

Managing Publications and Bookmarks with BibSonomy

Dominik Benz, Folke Eisterlehner, Andreas Hotho,
Robert Jäschke, Beate Krause, Gerd Stumme
University of Kassel & L3S Research Center
{lastname}@cs.uni-kassel.de

ABSTRACT

In this demo we present BibSonomy, a social bookmark and publication sharing system.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
H.3.7 [INFORMATION STORAGE AND RETRIEVAL]:
Digital Libraries—*Collection, Systems issues*

General Terms

Management

Keywords

folksonomy, tagging, web 2.0, tools

1. INTRODUCTION

Complementing the Semantic Web effort, a new breed of so-called “Web 2.0” applications recently emerged on the Web. These include user-centric publishing and knowledge management platforms like Wikis, Blogs, and social resource sharing tools.

BibSonomy¹ is one of the social resource sharing tools that have acquired large numbers of users within the last years. The reason for their immediate success is the fact that no specific skills are needed for participating, and that these tools yield immediate benefit for the individual user (e.g., organizing ones bookmarks in a browser-independent, persistent fashion) without too much overhead. Social resource sharing systems all use the same kind of lightweight knowledge representation, called *folksonomy*.

In this paper we describe BibSonomy. It allows to share both bookmarks and publication meta data. A more detailed description of BibSonomy can be found in [1]. BibSonomy started as a student project at the Knowledge and Data Engineering Group of the University of Kassel² in spring 2005. The goal was to implement a system for organizing BIB_TE_X entries in a way similar to bookmarks in Delicious – which was at that time becoming more and more popular. BIB_TE_X is a popular literature management system for

¹<http://www.bibsonomy.org/>

²<http://www.kde.cs.uni-kassel.de/>

L_AT_EX, which many researchers use for writing scientific papers. We soon decided to integrate bookmarks as a second type of resource into the system. After we finished the first round of implementation for the user interface, we had to tune the SQL queries for scalability reasons. At the end of 2005, we announced BibSonomy first to some colleagues, later in 2006 to the public. Since then, the number of users has steadily grown. We implemented several useful features and redesigned the architecture to ease future developments. Our team and other research groups use BibSonomy or its data for research, and we have implemented our research results into the system, e.g. the FolkRank algorithm and tag recommendation methods – both for the benefit of the users and to directly measure the performance of our methods.

2. USER INTERFACE

A typical list of posts is depicted in Figure 1 which shows bookmark and publication posts 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 a search box), 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 that are showing posts; it allows for navigation in all dimensions of the folksonomy. Beside this kind of pages, systems like BibSonomy typically contain summary pages representing the content in form of a cloud. The page with the global tag cloud summarizes in a clear way the content of the system by the used tags. Note that on selected pages posts can be ordered by relevance as calculated by the FolkRank algorithm.

3. ARCHITECTURE

The basic building blocks of BibSonomy are an Apache Tomcat³ servlet container using Java Server Pages⁴ and Java Servlet⁵ technology and a MySQL⁶ database as backend.

Currently the project has several thousand lines of code and is using the Model View Controller (MVC) programming paradigm to separate the logical handling of data from the presentation of the data. This enables us to produce output in various formats (see Section 4), since adding a new output format is accomplished by implementing a JSP as a view of the model.

