

Secure tag-sharing over a messaging network

Jan Torben Heuer jan.heuer@uni-muenster.de

April 1, 2008

1 introduction

For some years now, people tag any kind of *resources* on the web. Examples are: *bookmarks* on sites like delicious¹ or Mr Wong², *photos* on Flickr³ and *publications* on Bibsonomy⁴ and Citeulike⁵. *Tagging* refers to the annotation of resources with one or more keywords. The used vocabulary is up to the creator, making it easier for him to organize and retrieve his information later on. The current situation where people tag their resources (bookmarks, photos, ...) on several sites, has two major drawbacks: First, there is no integration (tag-sharing) between the sites. If you search for tags you will have to first choose a site and then perform your query. Second, and this is becoming more and more important, your private data is stored on a remote server, managed by a potentially not trustworthy third-party. But today, privacy is a growing issue for social networks.

We argue that a server-based social network can't guarantee privacy. Our approach is to use a messaging network based on the open standard xmpp⁶ to provide a network layer for a client-to-client communication. While all the clients are *nodes* in our network, we call two nodes *trusted* when they authorize each other to exchange data. The user keeps his data private by deciding who can query his data or see his tagged resources. That is an advantage over server-based approaches. A query is sent to all trusted nodes and they forward again the query to their trusted nodes. It is a recursive search in the network. Searches are optimized by finding neighborhoods: If other users (nodes) tag similar (meaning same interpretation of vocabulary) their *virtual distance* will be reduced - and vice versa. This is achieved by collecting feedback from the user. (For example if the user also tagged a bookmark he got as searchresult).

¹<http://del.icio.us>

²<http://mrwong.com>

³<http://flickr.com>

⁴<http://bibsonomy.org>

⁵<http://citeulike.org>

⁶<http://www.xmpp.org> and <http://www.ietf.org/rfc/rfc3920.txt>

This thesis will base on tags described by the SCOT ontology⁷. Our approach is independent from the the tagged resource itself (photo, bookmark, publication).And the design of the network communication protocol is independent from the ontology: It bases on SPARQL ⁸ a rdf query language.

Hypothesis A infrastructure based on the xmpp network, the SCOT ontology and SPARQL as query language:

1. is able to keep the users' privacy by controlling who can access their data
2. is efficient although there is no central database. Users can perform a recursive search where trusted nodes forward the search to their trusted nodes.
3. can be optimized by introducing a *virtual distance* between nodes.

2 Challenges and expected results

How can existing information be included? In order to face the typical “cold-start problem” of communities, existing data will be integrated: Java-plugins or APIs for librarything⁹, flickr, delicious and bibsonomy exist and will allow the user to import existing data.

How can a recursive search be implemented? If a user only asks his trusted nodes, there will be few results. A recursive search returns more. Challenges are redundant queries (when trusted nodes of trusted nodes overlap) and termination of recursive searches.

Can the user improve the ranking by rating the results? The user's feedback is used to improve rankings. There is implicit feedback: A user clicks on a link or a user tags the same link with the same vocabulary. Additionally, we can ask the user for explicit feedback (rate the results). The connection between the users' feedback and the virtual distance of nodes is a challenge that will be explored.

⁷<http://scot-project.org/>

⁸<http://www.w3.org/TR/rdf-sparql-XMLres/>

⁹<http://www.librarything.com>