Project no. 34721

TAGora

Semiotic Dynamics in Online Social Communities

http://www.tagora-project.eu

Sixth Framework Programme (FP6)

Future and Emerging Technologies of the Information Society Technologies (IST-FET Priority)

D2.5 Final Report on tagging systems update and usage

Period covered: from 01/06/2008 to 31/08/2009	Date of preparation: 15/09/2009
Start date of project: June 1st, 2006	Duration: 39 months
Due date of deliverable: October 15th, 2009	Actual submission date: September 15th, 2009
Distribution: Public	Status: Final

Project coordinator: Vittorio Loreto
Project coordinator organisation name: PHYS-SAPIENZA
Lead contractor for this deliverable: UNIK

Executive Summary

This deliverable provides the final report about the development of the following systems:

BibSonomy – A social resource sharing system for bookmarks and publications

Tagster – Folksonomy Peer-to-Peer System for Sharing Multimedia Data

Ikoru – A Test-bed for Collaborative Tagging and Content-Based Analysis

Zexe.net – A Community Memory for Representing Daily Experiences using Folksonomies

NoiseTube – A Community Memory representing the exposure of people to noise pollution

TAGnet – A tool for awareness and management of personal metadata

MyTag – A tool for integrating folksonomies

Live Social Semantics – A system for supporting real-world social networking at conferences

The aim of these systems in Tagora is twofold: they shall support the collection of real-world user data for experiments, and provide a platform for experimenting how different kinds of user interaction influence the evolution of the resulting data over time.

The first system, *BibSonomy*, allows users collaborative organizing and sharing of bookmark collections and publication lists. A basic version of BibSonomy has been online before the start of the project. Within Tagora, we have extended its functionality to attract a significant number of users, and have provided means for a systematic generation of data for experiments.

Tagster is a system for collaboratively organizing and sharing multimedia data in a peer-to-peer network. It is completely decentralized and provides the same functionalities as common centralized folksonomy systems like Flickr, delicious or BibSonomy. Besides the basic tagging and browsing support it is also used as a test bed for implementing new features like distributed management of metadata statistics. The tagging data gathered in the system will following be used for further analysis and comparison with the dataset from centralized tagging systems.

The *Ikoru* system, developed at Sony CSL, is primarily used to experiment with collaborative tagging and content-based analysis. The project consists of a server-side component and a Web interface, which can be viewed at http://www.ikoru.net.

The Zexe.net initiative focused on finding new ways to use tagging for the benefit of off-line communities facing real-world issues related to accessibility and sustainability. *canal*MOTOBOY* and *GENEVE*accessible* are the latest deployments of the Zexe.net platform. Both of these projects made intensive use of collaborative tagging. In *canal*MOTOBOY*, 15 motorcycle messengers in São Paulo, Brazil, used multimedia mobile phones to capture images and videos of their daily lives, describe the contents using tags and published them on the web. The *GENEVE*accessible* project involved handicapped people in Geneva, Switzerland. They used multimedia phones equipped with GPS receivers to create maps of their city's accessibility. They used tags to describe the images of obstacles they find in their way. By publishing these (geo-)tagged images on the web, they effectively built an intelligent, collaborative map which is immediately available to the public. These projects have enabled members of TAGora to study the dynamics of tagging in small-scale groups with shared interests. In particular, these projects provide two contrasting examples. While *canal*MOTOBOY* was totally open-ended, *GENEVE*accessible* had very specific goals. We have studied how these different scopes affect the projects' folksonomies.

NoiseTube is the natural successor of the Zexe.net platform. In the NoiseTube platform the application of collaborative tagging is extended to new type of resource: the exposure of individual citizens to pollution. NoiseTube also draws on the concept of tagging generation, based on primitive low level-features from the Ikoru platform. Concretely the Zexe.net platform was extended to support the sensing (as in measuring) and tagging of occurrences of noise pollution in cities using the mobiles phones of citizens as sensor and tagging instruments. The project consists of a web portal (which can be visited at http://www.noisetube.net) and an application for mobile phones.

The *TAGnet* system is a prototype (not yet available on the web) designed to provide users with a reflexive tool to expose regularities and patterns in their own tag-based annotations. Tagging patterns can reveal a lot about a user's experience, her interests and her emergent conceptualizations, but users are not aware of these patterns until these regularities are made explicit by means of data analysis and state-of-the-art visualization. TAGnet currently focuses on Flickr users, providing them with a "semantic mirror". It is conceived as a web application that provides users with actionable meaning on their own metadata. In perspective it will be also used as a tool to explore emergent conceptualizations and tag ranking strategies. This application was not initially foreseen, and was set up to exploit the results of WP3 and WP4 on the structure of tag co-occurrence networks.

The *MyTag* system aims at solving the limitations of current tagging platforms by enabling cross-media search across images, video, and social bookmarks. It offers transparent access to different single-media platforms currently including Flickr, YouTube, and del.icio.us. The search function can be personalized in two directions. First, MyTag users can restrict the search to the resources uploaded by the user him/herself. Second, the website uses an implicit user feedback mechanism to personalize the output of a query to MyTag by ranking results according to the user's personomy. The personomy is built without additional effort by the previous queries entered by the user, in contrast with other tagging platforms, such as Flickr or del.icio.us, where an explicit feedback is required in order to personalize the ranking of the results presented to the user.

Live Social Semantics is a platform for gathering and integrating data from social networking sites, the semantic web, and RFID devices. Live Social Semantics provides a suite of services for conference attendees, to facilitate discovering people will similar interests, to recommend talks, to find friends, and to race and log their real-world networking activities (face-to-face contacts). The system was successfully deployed at two international conferences, and acted as a great demonstrator of various TAGora technologies and products. Because of the success of Live Social Semantics so far, and the value it provides to conference attendees, we have been asked to rerun it at ESWC 2010, as well as at other events, such as EU ICT 2010, and WWW 2012.

Contents

List of Figures

Chapter 1

BibSonomy – A Social Resource Sharing System for Bookmarks and Publications

This chapter describes the extended functionality of BibSonomy that we have implemented during the project in order to increase user activities. This strategy has been successful, as indicated by Figure 1.1, which shows the increase of the number of users. New features are announced on a weekly basis on http://bibsonomy.blogspot.com/.