³<http://tomcat.apache.org/>

⁴<http://java.sun.com/products/jsp>

⁵<http://java.sun.com/products/servlets>

⁶<http://www.mysql.com/>

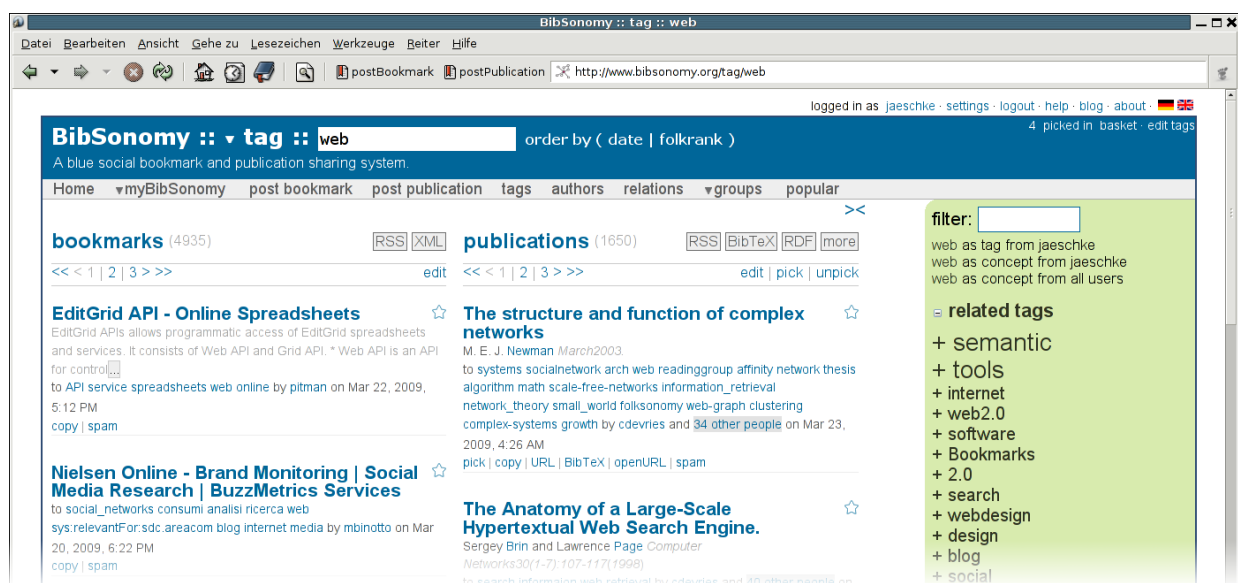


Figure 1: BibSonomy displays bookmarks and Bib_TE_X based bibliographic references simultaneously.

The central database schema of BibSonomy is based on four tables: One for bookmark posts, one for publication posts, one for tag assignments (*tas*) and one for *relations* between tags. Two further tables store information regarding *users* and *groups*.

The posts tables are connected with the *tas* table by the key *post.id*. The scheme is not normalized – on the contrary we have added a high amount of redundancy to speed up queries. For example, besides storing group, user name and date in the posts table, we also store it in the *tas* table to minimize the rows touched when selecting rows for the various views. Furthermore, several other tables hold counters (i. e., how many people share one resource, how often a tag is used, ...). Finally a large set of indexes (12 in the *tas* table alone) build the basis for a fast answering of queries.

Overall, we spent a large amount of work for investigating and optimizing SQL queries and table schemes and tested both with folksonomy data of up to 8.000.000 posts. At the moment, we need no special caching or physical distribution of the database to get reasonable response times.

4. FEATURES

The most simplistic but also most laborious way to add posts to BibSonomy is by entering their metadata manually into form fields. To lower the efforts to get data into BibSonomy, it supports various ways to import resources from files and web pages e.g. file import from Bib_TE_X or Endnote⁷ or by so called “scrapers” which allow to automatically extract publication metadata from digital libraries like SpringerLink⁸. Nevertheless – forms are still used to edit posts.

Exporting publication references in Bib_TE_X format is accomplished by preceding the path of a URL showing publication posts with the string `/bib` – this returns all publications shown on the respective page in Bib_TE_X format. For ex-

ample the page <http://www.bibsonomy.org/bib/search/text+clustering> returns a Bib_TE_X file containing all literature references which contain the words “text” and “clustering” in their fulltext.

More general, every page which shows posts can be represented in several different ways by preceding the path part of the URL with a specific string to specify the export format, e.g. `/xml` for bookmarks in XML format or `/publ` for publications in a simple HTML format suited for the integration into a web page (for an integration example see <http://www.kde.cs.uni-kassel.de/pub>). For an overview of the available export formats for publications, one can use the `/export` path extension which is also linked on all web pages showing publication posts. The export feature allows to generate publication lists for external websites, e.g. for personal and institute webpages or for project pages.

Experience has shown, that an Application Programming Interface (API) is crucial for a system to gain success. Hence we have implemented a lightweight REST API⁹ which can be used and accessed also by less experienced programmers. We use the API for the integration of JabRef¹⁰. The catalog of the library of Cologne uses the API to access tagging information for its books.

There are several other valuable features like the publication basket, the duplicate detection mechanism, a tag editor, or the integration with other systems. A description of those features can be found in [1].

5. REFERENCES

- [1] A. Hotho, R. Jäschke, C. Schmitz, and G. 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.

⁹<http://www.bibsonomy.org/help/doc/api.html>

¹⁰<http://bibsonomy.blogspot.com/2009/02/feature-of-week-bibsonomy-plugin-for.html>

⁷<http://www.endnote.com/>

⁸<http://www.springerlink.de/>