

Social Web Communities

Executive Summary of the Dagstuhl Seminar

Harith Alani,¹ Steffen Staab,² Gerd Stumme³

¹ Address Harith

² Research Group ISWeb — Information Systems and Semantic Web,
University of Koblenz-Landau, Univ.str. 1, 56016 Koblenz, Germany
<http://isweb.uni-koblenz.de>

³ Knowledge & Data Engineering Group, Department of Mathematics and Computer Science,
University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany
<http://www.kde.cs.uni-kassel.de>

Abstract. Blogs, Wikis, and Social Bookmark Tools have rapidly emerged on the Web. The reasons for their immediate success are that people are happy to share information, and that these tools provide an infrastructure for doing so without requiring any specific skills. At the moment, there exists no foundational research for these systems, and they provide only very simple structures for organising knowledge. Individual users create their own structures, but these can currently not be exploited for knowledge sharing. The objective of the seminar was to provide theoretical foundations for upcoming Web 2.0 applications and to investigate further applications that go beyond bookmark- and file-sharing.

The main research question can be summarized as follows: How will current and emerging resource sharing systems support users to leverage more knowledge and power from the information they share on Web 2.0 applications? Research areas like Semantic Web, Machine Learning, Information Retrieval, Information Extraction, Social Network Analysis, Natural Language Processing, Library and Information Sciences, and Hypermedia Systems have been working for a while on these questions. In the workshop, researchers from these areas came together to assess the state of the art and to set up a road map describing the next steps towards the next generation of social software.

1 Topic of the Seminar

Within the last two years, social software on the Web, such as Flickr, Delicious, Bibsonomy, Facebook, etc., has received a tremendous impact with regard to hundred of millions of users. A key factor to the success of social software tools in the Web is their grass-roots approach to sharing of information between users: there are no limitations on the kind of tags users may select. The resulting structures are often called ‘folksonomies’, that is, ‘taxonomies’ created by ‘folks’.

Such systems are also considered to realize a Web version 2.0. The reason is that the initial use of the Web could be characterized by many users consuming what a comparatively small set of producers had developed, whereas with social software on the Web, everyone becomes a prosumer, i.e. someone who produces *and* consumes content. The

success of this approach is visible with applications like flickr,⁴ which had approximately 250,000 users in April 2006. In the reference sharing systems CiteULike⁵ and Connotea,⁶ researchers and others insert, tag, and recommend scientific references in a shared knowledge space. This indicates a currently ongoing grass-root creation of knowledge spaces on the Web which is closely in line with “the 2010 goals of the European Union of bringing IST applications and services to everyone, every home, every school and to all businesses” [5].

The reason for the apparent success of the upcoming tools for web cooperation (wikis, blogs, etc.) and resource sharing (social bookmark systems, photo sharing systems, etc.) lies mainly in the fact that no specific skills are needed for publishing and editing. As these systems grow larger, however, the users will feel the need for more structure to better organize their resources and enhance search and retrieval. For instance, approaches for tagging tags, or for bundling them, are currently discussed on the corresponding news groups. We anticipate that resource sharing systems, together with wikis and blogs, are only first appearances of an emerging family of Web 2.0 tools.

2 Research Objectives

The main objective of the seminar was to answer the question: *How will current and emerging resource sharing systems support untrained users in sharing knowledge on the Web within the next few years?* This question can be split down in (at least) three tasks:

1. Which new modes for knowledge sharing can be established on the Web, that are suitable for a general audience?
2. Which other types of information (beside just tags) would increase the usefulness of knowledge sharing on the Web?
3. How can methods and techniques from knowledge engineering and management be adopted and/or developed to suit these modes in order to allow a minimal-invasive use?

The following research areas — and in particular joint activities between them — have the potential, based on years of previous work, to contribute to these tasks:

- Semantic Web
- Knowledge Discovery
- Information Retrieval
- Information Extraction
- Social Network Analysis
- Natural Language Processing
- Library and Information Sciences
- Hypermedia Systems

⁴ <http://www.flickr.com/>

⁵ <http://www.citeulike.org>

⁶ <http://www.connotea.org>

– Complex Systems Science.

In order to set up any online community service, the whole knowledge life cycle has to be considered including issues of interaction design, self-explanatory user interfaces and issues with the complexity of the knowledge representation paradigm. The success of Web 2.0 systems has shown that it is fruitful to start with a very lightweight folksonomy/ontology representation. They also demonstrate that over time users discover the need for richer representations: In Flickr, users have invented a syntax for representing the geographical latitude and longitude of images resembling the structure of a Semantic Web representation in RDF. In Facebook, plugin developers allow their users to define new properties of facebook participants ('good athlete', 'dances well' etc.) and allow for elaborate comparisons of people with such attributions.