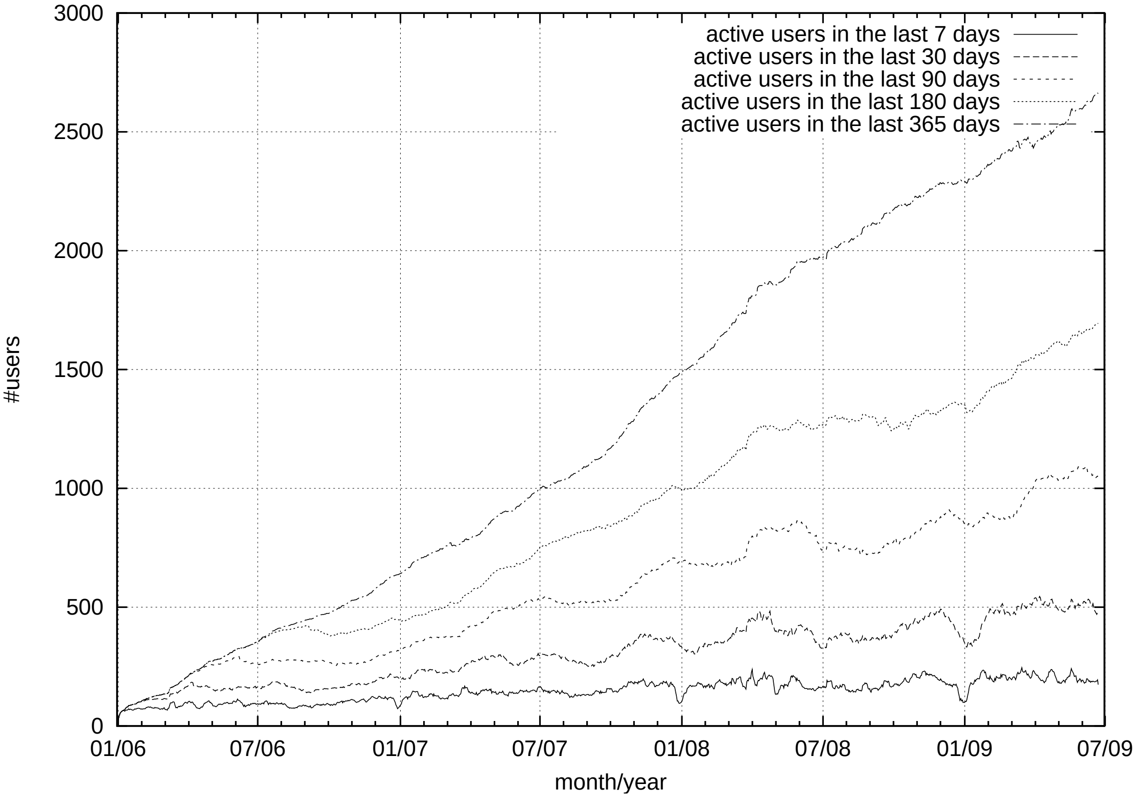


Figure 1.1: Growth of BibSonomy

1.1 Supporting community building in BibSonomy

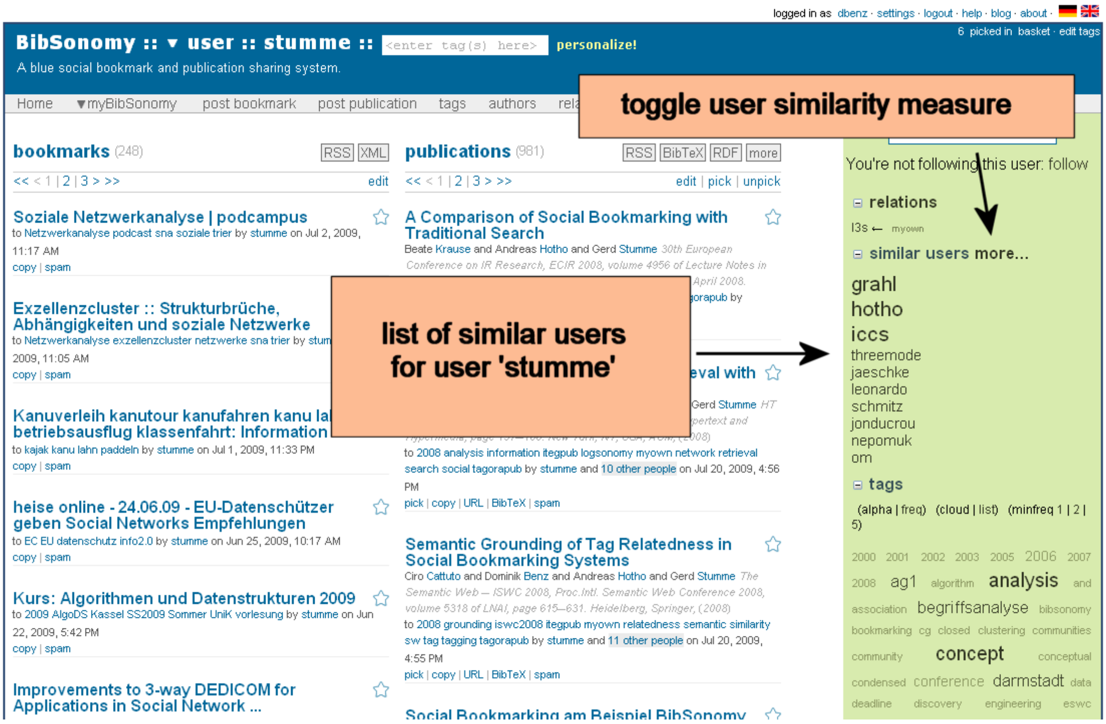
1.1.1 Explicitly defined communities/groups

BibSonomy contains more than 100 groups, which actively sharing and collecting resources. The groups are mostly research groups or participants of European projects. To facilitate the idea of group working, the system offers a common tag cloud for each group, which is characterized by a filter to restrict the view on the tags. In this way the sharing of resources is restricted to a group. To become a group member, the users have to apply for the group of interest. The group administrator will be informed by email and can accept or reject the request. After a user has joined a group, he can see all entries of this group (i. e., also the group internal resources).

1.1.2 Following similar users by personalized user pages

Apart from the possibility to define explicit groups, BibSonomy has a built-in mechanism to support users in keeping track of interesting content provided by "similar users". A similar user in our sense is hereby a user with similar interests, e.g. in similar research topics like the target user. This support mainly comprises three components: (i) Computation of similar users for a target user, (ii) personalized ranking of the resources provided by these similar users for the target user, and (iii) possibility to establish an explicit "follower" link between the target user and a similar user. Most of this work has been designed and implemented during the visit of Vito Servedio from the University "La Sapienzia", Rome, in Kassel.

Computing similar users: We are running a daily update job, which computes similar users for each active BibSonomy user. We consider a user to be "active" when he has at least 50 tag assignments, which corresponds to roughly 15 tagged resources in average. For each active user, this mechanism computes the most similar users based on several similarty measures:


toggle user similarity measure
list of similar users for user 'stumme'

logged in as dbenz · settings · logout · help · blog · about ·

6 picked in basket · edit tags

BibSonomy :: ▾ user :: stumme :: <enter tag(s) here> personalize!

A blue social bookmark and publication sharing system.

