Proceedings of the International Conference for Internet Technology and Secured Transactions (ICITST-2008), June 23–28, 2008, DIT, Dublin, Ireland. ISBN 0-9546628-2-2 www.icitst.org/2008 © e-Centre for Infonomics

ICITST-2008
3rd International Conference for Internet Technology and Secured Transactions
June 23–28, 2008, Dublin, Ireland (www.icitst.org)

Ontological Data Freshness on the Social Web

Alexandra Roshchina*, John Cardiff*, Paolo Rosso**, Alexander Troussov***

* Institute of Technology Tallaght, Dublin, Ireland; sasharo@itnet.ie, John.Cardiff@ittdublin.ie
** Natural Language Engineering Lab, Universidad Politécnica de Valencia, Spain; prosso@dsic.upv.es
*** IBM LanguageWare, Dublin Software Lab, Ireland; atrousso@ie.ibm.com

Abstract. The construction process of the traditional heavyweight ontology precisely describing a specific area is a time consuming task. But in the field of constantly changing dynamic areas like the Web it is impossible to produce a complete ontology to accurately reflect any particular domain of interest. If information in the domain is changing rapidly the corresponding ontology should be constantly enriched with newly emerged concepts and relations. The problem of ontology enrichment becomes even more crucial now with the emergence of social networking services and e-learning domains with highly dynamic content. In the paper we propose an experiment aimed at constructing ontology of interests based on the data provided by the Delicious online social service. This ontology will then be used as raw material for our main goal of addressing the challenge of improving or enriching the ontological structure by developing techniques and mechanisms for capturing and representing the "hidden information" in ontology.

Keywords. Semantic Web, ontology of interests, Galaxy, ontology enrichment, Social Web

1 OVERVIEW: THE SOCIALIZATION OF THE WEB

During the years of its existence the Web has changed a lot due to the process of socialization that it has undergone practically unnoticed by most of its users. Many popular social services have appeared on the Web including blogs, wikis, file sharing tools (like Flickr, http://flickr.com, and Delicious, http://del.icio.us), etc.

Users can publish web content with these new technologies without knowledge of even HTML. With their help the Web has gained new features changing the ways of communication and interaction between people. The socialization of the Web occurred in a way that is practically unnoticed for most of the users but its consequences are significant. The traditional approach of firstly gathering information for a site and then displaying it for users has given way to freely editable pages thus giving people the opportunity to be the creators of the content. With the contribution of each community member overall knowledge available broadens from being the simple sum of all experiences of individuals.

Further development of the Internet is now associated with bringing semantics forming the Social Semantic Web [1]. This revolutionary approach is not going to change the existing Web completely but will be an extension of the Web in which the semantics of information and services on the web are defined [2]. By analyzing semantic connections between objects and concepts on the Web semantic agents will select only those facts that are really relevant to the user's request.

This paper is organized as follows. First, we give the definition for the ontology and an overview of existing ontologies on the Web addressing the task of ontology enrichment. Second, we provide the description of our experiment. We conclude with a greater vision for our work.

2 ONTOLOGY CREATION AND THE SOCIAL WEB

Because the main idea of the Semantic Web is to make the content of the pages meaningful for machine agents, the representation of the background knowledge becomes an important issue performing the shared interpretation between various applications [1]. This knowledge is traditionally defined in ontology, which is a formal shared description of concepts and relations between them within a specific domain of interest [3].

A lot of projects aimed at constructing ontologies have recently appeared on the Web. Many of them exist in the public domain so they can be used by developers (like DMOZ, http://www.dmoz.com, OpenCyc, http://www.opencyc.org/, etc).

With the information on the Web changing rapidly the issue of ontology modification over time and therefore the necessity of its maintenance is vital. Recently considerable interest has been paid to folksonomies emerging in the process of collaborative tagging (people annotate pieces of content using one-word descriptions or tags) [4]. Attaching tags to objects has become a new solution to the problem of classi-

fication traditionally carried out by specially trained professionals or by authors. The main advantage of folksonomy is that tags are not imposed but appear spontaneously reflecting the real vocabulary of users. Folksonomy is often created automatically from the set of all tags within the system using popularity based approaches, co-occurrence algorithms and clustering techniques.

There are a number of ontology creation projects oriented primarily towards Semantic Web applications. For example, FOAF (Friend of a Friend RDF-vocabulary), originally created by Dan Brickley and Libby Miller, is aimed at describing people, relations between them, their interests [5]. FOAF ontology has become very popular with the emergence and fast spreading of social on-line networking services (like MySpace, Facebook, etc.). Another project SIOC (Semantically-Interlinked Online Communities initiative) is working in the area of bringing semantics to the social communities like blogs, forums, mailing lists, etc for interconnecting these isolated services, etc [6].

