubile, G. Semeraro, P. Basile, D. Gendarmi. Recommending smart tags in a social bookmarking system. In Proc. of The European Semantic Web Conference 07, June 3-7, 2007, Innsbruck, Austria.
16. G. Koutrika, P. Heymann, H. Garcia-Molina. Can social bookmarking improve web search? In Proc. of The WSDM08, February 11-12, 2008, Palo Alto, California.
17. L. Specia, S. Angeletou, M. Sabou, E. Motta. Bridging the gap between folksonomies and the semantic web: An experience report. In Proc. of The European Semantic Web Conference ESWC 2007, June 7, 2007, Innsbruck, Austria.
18. G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, S. Auer, C.Bizer. Dbpedia: a nucleus for a web of open data. In Proc. of The International Semantic Web Conference 07, November 11-15, 2007, Busan, Korea.
19. Jianchang Mao Zhichen Xu, Yun Fu, Difu Su. Towards the semantic web: Collaborative tag suggestions. In Proc. of The World Wide Web conference 06, April 23-26, 2006, Edinburgh, Scotland.