Home ▾myBibSonomy post bookmark post publication tags authors rela

bookmarks (248) RSS XML

<< < 1 | 2 | 3 > >> edit

Soziale Netzwerkanalyse | podcampus
to Netzwerkanalyse podcast sna soziale trier by stumme on Jul 2, 2009, 11:17 AM
copy | spam

Exzellenzcluster :: Strukturbrüche, Abhängigkeiten und soziale Netzwerke
to Netzwerkanalyse exzellenzcluster netzwerke sna trier by stum
2009, 11:05 AM
copy | spam

Kanuverleih kanutour kanufahren kanu la
betriebsausflug klassenfahrt: Information
to kajak kanu lahn paddeln by stumme on Jul 1, 2009, 11:33 PM
copy | spam

heise online - 24.06.09 - EU-Datenschützer geben Social Networks Empfehlungen
to EC EU datenschutz info2.0 by stumme on Jun 25, 2009, 10:17 AM
copy | spam

Kurs: Algorithmen und Datenstrukturen 2009
to 2009 AlgoDS Kassel SS2009 Sommer UniK vorlesung by stumme on Jun 22, 2009, 5:42 PM
copy | spam

Improvements to 3-way DEDICOM for Applications in Social Network ...

publications (981) RSS BibTeX RDF more

<< < 1 | 2 | 3 > >> edit | pick | unpick

A Comparison of Social Bookmarking with Traditional Search
Beate Krause and Andreas Hotho and Gerd Stumme *30th European Conference on IR Research, ECIR 2008, volume 4956 of Lecture Notes in*
April 2008.
orapub by

eval with

Gerd Stumme *HT*
pertext and
008)
to 2008 analysis information itegpub logsonomy myown network retrieval search social tagorapub by stumme and 10 other people on Jul 20, 2009, 4:56 PM
pick | copy | URL | BibTeX | spam

Semantic Grounding of Tag Relatedness in Social Bookmarking Systems
Ciro Cattuto and Dominik Benz and Andreas Hotho and Gerd Stumme *The Semantic Web — ISWC 2008, Proc.Intl. Semantic Web Conference 2008, volume 5318 of LNAI, page 615—631. Heidelberg, Springer, (2008)*
to 2008 grounding iswc2008 itegpub myown relatedness semantic similarity sw tag tagging tagorapub by stumme and 11 other people on Jul 20, 2009, 4:55 PM
pick | copy | URL | BibTeX | spam

Social Bookmarking am Beispiel BibSonomy

You're not following this user: follow

⊟ relations

l3s ← myown

⊟ similar users more...

grahl
hotho
iccs
threemode
jaeschke
leonardo
schmitz
jonducrou
nepomuk
om

⊟ tags

(alpha | freq) (cloud | list) (minfreq 1 | 2 | 5)

2000 2001 2002 2003 2005 2006 2007 2008 ag1 algorithm analysis and association begriffsanalyse bibsonomy bookmarking cg closed clustering communities community concept conceptual condensed conference darmstadt data deadline discovery engineering eswc

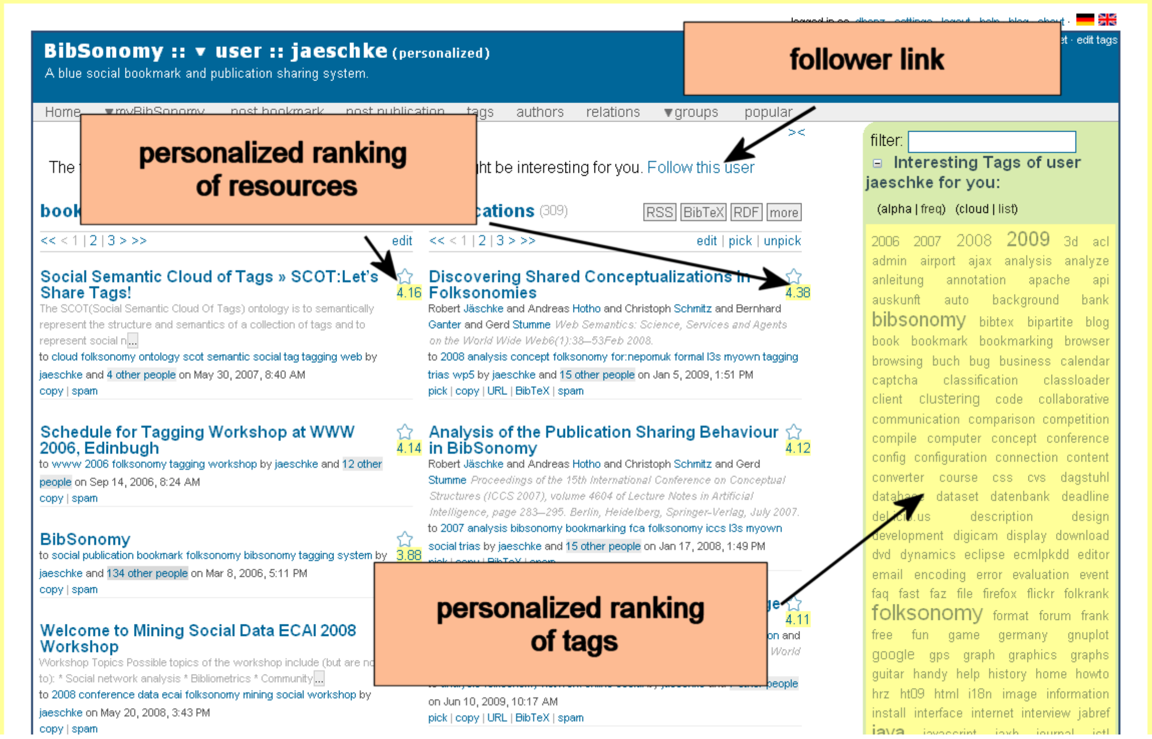
Figure 1.2: Recommendation of similar users on a user page in BibSonomy

- *FolkRank*: As described in the next section, *FolkRank* is a graph-based ranking algorithm for folksonomies (Hotho et al., 2006). With appropriate starting parameters, it is able to rank all folksonomy users with respect to their “relevancy” to a target user. So the “FolkRank similarity” is the first measure we use to compute similar users.

- *Distributional measures*: Another standard approach is to represent users by their tag clouds, or more precisely by feature vectors whose dimensions correspond to the tags used by a certain user. Within this vector space, one can apply standard measures from Information Retrieval to compute user similarity. We implemented the Cosine, Jaccard, and a TF/IDF-like similarity measure.

On a user page of a given user (let’s say *stumme*) within BibSonomy, we display the 10 most similar users. We allow the user to toggle between the different measures to identify which measure yields the most interesting users for him. Figure 1.2 shows an example.

Personalized ranking of resources: After having identified similar users, the target user can inspect the resources of the recommended users on their personalized user pages. The basic idea is to display the most relevant posts for the target user on top; we have adopted the straightforward approach that a post is relevant if it is tagged with tags which are important (i.e., frequently used) by the target user. Figure 1.3 displays such a personalized page.



