Project Number 001907

DELIS
Dynamically Evolving, Large-scale Information Systems

Integrated Project
Member of the FET Proactive Initiative Complex Systems

Deliverable D2.4.6

Demonstrator for data management platform using semantically-enhanced social tags
Start date of the project: January 2004
Duration: 50 months
Project Coordinator: Prof. Dr. math. Friedhelm Meyer auf der Heide
Heinz Nixdorf Institute, University of Paderborn, Germany
Due date of deliverable: December 2007
Actual submission date: December 2007
Dissemination level: PU – public

Work Package 1.1: Efficient discovery of network topology and relationships
Due date of deliverable: December 2007
Actual submission date: December 2007
Dissemination level: PU – public

Work Package 2.4: Integrated Application Testbed
Participants: RadioLabs (RAL), Universita di Roma “Tor Vergata”, Italy
Technical University Berlin (TUB), Germany
Authors of deliverable: Giovanni Cortese (g.cortese@computer.org)
Luca Lorenz (llorenz@gmail.com)
Felix Heine (fh@upb.de)
1 Introduction

This report describes the deliverable D2.4.6, "Demonstrator for data management platform using semantically-enhanced social tags". The deliverable is concerned with the development of a software application using the data management platform developed in WP2.3. In addition to the software, we delivered a technical report which provides a detailed description of the application design.

2 Overview of the Application

Social networking sites such as Flickr or Delicious are a very popular, new class of applications. More specifically, social bookmarking sites allow Internet users to store, organize, share and search bookmarks of web pages. In a social bookmarking system, users save links to web pages that they want to remember and/or share. Most social network sites today are implemented using a centralized storage architecture.

Peermarks is an application we developed to support social bookmarking, with the main purpose to study suitability of the peer-to-peer data management technologies we are researching. Peermarks allow users to share bookmarks through an XML and RDF-based, fully distributed datastore. The application allows users to classify web resources bottom-up (using tags) or to use a hierarchical categorization scheme. Peermarks supports both XML data storage and RDF-based inferencing. In Peermarks, users share bookmarks along with their classification system. We suggest the proposed system allows a global classification system to emerge from individual user categories.

Social bookmarking involves saving bookmarks to an online service and “tagging” them with keywords the user creates instead of saving the bookmarks in a browser’s favorites list. The collection of bookmarks can be made accessible to other users who may copy bookmarks to their own collection. Social bookmarking enables to discover other people who are interested in a topic and know about excellent web resources that one may not have found by using a search engine.

Tags can be thought of as keywords that allow ad hoc classification and sorting of a variety of types of information. Tagging, in the context of social bookmarking, is applied to URLs. Other popular web 2.0 sites (e.g. Flickr) apply the tagging approach to photos. A well understood problem with tag-based approaches to categorization of bookmarks (in general, of content) is that no consistent oversight exists to how a resource is tagged—e.g. the tag London could stand for City of London or Jack London. A tag is just a tag—not a concept.

Traditional means of organizing information elements have generally relied on well-defined and pre-declared schemas ranging from simple controlled vocabularies to taxonomies to thesauri to full-blown ontologies. A controlled approach to cataloguing allows for both the validation and quality control of known terms to be registered within an information system. By contrast, tag-based systems create dynamic categorization systems whereby the user annotates links with whatever terms seem most relevant. Links are generally annotated with ‘tags’, which are free-form labels assigned by the user and not drawn from any controlled vocabulary. This is very much a ‘bottom-up’ (or personal) approach compared with the traditional ‘top-down’ (or organizational) structured means of classification [DLIB]. This unstructured (or better, freely structured) approach to classification with users assigning their own labels is variously referred to as a ‘folksonomy’ or ‘social classification’, which is distributed and cooperative in nature.

Tags generally produce a flat namespace, rather than the hierarchical structures that a taxonomy or other formal classification system usually provides. This, of course, has its upsides and downsides. There is considerable debate in the specialized literature on the pros and cons of flat vs. hierarchical classification systems. Attempts are ongoing to introduce structure within tags. For example, tagging of tags (see for an example ‘tag bundles’ in del.icio.us) could help create hierarchical folksonomies. [BlogPind] suggests various possibilities to improve tag-based systems:
• ‘suggest tags for me’
• find synonyms automatically
• help me use the same tags others use
• infer hierarchy from the tags
• make it easy to adjust tags on old content

Currently only the last option appears to be in common use, presumably because it is the easiest to implement.

The application we developed, although our main aim was technological i.e. to test suitability of distributed query and inferencing techniques to a novel application fields, proposes also a fresh approach, which we believe a) helps users organize tags in groups, b) through sharing and collaboration, promotes cooperative, informal creation of taxonomies. The application design is described in the technical report [DELIS-TR-0602].

Peermarks is basically designed to support a user, who searches the Web, to

• organize bookmark information about interesting Web resources for later recall
• discover potentially useful resources already ‘filtered’ by other users in the social network

In order to organize and classify bookmarks and to support later search and recall, a social bookmarking applications must use some categorization technique. In Internet-scale environments, where the classification is forcibly open or very large, it is very difficult or unfeasible to use an approach to categorization of resources based on controlled vocabularies.

Peermarks takes a mixed approach to categorization of resources for later retrieval.

- Web Resources can be bookmarked and associated to search terms in a bottom-up fashion by associating them to tags.

- In addition, the user can associate a web resource to a category from one or several classifications. Informally speaking, a category represents a concept, while a tag is just a name, possibly one of the several which can be related to the concept. Categories are arranged in a tree-like fashion, and retrieval works as one would intuitively expect - when querying resourcing matching a category (e.g., winter sports), all resources associated to the category and its descendants in the tree are returned.

- Last, tags can be related to similar tags by relating them to a category.

As most social bookmarking tools, users can share their bookmarks and import bookmarks from other participants in the application. Peermarks makes a step further, by supporting users in collaborative creation of taxonomical and classification knowledge. This is achieved by allowing users to export their personal classification knowledge, which is merged by Peermarks into the network-wide, global classification knowledge. Users can then query the network classification and move selected parts of the network classification into their personal classification.

3 Contributions to Deliverable

The main contribution to section 2 is a software and is also described in full detail in [DELIS-TR-0602].
References

