Organizing Publications and Bookmarks in BibSonomy

Robert Jäschke[1,2]

Miranda Grahl[1]

Andreas Hotho[1]

Beate Krause[1,2]

Christoph Schmitz[1]

Gerd Stumme[1,2]

[1]Knowledge & Data Engineering Group, Dept. of Mathematics and Computer Science
Univ. of Kassel, Wilhelmshöher Allee 73, D-34121, Kassel, Germany
{lastname}@cs.uni-kassel.de

[2]Research Center L3S, Univ. of Hannover,
Appelstr. 9a, D-30167 Hannover, Germany

ABSTRACT

BibSonomy is a web-based social resource sharing system which allows users to organise and share bookmarks and publications in a collaborative manner.

Apart from standard folksonomy features such as an intuitive user interface, navigation along all dimensions, or browser integration via RSS feeds, BibSonomy provides tag hierarchies, group management and privacy features, and numerous import and export functions.

Categories and Subject Descriptors

H.3.4 [**Information Systems**]: Systems and Software

Keywords

folksonomies

1. SHARING BOOKMARKS AND PUBLICATIONS WITH BibSonomy

Social resource sharing systems are web-based systems used to manage resources on the web in a collaborative way. Users can describe the resources with arbitrary words, so-called *tags*. The systems can be distinguished according to what kind of resources are supported. Flickr, for instance, allows the sharing of photos, del.icio.us the sharing of bookmarks, CiteULike[1] and Connotea[2] the sharing of bibliographic references, and 43Things[3] even the sharing of personal goals and resolutions. Our own system, *BibSonomy*,[4] can be used for sharing bookmarks and BibTEX entries simultaneously. In their core, these systems are all very similar. Once a user is logged in, he can add a resource to the system and assign arbitrary tags to it.

The collection of all users' tag assignments is called a *folksonomy*. The word 'folksonomy' (coined by Vander Wal in [10]) is a blend of the words 'taxonomy' and 'folk', and stands for conceptual structures created by the people. Folksonomies are thus a bottom-up complement to more formalized Semantic Web technologies, as they rely on *emergent semantics* [8, 9] which result from the converging use of the same vocabulary.

A typical user interface allows for exploration of the folksonomy in all dimensions: for a given user one can see all resources he has uploaded, together with the tags he has assigned to them; when clicking on a resource one sees which other users have uploaded this resource and how they tagged it; and when clicking on a tag one sees who assigned it to which resources (see Figure 1).

After an introduction to the user interface and architecture of BibSonomy, we give an overview about some of its advanced features. For the underlying data structure of the system and more details, we refer to [2].

1.1 User Interface

BibSonomy with its more than 5,000 registered users allows to share bookmarks (i.e., URLs) as well as publication references. The data model of the publication part is based on BibTEX [6], a popular literature management system for LATEX [5]. A typical list of posts is depicted in Figure 1 which shows bookmark and publication posts in a column layout containing the tag *web*. The page is divided into four parts: the header (showing information such as the current page and path, navigation links and search boxes), two lists of posts – one for bookmarks and one for publications – each sorted by date in descending order, and a list of tags related to the posts. This scheme holds for all pages showing posts and allows for navigation in all dimensions of the folksonomy.

A detailed view of one bookmark post from the list in Figure 1 can be seen in Figure 2. The first line shows in bold the title of the bookmark which has the URL of the bookmark as underlying hyperlink. The second line shows an optional description the user can assign to every post. The last two lines belong together and show detailed information: first, all the tags the user has assigned to this post (*web, service, tutorial, guidelines* and *api*), second, the user name of that user (*hotho*) followed by a note, how many users tagged that specific resource. These parts have underlying hyperlinks, leading to the corresponding tag pages of the user (/user/hotho/web [5], /user/hotho/service, ...),