BibSonomy :: ▾ user :: jaeschke (personalized)
A blue social bookmark and publication sharing system.
edit tags
Home
▾myBibSonomy
post bookmark
post publication
tags
authors
relations
▾groups
popular
>&lt;
RSS
BibTeX
RDF
more
filter:
(alpha | freq) (cloud | list)

follower link

personalized ranking of resources

personalized ranking of tags

The … ht be interesting for you. Follow this user

book…

<< < 1 | 2 | 3 > >> edit

Social Semantic Cloud of Tags » SCOT:Let’s Share Tags! 4.16
The SCOT(Social Semantic Cloud Of Tags) ontology is to semantically represent the structure and semantics of a collection of tags and to represent social n...
to cloud folksonomy ontology scot semantic social tag tagging web by jaeschke and 4 other people on May 30, 2007, 8:40 AM
copy | spam

Schedule for Tagging Workshop at WWW 2006, Edinbugh 4.14
to www 2006 folksonomy tagging workshop by jaeschke and 12 other people on Sep 14, 2006, 8:24 AM
copy | spam

BibSonomy 3.88
to social publication bookmark folksonomy bibsonomy tagging system by jaeschke and 134 other people on Mar 8, 2006, 5:11 PM
copy | spam

Welcome to Mining Social Data ECAI 2008 Workshop
Workshop Topics Possible topics of the workshop include (but are no… to): * Social network analysis * Bibliometrics * Community...
to 2008 conference data ecai folksonomy mining social workshop by jaeschke on May 20, 2008, 3:43 PM
copy | spam

…cations (309)

<< < 1 | 2 | 3 > >> edit | pick | unpick

Discovering Shared Conceptualizations in Folksonomies 4.38
Robert Jäschke and Andreas Hotho and Christoph Schmitz and Bernhard Ganter and Gerd Stumme *Web Semantics: Science, Services and Agents on the World Wide Web6(1):38—53Feb 2008.*
to 2008 analysis concept folksonomy for:nepomuk formal l3s myown tagging trias wp5 by jaeschke and 15 other people on Jan 5, 2009, 1:51 PM
pick | copy | URL | BibTeX | spam

Analysis of the Publication Sharing Behaviour in BibSonomy 4.12
Robert Jäschke and Andreas Hotho and Christoph Schmitz and Gerd Stumme *Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007), volume 4604 of Lecture Notes in Artificial Intelligence, page 283—295. Berlin, Heidelberg, Springer-Verlag, July 2007.*
to 2007 analysis bibsonomy bookmarking fca folksonomy iccs l3s myown social trias by jaeschke and 15 other people on Jan 17, 2008, 1:49 PM

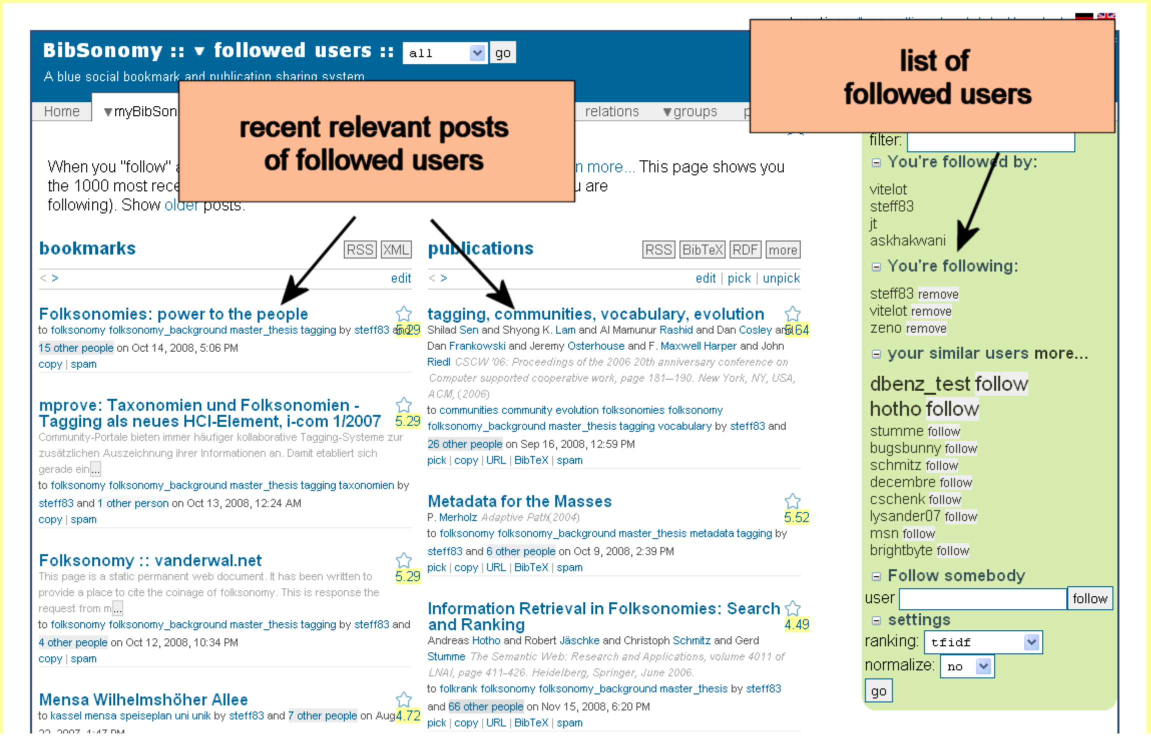
…ge 4.11
…on and
…World
…people
on Jun 10, 2009, 10:17 AM
pick | copy | URL | BibTeX | spam

Interesting Tags of user jaeschke for you:

2006 2007 2008 2009 3d acl admin airport ajax analysis analyze anleitung annotation apache api auskunft auto background bank bibsonomy bibtex bipartite blog book bookmark bookmarking browser browsing buch bug business calendar captcha classification classloader client clustering code collaborative communication comparison competition compile computer concept conference config configuration connection content converter course css cvs dagstuhl database dataset datenbank deadline del.icio.us description design development digicam display download dvd dynamics eclipse ecmlpkdd editor email encoding error evaluation event faq fast faz file firefox flickr folkrank folksonomy format forum frank free fun game germany gnuplot google gps graph graphics graphs guitar handy help history home howto hrz ht09 html i18n image information install interface internet interview jabref java javascript jaxb journal jctl

Figure 1.3: Personalized user page in BibSonomy

Creating an explicit “follower” link: If our recommendation was successful, the target user is interested in the resources of the recommended similar user(s). In such a case, the target user can “follow” one or more similar users by creating an explicit “followers”-link (see Figure 1.3). We provide an additional page which summarizes all recent posts of all users followed by the target user, once again ranked personally. By subscribing e.g. to the RSS feed of this *followers page*, the target user can easily keep track of interesting content added by other users with similar interests. See Figure 1.4 for an impression of the followers page.



BibSonomy :: ▾ followed users :: all go

A blue social bookmark and publication sharing system

Home ▾myBibSon relations ▾groups

When you "follow" a n more... This page shows you the 1000 most rece u are following). Show older posts.

bookmarks RSS XML

< > edit

Folksonomies: power to the people 5.29
to folksonomy folksonomy_background master_thesis tagging by steff83 and 15 other people on Oct 14, 2008, 5:06 PM
copy | spam

mprove: Taxonomien und Folksonomien - Tagging als neues HCI-Element, i-com 1/2007 5.29
Community-Portale bieten immer häufiger kollaborative Tagging-Systeme zur zusätzlichen Auszeichnung ihrer Informationen an. Damit etabliert sich gerade ein...
to folksonomy folksonomy_background master_thesis tagging taxonomien by steff83 and 1 other person on Oct 13, 2008, 12:24 AM
copy | spam

Folksonomy :: vanderwal.net 5.29
This page is a static permanent web document. It has been written to provide a place to cite the coinage of folksonomy. This is response the request from m...
to folksonomy folksonomy_background master_thesis tagging by steff83 and 4 other people on Oct 12, 2008, 10:34 PM
copy | spam

