CitationBase: A social tagging management portal for references

Martin Hofmann
Department of Computer Science, University of Innsbruck, Austria
m_ho@aon.at

Ying Ding
School of Library and Information Science, Indiana University, Bloomington, USA
dingying@indiana.edu

Abstract

Social tagging is one of the major phenomena brought by the social media and technologies in the Web2.0. It allows users to organize and share their information and online resources on the Web. It enables the content management for the communities as well as for individual person. In order to explore the possibilities of applying ontology and semantic web data integration approaches to similar area, we implement a system called CitationBase which aims to provide the reference management for community and individual users based on the existing technologies of social web and the semantic web.

Keywords: social tagging, citation, web2.0, social web, semantic web

Introduction

Web 2.0 changes the current Web into a Social Web where ordinary users can meet, collaborate and share information by using tagging, bookmarking and annotating. Some popular examples are bebo, Delicious, Digg, Flickr, Google maps, Skype, Technorati, Wikipedia, last.fm and so on. It provides easy-to-use social software for massive online publishing and sharing. Social software lets people rendezvous, connect or collaborate by use of a computer network and furthermore leads to the direction of the creation of shared and interactive spaces.

There are a couple of existing efforts trying to bring the Web to its next generation. The Semantic Web is one of the efforts embedded significantly in artificial intelligence area. It has the long-term vision to make the Web as the global brain of human and machine by representing data in machine understandable way and automating the mediation of data and services. Meanwhile, Web 2.0 represented Social Web has successfully motivated users to share information and collaborate with each other (Hichcliffe, 2006; O'Reilly, 2005).

Web 2.0 is not completely different from the Semantic Web. As Sir Tim Berners-Lee mentioned "the Semantic Web is an extension of the current Web in which information is given well-defined meaning,

better enabling computers and people to work in cooperation [1] ". Web 2.0 not only extends the communication dimensions (publishing and commenting) but also tries to add extra contextual information to the current Web in a social and informal way (e.g. tagging, bookmarking and annotating). The power of the Semantic Web lies in the potential for interoperability through some well-defined metadata in machine understandable way with logical reasoning support. Layered design principle in the Semantic Web paves the way for reuse (Antoniou and van Harmelen, 2004). Web 2.0 provides scalable information sharing platform, while the Semantic Web adds valuable machine understandable metadata to enable efficient and automatic way of information sharing and cross-portal communication and collaboration. The combination of the Semantic Web with Web2.0 forms a new momentum for the next web weave coined as Web 3.0 in the New York Time[2].

Social tagging is one of the major phenomena brought by the social media and technologies in the Web2.0. It allows users to organize and share their information and online resources on the Web (Li, et. al., 2008; Lin, et al., 2008). It enables the content management for the communities as well as for individual person. For managing references, there are already some available social network websites, such as CiteULike[3] and Bibsonomy[4]. In order to explore the possibilities of applying ontology and semantic web data integration approaches to similar area, we implement a system called CitationBase which aims to provide the reference management for community and individual users based on the existing technologies of social web and the semantic web.

CitationBase

CitationBase provides the central reference space where community members can go there to browse and search interesting references which are either imported through DBLP dumps or shared by the community members. User himself can also have his own personal content space where he can manage his own favorite references. All users are allowed to tag any public references shared in the central reference space and his own personal references (Marlow, et. al., 2006) (see Fig 1).
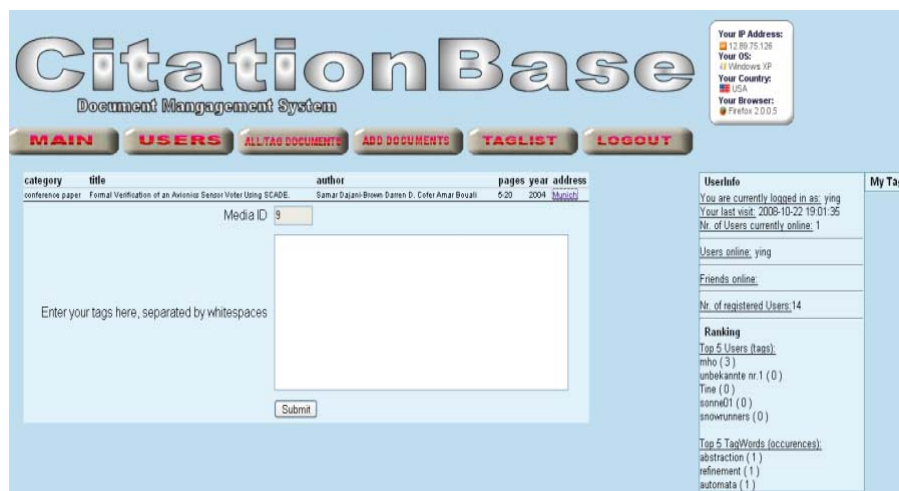

