Privacy in the social semantic web: a tag-sharing network

Jan Torben Heuer

Institute for Geoinformatics, University of Muenster
jan.heuer@uni-muenster.de

Social networks are successful “web 2.0” applications. The mixture of collaborative content-generation, interactive user interfaces and the ability of linking with other members in the network attracts a lot of users. Flickr (photos), Delicious (bookmarks) or Bibsonomy (publications) are current examples. All sites have in common that a user can attach keywords to the photo or bookmark he submitts. This step is called *tagging* and 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 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 entity.

At this point we suggest to use our social semantic network infrastructure. We use a friend-to-friend like communication infrastructure basing on the *extensible messaging and presence protocol* xmpp [1]. This protocol allows for communication between clients that know and trust each other. Extra security can be added by signing and encrypting the messages between clients [2]. All content is stored at the client and only the user decides who can access which content. We present a social tagging application where users can annotate their (private) bookmars and share them if they want. Bookmarks are represented by the TagOntology [3] as rdf triples and can be queried over the network using SPARQL [4]. Incoming queries within the network are forwarded to friends (allowing recursive searches). A distance between friends is defined to guarantee termination: Each time a query is sent or forwarded to a friend, the distance to the receiver is added. Searches will terminate if a given maximum distance is reached. When a query has matched, the result will be sent back together with the current distance. Once, all results have been collected, they can be ranked. Two parameters can be taken into account: *How many friends bookmarked a resource?* and *what is its closest distance?* . Besides the privacy aspect, this approach has another advantage over traditional

server-based approaches: Higher precision and recall [5]. We argue that results from closer friends are more relevant because close friends (or co-workers) have more likely the same intended meaning of a vocabulary (tag). Finally, a user's feedback can be used to recalculate his friends' distance: If a user often clicks on the bookmarks provided by one friend or if he also bookmarks the same page with the same tags, the distance to the friend will be reduced and his search results will be ranked higher in future.

Our approach shows how traditional social semantic applications can be build upon a network infrastructure that guarantees privacy. Additionally, the recursive search can even improve the precision of search results.

REFERENCES

[1] Extensible messaging and presence protocol (xmpp): Core. Technical report, IETF proposed standard, RFC 3920, Oct 2004.

[2] Richard Newman, Danny Ayers, and Seth Russell. Tag ontology. http://www.holygoat.co.uk/owl/redwood/0.1/tags/, December 2005.

[3] Ian Paterson. Simplified encrypted session negotiation. Technical Report XEP-0217, XMPP Standards Foundation, May 2007.

[4] Eric Prud'hommeaux and Andy Seaborne. SPARQL query language for RDF. Technical report, World Wide Web Consortium, 2005.

[5] C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 1979.