Mensa Wilhelmshöher Allee 4.72
to kassel mensa speiseplan uni unik by steff83 and 7 other people on Aug 22, 2007, 1:47 PM

publications RSS BibTeX RDF more

< > edit | pick | unpick

tagging, communities, vocabulary, evolution 5.64
Shilad Sen and Shyong K. Lam and Al Mamunur Rashid and Dan Cosley and Dan Frankowski and Jeremy Osterhouse and F. Maxwell Harper and John Riedl *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, page 181—190. New York, NY, USA, ACM, (2006)*
to communities community evolution folksonomies folksonomy folksonomy_background master_thesis tagging vocabulary by steff83 and 26 other people on Sep 16, 2008, 12:59 PM
pick | copy | URL | BibTeX | spam

Metadata for the Masses 5.52
P. Merholz *Adaptive Path(2004)*
to folksonomy folksonomy_background master_thesis metadata tagging by steff83 and 6 other people on Oct 9, 2008, 2:39 PM
pick | copy | URL | BibTeX | spam

Information Retrieval in Folksonomies: Search and Ranking 4.49
Andreas Hotho and Robert Jäschke and Christoph Schmitz and Gerd Stumme *The Semantic Web: Research and Applications, volume 4011 of LNAI, page 411-426. Heidelberg, Springer, June 2006.*
to folkrank folksonomy folksonomy_background master_thesis by steff83 and 66 other people on Nov 15, 2008, 6:20 PM
pick | copy | URL | BibTeX | spam

filter:

⊟ **You're followed by:**
vitelot
steff83
jt
askhakwani

⊟ **You're following:**
steff83 remove
vitelot remove
zeno remove

⊟ **your similar users more...**
dbenz_test follow
hotho follow
stumme follow
bugsbunny follow
schmitz follow
decembre follow
cschenk follow
lysander07 follow
msn follow
brightbyte follow

⊟ **Follow somebody**
user follow

⊟ **settings**
ranking: tfidf
normalize: no
go

Figure 1.4: The “followers page” of BibSonomy summarizes recent relevant posts from followed users.

1.1.3 Searching enhances implicit community building.

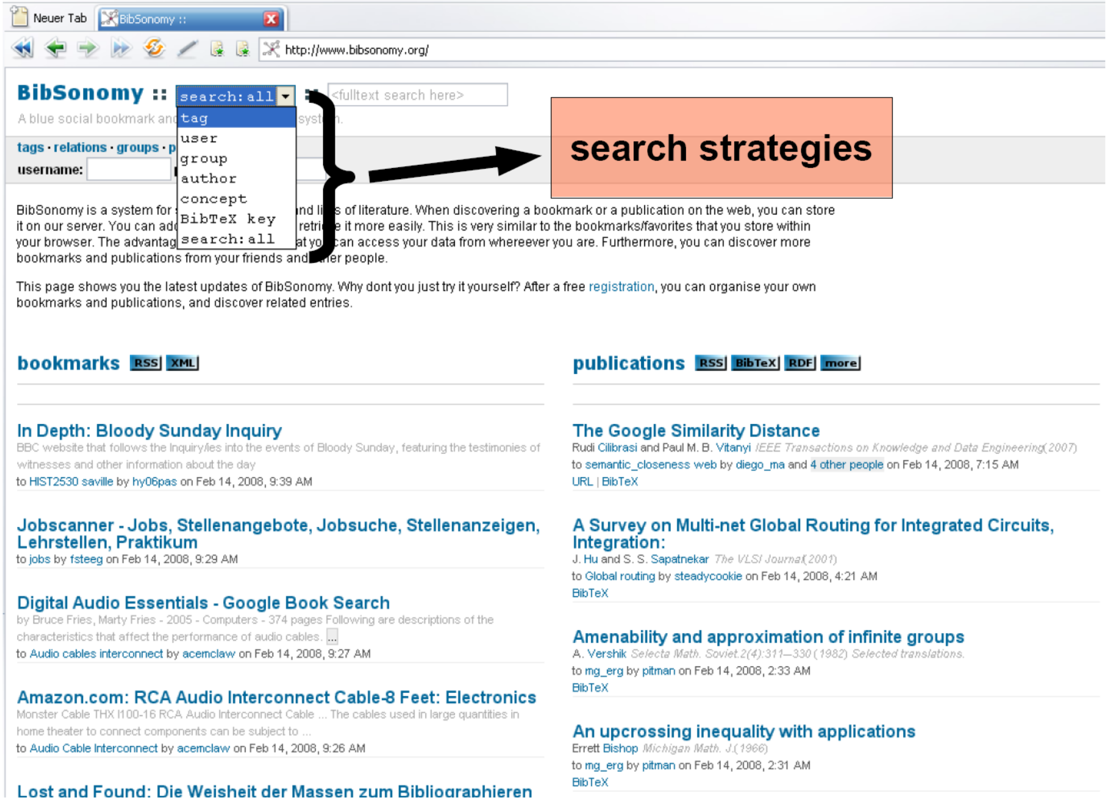
To find content from users who used similar tags as oneself, BibSonomy supports more user flexibility regarding different search mechanisms. The following search strategies were implemented after 1st of June 2007 and are available online for navigation (see Fig. 1.5).

Searching by author BibSonomy allows users to search in publications via author names. The author search is implemented based on the MySQL full text search feature of the MyISAM database engine. The system copies all the author information of a publication into a text field of a MyISAM table, yielding a very fast search functionality.

The simplest way to search for an author is to search for the last name. That is called an ‘author page’. The author search results in a list with all publications together with a tag cloud describing the topics of the author based on the tags that are attached to his publications in the system.

Searching by concepts BibSonomy allows users to structure the content via *SUPERTAG <- SUBTAG* relations (see Fig. 1.6). When ‘concept’ is selected in the pull-down menu in Fig. 1.5, then the search does not result only in the resources that are annotated with the given tag, but returns also all resources which are annotated by at least one subtag.

Ranking Another major extension contained in BibSonomy is a mechanism to rank resources and users for a given tag by relevance. To this end, we have implemented the *FolkRank* algorithm (Hotho et al., 2006). Its idea is similar to Google’s PageRank algorithm,



A blue social bookmark and syst n.

BibSonomy is a system for s nd li s of literature. When discovering a bookmark or a publication on the web, you can store it on our server. You can add retri e it more easily. This is very similar to the bookmarks/favorites that you store within your browser. The advantag at yo can access your data from whereever you are. Furthermore, you can discover more bookmarks and publications from your friends and er people.