---

[1]http://www.citeulike.org

[2]http://www.connotea.org

[3]http://www.43things.com

[4]http://www.bibsonomy.org

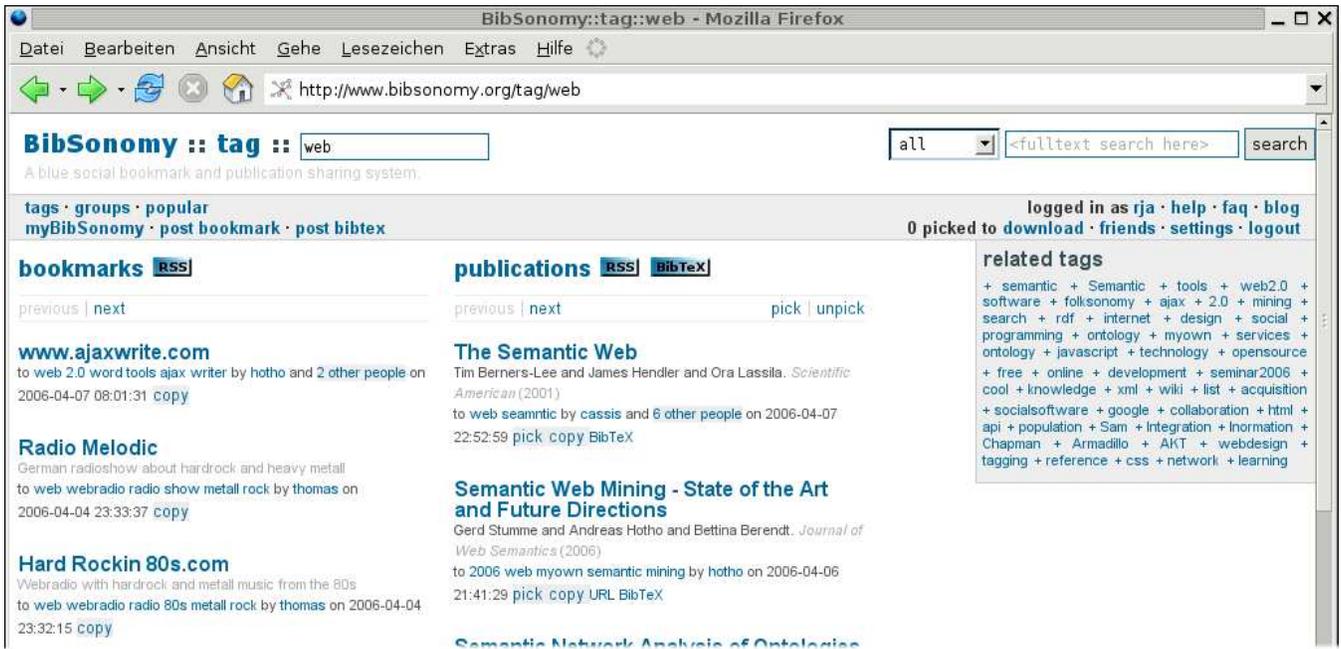[5]All paths given in parentheses are relative to http://www.

Figure 1: BibSonomy displays bookmarks and BIBTEX based bibliographic references simultaneously.

the users page (`/user/hotho`) and a page showing all four posts (i. e., the one of user *hotho* and those of the 3 other people) of this resource (`/url/r`, where r is a hashed representation of the resource). The last part shows the posting date and time followed by links for actions the user can do with this post – depending on whether this is his own post (*edit*, *delete*) or another user's post (*copy*).

REST web services

Good intro to the REST "architecture"

to web service tutorial guidelines api rest by hotho and 3 other people on 2006-04-04 16:11:47 copy

Figure 2: Detail showing a single bookmark post

Semantic Network Analysis of Ontologies

Bettina Hoser and Andreas Hotho and Robert Jäschke and Christoph Schmitz and Gerd Stumme. *Proceedings of the 3rd European Semantic Web Conference \emph{(accepted for publication)}* (2006)

to web 2006 social ontology myown semantic analysis network sna by hotho and 1 other person on 2006-04-06 21:32:23 pick copy URL BibTeX

Figure 3: Detail showing a single publication post

The structure of a publication post displayed in BibSonomy is very similar, as shown in Figure 3. The first line shows again the title of the post, which equals the title of the publication in BIBTEX. It has an underlying link leading to a page which shows detailed information on that post. This line is followed by the authors or

bibsonomy.org.

editors of the publication, the journal or book title and the year. The following lines show the tags assigned to this post by the user, whose user name comes next, followed by a note on how many people tagged this publication. As described for bookmark posts, these parts link to the respective pages. After date and time of the posting follow the actions the user can do, which in this case include picking the entry for later download, copying it, accessing the URL of the entry or viewing the BIBTEX source code.

1.2 Additional Features

This section briefly describes some extensions of BibSonomy which go beyond the basic folksonomy model and have evolved during the practical use of the system.

1.2.1 Tag Hierarchy

Tagging gained so much popularity in the past two years because it is simple and no specific skills are needed for it. Nevertheless the longer people use systems like BibSonomy, the more often they ask for options to structure their tags. A user specific *binary relation* $\prec$ between tags as described in our model of a folksonomy (cf. [2] for details) is an easy way to arrange tags.

Therefore we included this possibility in BibSonomy and extended it further to use it for conceptual navigation. For instance, it is possible, given a tag, to show all posts with one of the subtags of the given tag.