3 EXPERIMENT: ONTOLOGY OF INTERESTS

3.1 Description

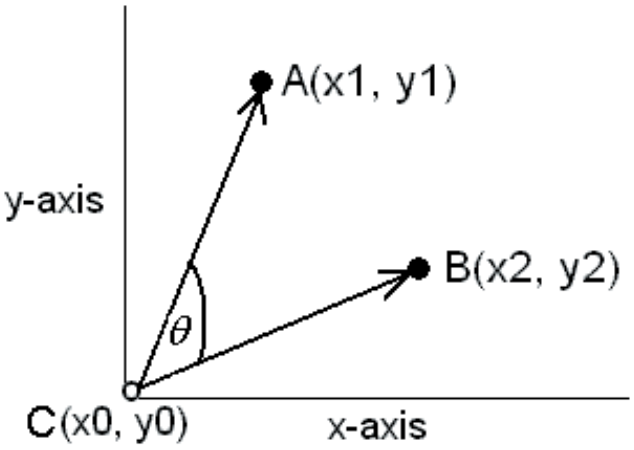
From the data available at Delicious the list of users will be gathered with the circle of their friends and fans to construct the social network. From parsing HTML pages describing those people a set of tags they use will be collected viewed as their interests. We will use the APML (Attention Profiling Markup Language, http://apml.pbwiki.com) format to work with users' interests. This standard provides a way of saving users' preferences: what he/she likes (from the Web browsing history) and also how much he/she is interested in those issues. A number of on-line services are already using it including Cluztr (http://cluztr.com) social network, Engagd (http://www.engagd.com) tool, etc. Construction of the APML profile will help the user to find the information on the Web that will match his/her personal interests more easily and accurately.

To construct the ontology of interests each interest will be represented as a vector $I(i_1, i_2,..i_n)$ where i_k is a weight of this interests for the user k and n is a number of users in the network. A relation between two interests will be established according to their cosine similarity (a distance between two vectors in n-dimensional space) (see Equation 1 and Figure 1) [7]:

$$sim(I_i, I_j) = \frac{(I_i, I_j)}{\|I_i\| \times \|I_j\|} = \frac{\sum_{k=1}^{n} i_{ik} \times i_{jk}}{\sqrt{\sum_{k=1}^{n} i^2_{ik}} \times \sqrt{\sum_{k=1}^{n} i^2_{jk}}} \quad (1)$$

The resulting ontology with the network of users will be stored in Galaxy to provide recommendations for people through the web interface (people matching you with a variable number of interests and those who are nearest to you in the network of social connections) (see Figure 2).

Figure 1. Distance between two vectors in two-dimensional space



3.2 Delicious

To provide us with the experimental data the social bookmarking service Delicious will be used. Through the use of tags, semantics are brought to the content of the site. Allowing people to annotate the content they are uploading to the social system (as well as associating the existing item with the concept represented by the tag) is effective in a number of ways:

- Users are able to organize data as a part of Personal Information Management. They can save links they find useful and interesting, share them with others and add tags to annotate the content of the web site.
- It provides a new approach of classifying objects done by a wide group of users in a collaborative manner.
- Tagging systems reflects the real terminology and vocabulary of people.

As users store in the system data that is somehow significant to them, tags used for annotation can be viewed as users' interests. Thus by parsing tag clouds (a visualized set of user-generated tags where the frequency of the tag is reflected by size or color) of a large amount of people the ontology of interests can be dynamically constructed and updated constantly with the appearance of new tags and members of the online community [8].