This page shows you the latest updates of BibSonomy. Why dont you just try it yourself? After a free registration, you can organise your own bookmarks and publications, and discover related entries.

bookmarks

In Depth: Bloody Sunday Inquiry
BBC website that follows the Inquiry/ies into the events of Bloody Sunday, featuring the testimonies of witnesses and other information about the day
to HIST2530 saville by hy06pas on Feb 14, 2008, 9:39 AM

Jobscanner - Jobs, Stellenangebote, Jobsuche, Stellenanzeigen, Lehrstellen, Praktikum
to jobs by fsteeg on Feb 14, 2008, 9:29 AM

Digital Audio Essentials - Google Book Search
by Bruce Fries, Marty Fries - 2005 - Computers - 374 pages Following are descriptions of the characteristics that affect the performance of audio cables. ...
to Audio cables interconnect by acemclaw on Feb 14, 2008, 9:27 AM

Amazon.com: RCA Audio Interconnect Cable-8 Feet: Electronics
Monster Cable THX I100-16 RCA Audio Interconnect Cable ... The cables used in large quantities in home theater to connect components can be subject to ...
to Audio Cable Interconnect by acemclaw on Feb 14, 2008, 9:26 AM

Lost and Found: Die Weisheit der Massen zum Bibliographieren

publications

The Google Similarity Distance
Rudi Cilibrasi and Paul M. B. Vitanyi *IEEE Transactions on Knowledge and Data Engineering*(2007)
to semantic_closeness web by diego_ma and 4 other people on Feb 14, 2008, 7:15 AM
URL | BibTeX

A Survey on Multi-net Global Routing for Integrated Circuits, Integration:
J. Hu and S. S. Sapatnekar *The VLSI Journal*(2001)
to Global routing by steadycookie on Feb 14, 2008, 4:21 AM
BibTeX

Amenability and approximation of infinite groups
A. Vershik *Selecta Math. Soviet.2(4):311—330 (1982) Selected translations.*
to mg_erg by pitman on Feb 14, 2008, 2:33 AM
BibTeX

An upcrossing inequality with applications
Errett Bishop *Michigan Math. J.*(1966)
to mg_erg by pitman on Feb 14, 2008, 2:31 AM
BibTeX

Figure 1.5: Search strategies in Bibsonomy

tagging ← folksonomy tagcloud

Figure 1.6: *SUPERTAG* <- *SUBTAG* relation in Bibsonomy.

i.e. it analyzes the link structure between users, tags and resources in order to calculate the relevance (more details in Deliverable 3.2).

1.2 Integration with 3rd party products

We made several steps to combine BibSonomy with third party products.

CiteSmart MireSoft has a product called CiteSmart, which is a citation software. It nicely integrates with Word and builds a bridge between web-based tools like BibSonomy or Connotea. It is easing the way to takeover the data from web application and it is able to produces references in various formats for articles written in Word. In this way, it allows to easily write scientific articles and supports the scientific work of researchers. We support this partnership both to broaden our community and to make BibSonomy more valuable for its users.

Zope Zope[1] is an open source application server for building content management systems, intranets, portals, and custom applications. Publication lists, link lists, and tag clouds can be dynamically integrated into Zope web pages, using the *KebasData* product (Zope, 2002).

[1] http://www.zope.org/

Typo3 Extension Typo3 is a popular open-source content management system, used by a large number of private and corporate websites. It offers among others a generic extension architecture, which enables developers to add custom functionality to Typo3-based websites. For many websites in academic contexts (e.g. personal homepages of researchers, universities, research projects, ...), an important building block is an up-to-date publication list. Maintaining these lists manually is a tedious task. The core concept of our Typo3 extension is to keep all references cleanly stored inside BibSonomy (leveraging all useful BibSonomy features like import from different formats, scraping services, ...) and to generate automatically a publication list from this data.

Wiki -and Webblog Software BibSonomy also supports Wiki -and Webblog Software. To integrate BibSonomy data into an XWiki-Page, one has to install the XWiki RSS Feed Plugin (XWi, 2004). Then, the data can be imported as RSS Feed. To import BibSonomy data to a WordPress blog, the WordPressBlog plug-in [2] has been implemented.

WordPress Bloggers who are using WordPress can integrate data from BibSonomy into their posts - for instance your tag cloud, or your last three publications (or all of them). Conversely, your blog posts will (almost) automatically be published on BibSonomy. A more general way of including BibSonomy content into your system is BibSonomy's JSON feed. JSON (JavaScript Object Notation) is a lightweight data-interchange format, which is now available for all BibSonomy pages.

Digital Libraries meet BibSonomy The Library of the University of Cologne [3] was the first 3rd-party organization that incorporated BibSonomy's services: When searching for books and articles, the results can be stored with one mouse click in a personal bibliography collection at BibSonomy.

The Library of the Institute of Information Sciences at the Saarland University, Saarbruecken [4] also integrated BibSonomy into their literature research interface. In addition to the features provided by the KUG library (i.e. the direct posting of search results) links are provided to retrieve further articles from BibSonomy by author name.

The Library of the University of Heidelberg[5] was following this service, and a similar implementation at the University of Kassel[6] is coming up.

Moodle Moodle is a popular e-learning platform for students and lectures (Dougiamas, 1999). BibSonomy can be integrated to enhance course descriptions and e-learning projects by providing the corresponding literature also via RSS-feeds.[7]

GoogleSonomy GoogleSonomy is a new firefox addon which integrates search results from BibSonomy directly in your Google search. The addon is customizable so that you can decide whether to search in your personal publications and/or bookmarks, or to search over all BibSonomy posts.The extension is available from the Mozilla Addon Page.

Zotero BibSonomy now also allows to export citation information to Zotero. Zotero is a Firefox extension, which helps you to collect, manage and cite publications locally in your browser. The other way around is not fully automized yet. However, there is a copy and paste workaround.

[2] http://www.christianschenk.org/projects/wordpress-bibsonomy-plugin/
[3] http://kug.ub.uni-koeln.de/
[4] http://is.uni-sb.de/vibi/suchen.html
[5] http://katalog.ub.uni-heidelberg.de
[6] http://opac.bibliothek.uni-kassel.de
[7] http://educampus.uni-kassel.de/

OpenID As alternative login procedure, BibSonomy now also supports OpenID, which is an open, decentralized standard, allowing users to log onto many different services on the web using the same identity identification (single sign-on). This kind of authentication is provided by a growing number of websites, including large ones like AOL, Google, Microsoft, MySpace, Yahoo and many others. You may even have an OpenID without knowing so, e.g. when you have a Flickr account. Why not using it for logging in to BibSonomy as well?

Scrapers The family of scrapers for automatically extracting references from digital libraries or publishers' websites has been extended, allowing you to store publication metadata automatically from over 60 sites. The scraping service can be used independently from BibSonomy for other purposes by everyone needing access to bibliographic metadata.

1.3 BibSonomy as Web Service

BibSonomy's REST API BibSonomy now has an application programming interface (API) which allows external applications to interact with BibSonomy. It is restful API which provides a simple access to all data of BibSonomy. One example application which uses the API to access BibSonomies data is the stand alone BibTeX Manager JabRef[8]. It is an open source tool to manage bibliographic metadata. We have extended the tool by using the client part of API implementation to establish a connection to BibSonomy. The client is able to store and to retrieve references from BibSonomy. The user tag cloud or the tag cloud of the system is shown after the start and offers an intuitive browsing interface. Tags can be used in the usual way to search for references directly in Jabref, and the retrieved reference list can be imported into Jabref's internal library with one click. This also means that references can be used on a laptop without having network connectivity, and will be synchronized with BibSonomy once reconnected.

For the more technically minded: the API is based on REST.[9] To get all users, the data are requested as "GET /users" over HTTP. To modify a particular user, the expression "PUT /users/<username>" is used, together with an appropriately formatted XML document containing the user data. The API documentation is available on http://www.bibsonomy.org/help/doc/api.html.

It is possible to code an application against BibSonomy in about any programming language, although you will have to write all the HTTP and XML wrangling yourself. For the Java language, BibSonomy is also offering a client library (available on http://www.bibsonomy.org/help/doc/javaclient.html). We expect that the community will setup client implementations for other programming languages like, python, perl or php, too.