Fig 1: Tagging function of the CitationBase

CitationBase allows two ways to contribute references. Reference can be imported from an external XML, RDF or text files (such as DBLP dumps) or entered by the users themselves based on provided templates. These templates are built up on the reference ontologies, which currently only contain book ontology for book references, conference paper ontology for conference paper and journal paper ontology for journal paper (Fig 2 and Fig 3). It will be further extended in our future work.
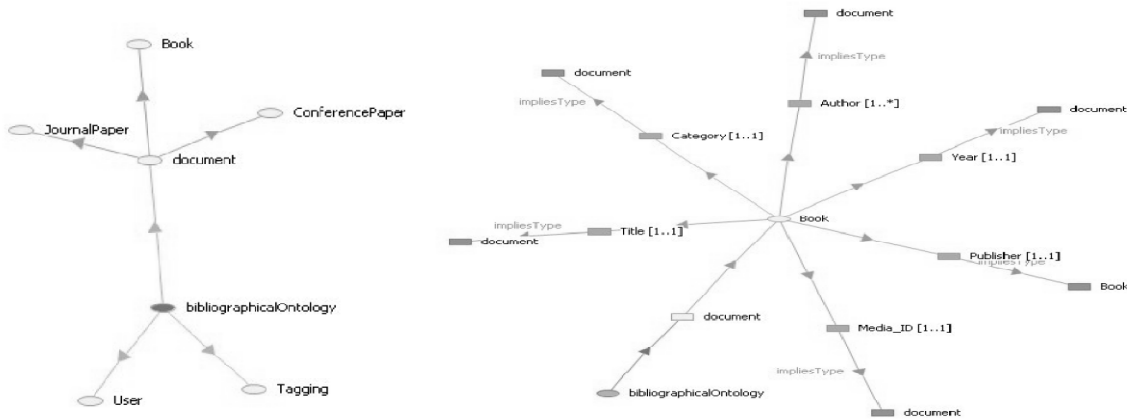

Fig 2: Overall ontology of the system and the book ontology.

Fig 3: Reference input templates

The key design idea of the CitationBase is to show the flexibility of data integration and mashup by using the semantic web technology. Based on the reference ontologies in the CitationBase, input data can be integrated together and linked with other LOD[5] bubbles. Interlinking with LOD data provides possibility of adding extra background information for the data, such as geographical distribution of the conference papers, the Wikipedia page of the key concepts in the paper titles, the FOAF data of the authors and so on. Current implementation of the CitationBase integrates the conference papers with the DBPedia and Geodata. It can show the map of the conference paper distributions based on the location of the

---

conferences (see Fig 4). Data can be exported based on the user defined needs which means that the users are allowed to drag around the components of the reference data and design their own output templates which fit to different requirement by different journals, conference proceedings and publishers.
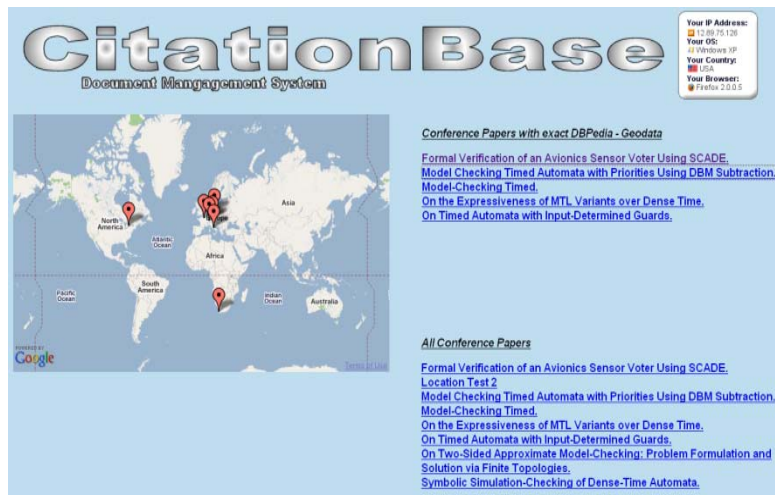