Figure 4 shows details of the relation management in BibSonomy. Relations can be edited manually, but usually they will be created in the normal tagging process by tagging a resource, e. g. as *eclipse->java*, expressing that the tags *eclipse* and *java* should be assigned to a particular resource, and that the pair *(eclipse, java)* should be inserted into the user's $\prec$ relation. When browsing the folksonomy, a user can decide if she wants to make use of the tag hierarchy, e. g. when querying for the tag *java*, resources that are tagged with *eclipse* can be included in the result set, even if they do not have the tag *java* themselves. In order to distinguish between simple tag queries and those involving subtags, we call the latter

rename/replace all your tags

In all posts which contain all of the tags in the first box these tags will be substituted by the tags in the second box.

tag(s) to replace:

This example shows the substitution of the tags folksonomy and project by the tags social, system and bookmarking. This would be done in all posts which contain the tags folksonomy *and* project.

new tag(s):

BibSonomy
to bibtex folksonomy project by schmitz on 2007-02-12 11:19:11

BibSonomy
to bibtex social system bookmarking by schmitz on 2007-02-12 11:19:11

also update relations: ☐

Submit Query

insert relations
relations to insert: [ ] → [ ] Reset Submit Query

delete relations
relations to delete: [ ] → [ ] Reset Submit Query

relations
boomerang ← trifly
distribution ← eggenberger polya
eclipse ← yoxos
email ← sieve
event ← conference workshop
folksonomy ← tagging
functionalprogramming ← scheme
gnome ← linux
graphtheory ← graphgenerator smallworld
hobby ← boomerang cinema movies music
java ← eclipse jnlp webstart
kassel ← unikassel
knowledge ← knowledgemanagement
latex ← auctex
linux ← gnome zsh
management ← knowledgemanagement
math ← generatingfunctions statistics
music ← bass blues guitar radio
p2p ← p2pir p2pkm
person ← researcher
plane ← infill kanna
programming ← c c++
functionalprogramming java perl python ruby
protocol ← rest
rdf ← kaon

Figure 4: Relation and tag editor, with relations in the sidebar

one a query for *java* as a *concept*.

While it is not enforced upon the user how the ≺ relation is to be used, we expect that most of the time, it will be used in order to express subsumptions in an *is-a* hierarchy. The actual use we are observing confirms this assumption.

1.2.2 Duplicate detection.

In particular for literature references there is the problem of *detecting duplicate entries*, because there are large variations in how users enter fields such as journal name or author. On the one hand it is desirable to allow a user to have several entries which differ only slightly. On the other hand one might want to find other users' entries which refer to the same paper or book even if they are not completely identical. Hence it is necessary to map these entries together to allow such browsing functionality.

To fulfill both goals we implemented two hashes to compare publication entries at different levels of granularity. One is used to warn the user if she posts very similar BIBTEX entries twice, possibly creating unwanted redundancy. The other one is used to aggregate BIBTEX entries that were posted by more than one user in a common view, providing an opportunity to pick the most complete one or copy over missing data to one's own version of the publication entry. The implemented solution does not allow for a common creation of a single (correct) entry in a wiki style as we want to allow every user to store BIBTEX in the way she likes without changes from other users.

1.2.3 Import of resources.

To encourage users to transition from other systems, we implemented an *import* functionality. For del.icio.us, this functionality also takes into account the del.icio.us bundles which are named sets of tags. We map bundles to relations. Furthermore it is possible to import bookmark files of the Firefox[6] web browser, where the typical folder hierarchy of the bookmarks can be added to the user's ≺ relation.

Existing BIBTEX entries can be imported by uploading files, past-

---

ing BIBTEX snippets into BibSonomy, or just marking a BIBTEX entry on a web page and hitting the *post* button. Furthermore, for numerous digital library services including the *ACM Digital Library*, *SpringerLink*, *arXiv*, and *CiteSeer*, automatic screen scrapers for publication metadata are provided.

For unstructured publication metadata, such as the publication lists often found on researchers' web pages, semi-automatic extraction using the *Mallet*[7] information extraction tool is supported.

1.2.4 Export facilities.

The content of any page in the BibSonomy user interface can be presented in different formats for export. Exporting BIBTEX is accomplished by preceding the path of an URL with the string `/bib` – this returns all publications shown on the respective page in BIBTEX format. For example the page http://www.bibsonomy.org/bib/search/text+clustering returns a BIBTEX file containing all literature references which contain the words "text" and "clustering" in their fulltext.

HTML-formatted[8] publication lists can be exported which can be easily included into personal homepages. For example, a user *schmitz* might want to tag his own publications with *myown* and the year of publication. This enables him to include the page http://www.bibsonomy.org/publ/user/schmitz/myown+2006 into his personal web page to get an automatically updated list of his publications on the web.