The discussed functionality of BibSonomy as Web Service works within the bounds of proper authorization. To be able to use the API, interested users can obtain an API key by checking the setting page within its user account.

Multi-language support As researchers and students of different nationalities work with BibSonomy, version 2.0 supports multilingual pages. Almost all non-posting pages of BibSonomy are now available in English and German.

1.4 Recommendation Framework

Allowing freely typed tags leads to problems which all tagging systems face. Golder et. al. [1] identified three major problems as polysemy, synonymy, and level variation. Polysemy refers to tags

[8] http://jabref.sourceforge.net/

[9] http://en.wikipedia.org/wiki/Representational_State_Transfer

which have several meanings (e.g. turkey), synonymy to situations, where several tags share a common meaning (above all morphological variations as 'web20', 'web2.0',...) and level variation to situations, where different people are using tags from different levels of abstraction (e.g. 'computer' vs. 'commodore 64'). One major approach for dealing with such problems evolved recently: Whenever a user wants to tag a resource, a set of (supposedly) appropriate tags is displayed and can be easily inherited for tagging. Beside consolidating a tagging system's vocabulary, these so called 'recommender systems' lower the cognitive effort needed for tagging a resource and thus advocate broader tagging of resources.

Different techniques for recommender systems were developed and evaluated, mainly by comparing the set of tags which a user assigned to a resource with those, which were suggested by the tag recommender (see Deliverable 4.6). Most evaluations used a dataset of a tagging system (e.g. all tagged resources up to a given date), by iterating over all entries and presenting only the resources to the recommender system (spitting the data set for training and testing if necessary). These 'off-line' settings not only ignore some constraints in real live applications (e.g. cpu usage and memory consumption), they also cannot take into account the effect of presenting a set of recommended tags to the user. To evaluate these effects, a recommender system must be integrated into a real live application.

We designed a framework, which allows integration and evaluation of different recommender systems into BibSonomy. These recommender systems can be either installed locally or remotely (connected and queried via http), thus allowing other research teams to integrate their recommender systems and giving a broad base for evaluation.

The framework's central component is a multiplexer where each tag recommender system is registered during initialization. Whenever a user wants to assign tags to a resource, each recommender is queried for recommended tags in parallel, spawning separate threads for each recommender query. All responses are collected and exactly one recommender's result is uniform randomly chosen and presented to the user. Finally, to allow machine learning techniques, we pass to each recommender the set of tags which the user assigned to the resource.

All incoming events and informations (including the information, when and where the user clicked on a button, tag or link) are logged for evaluation in a SQL database. For logging mouse events, we reused BibSonomy's existent click-log facility (see details below). Logging of recommender queries is done in the multiplexer: For each recommendation event we log date, user name, references to all recommender systems as well as all corresponding recommended tags. Finally, when the user submits the readily tagged post to BibSonomy, we store a reference from the recommendation event to the final post in BibSonomy's database. For evaluating time constraints, we also log processing time for each recommendation event and each recommender system.

We also capture those situations, where the user is unsatisfied with a set of recommended tags: A 'reload' button is displayed, which replaces the set of recommended tags with a different recommender's suggestion. This encourages the user to give us the desired feedback, as new tags are presented with low cognitive effort.

To minimize the effort needed for integration of remotely installed foreign recommender systems, we designed a small server application. This server is deployed on the remote machine and handles recommendation requests, passing them the recommender system and returning its result set to BibSonomy.

Because of the limited budget of Tagora, only the design of the recommendation framework could be done within the project. Its implementation was done during the extended life time of the project – and used in the ECML PKDD discovery challenge 2009[10] for dissemination of Tagora (see below) – but financed by a national follow-up project.

[10] http://www.kde.cs.uni-kassel.de/ws/dc09/online

1.4.1 ECML PKDD 2009 Discovery Challenge

Since 1999 the ECML PKDD embraces the tradition of organizing a Discovery Challenge, allowing researchers to develop and test algorithms for novel and real world datasets. This year's Discovery Challenge[11] presents a dataset from the field of social bookmarking to deal with the recommendation of tags. The results submitted by the challenge's participants are presented at an ECML PKDD workshop on September 7th, 2009, in Bled, Slovenia.

The provided dataset has been created using data of BibSonomy. The training data was released on March 25th 2009, the test data on July 6th. The participants had time until July 8th to submit their results. This gave researchers 14 weeks time to tune their algorithms on a snapshot of a real world folksonomy dataset and 48 hours to compute results on the test data.

To support the user during the tagging process and to facilitate the tagging, BibSonomy includes a tag recommender. When a user finds an interesting web page (or publication) and posts it to BibSonomy, the system offers up to five recommended tags on the posting page. The goal of the challenge is to learn a model which effectively predicts the keywords a user has in mind when describing a web page (or publication). We divided the problem into three tasks, each of which focusing on a certain aspect. All three tasks get the same dataset for training. It is a snapshot of BibSonomy until December 31st 2008. The dataset is cleaned and consists of two parts, the core part and the complete snapshot. The test dataset is different for each task.

Task 1: Content-Based Tag Recommendations. The test data for this task contains posts, whose user, resource or tags are not contained in the post-core at level 2 of the training data. Thus, methods which can't produce tag recommendations for new resources or are unable to suggest new tags very probably won't produce good results here.

Task 2: Graph-Based Recommendations. This task is especially intended for methods relying on the graph structure of the training data only. The user, resource, and tags of each post in the test data are all contained in the training data's post-core at level 2.

Task 3: Online Tag Recommendations. This is a bonus task which took place after Tasks 1 and 2. The participants had to implement a recommendation service which can be called via HTTP by BibSonomy's recommender infrastructure when a user posts a bookmark or publication. All participating recommenders are called on each posting process, one of them is chosen to actually deliver the results to the user. We can then measure the performance of the recommenders in an online setting, where timeouts are important and where we can measure which tags the user has clicked on.

Results. More than 150 participants registered for the mailing list which enabled them to download the dataset. At the end, we received 42 submissions – 21 for each of the Tasks 1 & 2. Additionally, 24 participants submitted a paper that can be found in the proceedings.

We used the F1-Measure common in Information Retrieval to evaluate the submitted recommendations. Therefore, we first computed for each post in the test data precision and recall by comparing the first five recommended tags against the tags the user has originally assigned to this post. Then we averaged precision and recall over all posts in the test data and used the resulting precision and recall to compute the F1-Measure as $\mathrm{f1m} = \frac{2 \cdot \mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$.

The winning team of Task 1 has an f1m of 0.18740, the second and third follow with 0.18001 and 0.17975. For Task 2, the winner achieved an f1m of 0.35594, followed by 0.33185 and 0.32461. The winner of Task 3 will be announced at the conference and later on the website of the challenge.

[11] http://www.kde.cs.uni-kassel.de/ws/dc09/

Lipczak et al. from Dalhousie University, Halifax, Canada are the winners of Task 1. With a method based on the combination of tags from the resource's title, tags assigned to the resource by other users and tags in the user's profile they reached an f1m of 0.18740 in Task 1 and additionally achieved the third place in Task 2 with an f1m of 0.32461. The system is composed of six recommenders and the basic idea is to augment the tags from the title by related tags extracted from two tag-tag–co-occurrence graphs and from the user's profile and then rescore and merge them.