Fig 4: Demo of geo-location of the conference papers

Discussion and Conclusion

CitationBase provides the key function of the portable references similar as Delicious which provides the portable bookmarks. It will further be enhanced by including other social network features so that domain experts, social trends and many others relationships can be identified through collective social tagging. CitationBase also demos the role of the ontology-based data for easy data integration. It eases the tasks for handling input and output data and also the integration with other available semantic data. CitationBase as a community portal for tagging references also provides a unique way to create social ontologies (i.e. folksonomy) as tags can be viewed as social metadata.

Folksonomy is generated collaboratively by community users to use tags to categorize content. In contrast to controlled vocabularies, folksonomies are a distributed classification system with low entry costs. Folksonomies and tag clouds can be viewed as shallow ontologies with rather loosely organized flat structures, while ontologies are defined carefully to reflect the parts of the world in formal, explicit and machine-processable ways (Mika, 2005). Ontologies are carefully defined to remove ambiguity while tag is a loose and implicit process where ambiguity still remains. The inference process applied to ontologies is logic based while the inferential process on tags is statistical based. Although folksonomies and tags are quite different comparing to ontologies, they share the same purpose of organizing and sharing web information. Folksonomies and tags offer social approaches to manage Web information. They have a big potential to reform the current ontological approach to make it scalable and easy-to-use. The combination of two is a good trial such as using FOAF to share bookmarks in Delicious. Instance-based mediation by data mining techniques could provide scalable mediation based on the tagging activities of community users. Besides, some existing widely-adopt social metadata can smooth integration processes since data are represented in pre-defined shared consensual way.

Current semantic efforts try to facilitate integration by using ontologies. But the current methodology for ontology generation, mediation and versioning is not routine enough to be deployed in daily life (Ding and Foo, 2002). As Shadbolt, Berners-Lee and Hall (2006) point out that ontology must be developed, managed and endorsed by practice communities. Communities and practice should be involved in

ontology engineering and Web 2.0 provides ideal platform and tools for community users to easily add social metadata and annotate data.

References:

Antoniou, G. and van Harmelen, F. (2004). *A Semantic Web Primer*. MIT Press (ISBN 0-262-01210-3).

Ding, Y. and Foo, S. (2002). Ontology Research and Development: Part 1 – A Review of Ontology Generation. *Journal of Information Science*, 28(2): 123-136.

Hinchcliffe, D. (2006). The State of Web 2.0. *Web Services Journal*. Available: http://web2.wsj2.com/the_state_of_web_20.htm (visited at Feb 10, 2009).

Li, X.; Guo, L.; Zhao, Y. (2008). Tag-based Social Interest Discovery. In *Proceedings of the World Wide Web 2008*, April 2008, Beijing, China.

Lin, Y.; Chi, Y.; Zhu, S.; Sundaram, H. and Tseng, B. (2008). FacetNet: A Framework for Analyzing Communities and Their Evolutions in Dynamic Networks. In *Proceedings of the World Wide Web 2008*, April 2008, Beijing, China.

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*.

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. In Proceedings of the Fourth International Semantic Web Conference (ISWC-2005), *Lecture Notes in Computer Science*, vol. 3729, Galway, Ireland, November 2005. Springer.

O'Reilly, T. (2005). What is Web 2.0. O'Reilly Network. Available: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (visited at Feb 10, 2009).

Shadbolt, N.; Berners-Lee, T. and Hall, W. (2006). The Semantic Web revisited. IEEE Intelligent Systems, 21(3): 96-101.