Other exports such as XML, RSS and BURST[9] feeds, RDF according to the SWRC ontology, BIBTEX and EndNote work similarly.

Furthermore, links can be provided to an OpenURL[10] resolver. Such a resolver allows every user to find any publication presented in BibSonomy in her own local library.

1.2.5 Group Management and Privacy.

---

[6]http://www.mozilla.com/firefox/

[7]http://mallet.cs.umass.edu/

[8]A small heuristic is applied to handle special letters e.g. with accents but LATEX commands are not used to format the output.

[9]http://www.cs.vu.nl/~pmika/research/burst/BuRST.html

[10]http://www.exlibrisgroup.com/sfx_openurl.htm

In many situations it is desirable to share resources only among certain people. If the resources can be public, then one could agree to tag them with a special tag and use that tag to find the shared resources. The disadvantage is, that this could be undermined by other users (or spammers) by using the same tag. To solve this problem and also to allow resources to be visible only for certain users, we introduced *groups* in BibSonomy which gives users more options to decide with whom they share their resources.

2. CONCLUSION AND OUTLOOK

2.1 Summary

BibSonomy is, to the best of our knowledge, the only folksonomy system currently online which combines bookmark and publication management in a common user interface.

In addition to standard folksonomy features, BibSonomy provides capabilities for structuring a user's tag cloud, group and privacy management, and various import and export options, including screen scrapers for popular publication services as well as for unstructured publications lists.

2.2 Ongoing and Future Work

There are several important topics which we will be address in the near future. As stated in the introduction folksonomies can be seen as a lightweight knowledge representation. One major goal is therefore the convergence with the semantic web effort which is also called Web 3.0. To reach this goal a more machine understandable tagging is needed which can be reached be using so called "machine tags"[11] but also by developing new methods to extract semantics from folksonomies. Our next steps in this direction but also to enhance the usability are as follows:

Ranking The FolkRank [3] ranking algorithm has been developed which allows for a topic-specific ranking of folksonomy resources. Incorporating a ranking scheme to enhance the simple reverse-chronological presentation of posts is ongoing work.

Conceptual Clustering and Community Detection We are currently investigating different approaches [4, 3, 7] for finding coherent clusters within the folksonomy; these clusters could be viewed as communities of users being interested in common topics.

Tag Recommender and Ontology Learning As users are providing tag-tag relations in BibSonomy, we are currently investigating techniques which enable the semi-automatic learning of the tag-tag relation. A first step in that direction is the learning of subsumption relations using association rules on the folksonomy data set [7]. The same techniques can also be used to generate tag recommendations.

API Experience has shown, that an Application Programming Interface (API) is crucial for a folksonomy system to gain success. It is something which has been requested by many people and which allows for easy interaction of BibSonomy with other systems. Hence we are currently investigating several approaches to add an API to BibSonomy. Most systems use lightweight APIs similar to the idea of REST [1] which can be used and accessed in a very straightforward and easy-to-implement fashion. Nevertheless, with SOAP[12] there exists a

standard for web services which should also be taken into account. Since the process of defining an API for BibSonomy has just started, this is still an open task.

3. REFERENCES

[1] Roy T. Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, University of California, Irvine, 2000.

[2] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. BibSonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, pages 87–102, 2006.

[3] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *Proceedings of the 3rd European Semantic Web Conference*, LNCS, pages 411–426, Budva, Montenegro, June 2006. Springer.

[4] Robert Jäschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter, and Gerd Stumme. Trias - an algorithm for mining iceberg tri-lattices. Hong Kong, December 2006. (to appear).

[5] Leslie Lamport. *LaTeX: A Document Preparation System*. Addison-Wesley, 1986.

[6] Oren Patashnik. BibTeXing, 1988. (Included in the BIBTEX distribution).

[7] Christoph Schmitz, Andreas Hotho, Robert Jäschke, and Gerd Stumme. Mining association rules in folksonomies. In *Proc. IFCS 2006 Conference*, pages 261–270, Ljubljana, July 2006.

[8] S. Staab, S. Santini, F. Nack, L. Steels, and A. Maedche. Emergent semantics. *Intelligent Systems, IEEE [see also IEEE Expert]*, 17(1):78–86, 2002.

[9] L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1(2):169–194, October 1998.

[10] Thomas Vander Wal. Folksonomy definition and wikipedia. November 2005.

---

[11] http://www.flickr.com/groups/api/discuss/72157594497877875/
[12] http://www.w3.org/TR/soap/