Obviously, there is a big need for a very easily understandable user interface with core functionalities, which may later be extended to include more complex knowledge representation facilities. By this way Web 2.0 systems also avoid the complexity overload that Semantic Web applications easily fall prey to when they target full coverage of underlying representations from the very start.

Based on these observations, we believe that the future lies in a convergence of stronger knowledge representation techniques (eg ontologies) with the grass-root approaches found in Web 2.0 systems. In order to support this convergence, contributions of all areas listed above are needed. At our Dagstuhl seminar, experts from these fields with interest in Web 2.0 topics gathered together. In small, mixed groups, the participants discussed the existing and potential contributions of their respective research areas to an overall picture of online community applications.

3 Workshop Schedule

The main objective of the workshop was to initiate discourses. Therefore, the major part of the week was assigned to work in small groups that were constituted from people of different areas. Only at the beginning, we had short presentations by individuals, to get to know each other.

Day 1: short presentations of the participants (approx. 10 min. each).

Day 2: splitting up in four working groups on specific topics.

Day 3: morning: plenary session with presentation of intermediate results of the working groups. Afternoon: excursion to Trier and to a winery.

Day 4: morning: continuation of the working groups. Afternoon: joint plenary session with the Perspectives Workshop "Virtual games, interactive hosted services and user-generated content in Web 2.0" on privacy issues in the Web 2.0.

Day 5: plenary session to gather the working group results. Discussion of next steps.

4 Working Groups and Panel Discussion

At the end of the first seminar day, each participants was asked to note two topics of interest. These topics were collected on a white board, and clustered interactively

by all participants. Five clusters resulted from this process, each becoming the topic of one working group. During the first part of the group phase, the groups finetuned their interests. One large group decided to split, and two groups dissolved and merged into the other groups. The resulting four groups worked together until the end of the seminar. In addition to the working groups, we had a panel discussion together with the Perspectives Workshop. The outcome of the working groups and of the panel discussion is briefly outlined below. More complete descriptions can be found in this volume.

Working group “The Berners-Lee Hypothesis: Power Laws and Group Structure in Flickr”. The hypothesis of Tim Berners-Lee, that the structure of online groups should conform to a power law distribution, was set in relation to the Dunbar number, a supposed limit for the number of social contacts an individual can have. It was also shown that public and private groups in Flickr show differences in their internal structure.

Working Group “The Evolution and Dynamics of Research Networks”. The group discussed the differences and relationships between social networks and homophily networks. While links in the former are explicit links between individuals (e. g., by the is-friend-of relation or the has-emailed-to relation), links in the latter result from some similarity in behavior (e. g., has-tagged-the-same resource-as or has-co-cited-the-same-paper-as). The working group discussed how homophily can lead to social relations, and in which settings such a correlation can be observed. Furthermore, the extension to dynamically evolving networks has been addressed.

Working Group “Mining for Social Serendipity”. At social events, where most of the participants do not know each other (e. g., at conferences), they tend to gather with people they already know. The group discussed how the identification of potentially interesting, but yet unknown persons can be supported by mining surprising relationships between people attending an event. As application scenario, the geo-locations of conference-related photos that were stored in Flickr have been used.

Working Group “Analyzing Tag Semantics Across Collaborative Tagging Systems”. The working group decided from the beginning to compare their different data analysis techniques in a hands-on approach on the same data sets. The aim was to analyze the semantic content of tags in the two social tagging systems Flickr and Delicious. Main findings were that tag context similarity can also be measured in a narrow folksonomy like Flickr, that Flickr and Delicious have little semantic overlap, that the tag-tag-space is structured into regions with high density, and that the order of the tags inside a post is semantically significant.

Joint Panel Discussion with the Perspectives Workshop “Virtual games, interactive hosted services and user-generated content in Web 2.0”. When users act as prosumers this does not only affect how knowledge is generated collectively, but it also affects the legal implications of such user generated content. The perspectives workshop targeted legal issues of such content and during the joint session, implications of content being generated, modified, distributed and owned with different licensing models was heavily discussed.

References

1. Berners-Lee, T.: Realising the full potential of the Web, 1997. <http://www.w3.org/1998/02/Potential.html>
2. Berners-Lee, T.: A roadmap to the Semantic Web, 1998. <http://www.w3.org/DesignIssues/Semantic.html>
3. Bud: IBM's Intranet and Folksonomy. In: The Community Engine. March 2005. http://thecomunityengine.com/home/archives/2005/03/ibms_intranet_a.html
4. Bud: xFolk: An xhtml microformat for folksonomy. In: The Community Engine. March 2005, http://thecomunityengine.com/home/archives/2005/03/xfolk_an_xhtml.html
5. European Commission: IST 2003-2004 Work Programme. http://www.cordis.lu/ist/workprogramme/fp6_workprogramme.htm