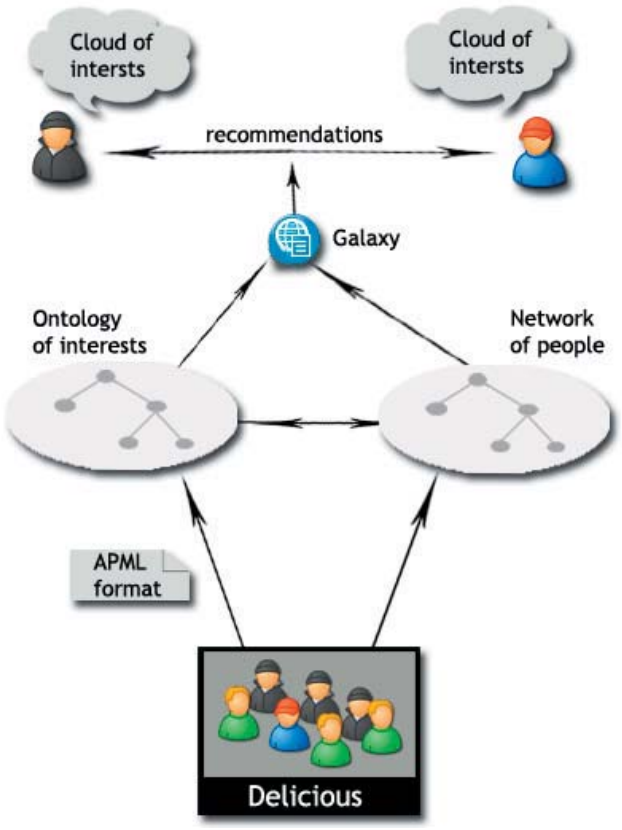
3.3 Galaxy

Galaxy (IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks) is a unified API provided by IBM addressing the tasks of maintaining multidimensional networks and ontology-based semantic processing. Galaxy was created within the NEPOMUK project (http://nepomuk.semanticdesktop.org) [9]. When dealing with all kinds of data based on a graph network representation (social networks, organization's structure, etc.), Galaxy can be extremely useful [10]. It can establish connections between different con-

Proceedings of the International Conference for Internet Technology and Secured Transactions (ICITST-2008), June 23–28, 2008, DIT, Dublin, Ireland. ISBN 0-9546628-2-2 www.icitst.org/2008 © e-Centre for Infonomics

cepts as the result of applying soft clustering techniques. It performs the spread of activation algorithm to discover the focus of the sub-graph forming around the input nodes. Those input nodes, according to the accepted metaphor, appear to be sources of light which is then propagated though the graph following its links. Those node(-s) that receives the largest amount of light (a combination of light from initial sources) is considered to be the focus for the performed query [11].

Figure 2. The visualization of the experiment



4 CONCLUSIONS AND FUTURE WORK

In our research we address the task of ontology enrichment which is vital for highly dynamic areas like the Web. The constantly updating Delicious folksonomy can be viewed as the source of appearance of new concepts reflected by tags. It also provides URL collections grouped by users' tags to represent the general belonging of the document to some category. This fact can be used in text mining tasks when the new term is found in the text. In this paper we have proposed a method of construction of the ontology of interests based on the information available at Delicious. This ontology can be used for future mapping of FOAF documents of Semantic Web researchers to it. As the information on the Web is changing continuously finding tendencies in scientific discussions is vital. Uncovering such tendencies and thus enriching the ontology of interests can be useful for researchers to stay informed about the activities taking place in the mainstream and provide them with possible directions of research.

The significance of this project is that we can provide users with a more complete picture of their domain of interest, not solely restricted to their present requirements. We can also suggest predicative links between concepts which may not be present in the ontology today, but which may arise in the future. We intend to validate the results of the project in the fields of e-learning and social networks.

The gathered information reflected by the enriched ontology can be then represented as a web application that will recommend people the most challenging areas of newly appearing research interests according to the interests of a particular researcher listed in his FOAF profile.

ACKNOWLEDGEMENTS

The work of Alexandra Roshchina was funded by the TSR Postgraduate R+D Training Program. The work of Paolo Rosso was partially funded by the MCyT TIN2006-15265-C06-04 research project.

REFERENCES

1. Gruber T. Collective knowledge systems: Where the Social Web meets the Semantic Web. J. Web Sem. 6(1): 4-13 (2008).
2. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. Scientific American, May 2001.
3. T. R. Gruber. A translation approach to portable ontology specification. Knowledge Acquisition 5, 1993.
4. Specia L., Motta E. Integrating Folksonomies with the Semantic Web. In proceedings of the 4th European Semantic Web conference (ESWC 2007), Innsbruck, Austria, Springer, 2007.
5. Ding L., Zhou L., Finin T., Joshi A. How the Semantic Web is being used: an analysis of FOAF documents. In Proceedings of the 38th International Conference on System Sciences, January 2005.
6. Harth A., Breslin J.G., O'Murchu I., Decker S. Linking semantically-enabled online community sites. Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and the (Semantic) Web (FOAF Galway), Galway, Ireland, pp. 19-29, September 2004.
7. Sarwar B., Karypis G., Konstan J., Riedl J. Item-based Collaborative Filtering Recommendation Algorithms. In Proceedings of the 10th International World Wide Web Conference, pp. 285-295, 2001.
8. Smith G. (2008), Tagging: people-powered metadata for the Social Web, New Riders Press.
9. Groza T., Handschuh S., Moeller K., Grimnes G., Sauermann L., Minack E., Mesnage C., Jazayeri M., Reif G., Gudjonsdottir R. The NEPOMUK Project - On the way to the Social Semantic Desktop. Proceedings of I-Semantics, pp. 201-211, 2007.
10. Judge J., Sogrin M., Troussov A. Galaxy: IBM Ontological Network Miner. CSSW 2007, 157-160.
11. Kinsella S., Harth A., Troussov A., Sogrin M., Judge J., Hayes C., Breslin J.G. Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects. The 4th Conference on Applications of Social Network Analysis (ASNA 2007), University of Zurich, Switzerland, September 2007.