The winners of Task 2, Rendle and Schmidt-Thieme from the University of Hildesheim, Germany, achieved an f1m of 0.35594 with a statistical method based on factor models. Therefore, they factorize the folksonomy structure to find latent interactions between users, resources and tags. Using a variant of the stochastic gradient descent algorithm the authors optimize an adaptation of the Bayesian Personal Ranking criterion. Finally, they estimate how many tags to recommend to further improve precision.

The second team of Task 1, Mrosek et al, harvests tags from sources like Delicious, Google Scholar, and CiteULike. They also employ the full-text of web pages and PDFs. The third team, Ju and Hwang, merges tags which have been earlier assigned to the resource or used by the user as well as resource descriptions by a weighting scheme. Finally, the second team of Task 2, Balby Marinho et al, uses relational classification methods in a semi-supervised learning scenario to recommend tags.

1.5 Spam Framework

BibSonomy's spam framework has been designed to automatically detect spam posts, and prevent those posts from being viewed by legal users or crawled by search engines. The framework has been developed using the insights we obtained from our first experiments with the BibSonomy Spam dataset (Krause et al., 2008).

On a regular basis (which can be defined by the system administrator), the system selects the classified users from the last months (currently from the last six months). We do not use the entire dataset as spamming behaviour changes over time. The system computes different user features (location, activity and profile based) to create a training dataset. A model is learned from this dataset. The machine learning algorithms to create the models are provided by Weka, an open source data mining software from the University of Waikato. The model is serialized and stored. It can be reused for the classification task. The administrator can select the classification algorithm the model is built on. As our experiments have shown, models tend to have a high false positive rate (i.e., non-spammers are classified as spammers). In a practical system, it might prevent legal users from continuing working with BibSonomy. The administrator can therefore introduce costs to penalize wrong classifications of non-spammers.

The classification task starts every three minutes. New users with at least one public post are collected from the database. A white-list checks, if the user's mail addresses or the IP addresses come from a well known university. Those users are directly marked as non-spammers. The remaining users are classified, using a model of the training phase. Four different categories for classification are available. The classifier predicts spam and non-spam with a certain confidence. Thus, we can create a category of a spam class, of an unsure spam class, of an unsure non-spam class and a non spam class. The division of spam / not sure spam or non-spam / not sure non-spam can be obtained with a threshold for the confidence level. For example, all users with a confidence higher than 0.9 can be classified as spam (sure). The ones with a lower confidence level are assigned to the unsure spam class.

Users with a label of one of the three categories (spam, spam not sure, spam sure) are marked as spammers. As a consequence, public posts of those users are set to spam posts and can only be viewed by the spammers themselves. Users where the classifier's confidence was below the threshold are again considered in the next classification procedure. To reduce processing time, however, they are only re-classified when they interacted with the system, e.g. submitted a new post.

The classification results are presented in an admin spam interface. Each category is represented by one tab. Administrators can go through the lists of classifications and change the decision of the classifier. This is especially interesting for the unsure categories, where the classifier can not decide safely. To facilitate the manual activity of identifying users, the admin interface dynamically provides user information with the help of Ajax technology. For example, one can view the user's last bookmarks, the classifiers prediction and confidential level, as well as profile information (e-mail, IP address, registration date).

The implemented spam admin interface can also be used to manually create spam datasets for research purposes. For example, in June 2009, different administrators (three of them were familiar with the system, one unfamiliar) independently reclassified 1000 users of the BibSonomy Challenge Dataset from 2008[12] to see the compliance rate of human annotators.

1.6 Logging Framework

ClickLog is an add-on for bibsonomy to detect and log user interaction on every bibsonomy web page. Every time a user clicks on an anchor within a page, Clicklog recognises it and computes different values which it submits to a log server. First, the clicked anchor element itself will be parsed. The attributes and their containing text are used to get values like anchor title, hyperreference and some id and class values to determine the area of this clicked link. Anchor tags in different areas like navigation, tag cloud, bookmark posts, and publication posts have different class definitions. Next, the position of the anchor in a list of anchors will be determined for later analysis. Next to this, some other meta data will be logged. Date, current webpage location, user agent of the web client, username, session id, complete http request header and at least the mouse position while clicking the anchor in two xy-value-pairs, mouse position in client window and position in document area, which are different, if the user scrolls the document.

To determine every click of a user, a function must be hooked into the browser's event handling. This clicklog function is called at the mouse click event and computes then all for logging the necessary information as described above. All assembled data will be sent with an asynchronous http request to a logging server. The Server sends, immediately after getting the data, the http response, and reorganizes and stores the received data subsequently only, in order not to slow down the interaction with the user. Ideally, the user does not observe any slowdown at all due to the asynchronous request handling in the background.

The logging server is written in java, and has a java script client component that is running in the browser of the user. The server part can either be included in BibSonomy or work as a stand alone server. The logging framework is running since October 2008, and stores approx. 70000 clicks with 115 million bytes per month.

The click log data can be used for analysing the usage behavior of the different users. Typical questions are: Are there types of links that are used below expectation? Are there specific user pages or tag pages that are accessed above average? Are they accessed from the general tag cloud or from some content-specific pages? How intensively are the relations used? Will new features be used, or are they ignored? How many external links are followed, and where to? How is the usage distributed over the two categories ob bookmarks and publications?

[12] http://www.kde.cs.uni-kassel.de/ws/rsdc08/

1.7 Dissemination of BibSonomy

Mailing List Our intension is to build up a denser network around BibSonomy research. Therefore, we have set up a mailing list *bibsonomy-research@cs.uni-kassel.de*. The list is intended for facilitating exchange about research issues and research around BibSonomy. The update of the BibSonomy dump every three months is also published on the mailing list. 285 members are participating in the mailing list.

Cooperation BibSonomy is currently used at the Fraunhofer Institute for Autonomous Intelligent Systems and at SAP Research for internally organizing publications. Primarily, the local implementation serves as first test phase and will be extended after a successful evaluation.

Conference Support The system contains all accepted papers of the conferences Statphys23,[13] ISWC+ASWC 2007,[14] and ESWC 2008,[15] together with the keywords (tags) that authors have associated with their papers (or that show up in the paper titles). To help conference participants find interesting publications, a web-front-end has been created which shows a tag cloud of the most important keywords. The colour of each tag indicates the track to which most abstracts annotated with that tag belong to. Clicking on a tag (keyword) will retrieve from BibSonomy the abstracts that have been tagged with it. Given the necessary BibTeX entries to store publication abstracts, metadata and associated keywords in BibSonomy, we can provide BibSonomy web front-ends presenting a conference's tag cloud and interests.

Discovery Challenge We were organising the Discovery Challenge of the ECML/PKDD 2008 conference.[16] The challenge comprises two tasks: learning tag recommendations, and detecting spam, both based on a BibSonomy dataset. The final results were published in September 2008. Because of its success, the conference organizers asked us to repeat the challenge in 2009; its preparation is currently ongoing. Both challenges were organised as Tagora outreach.

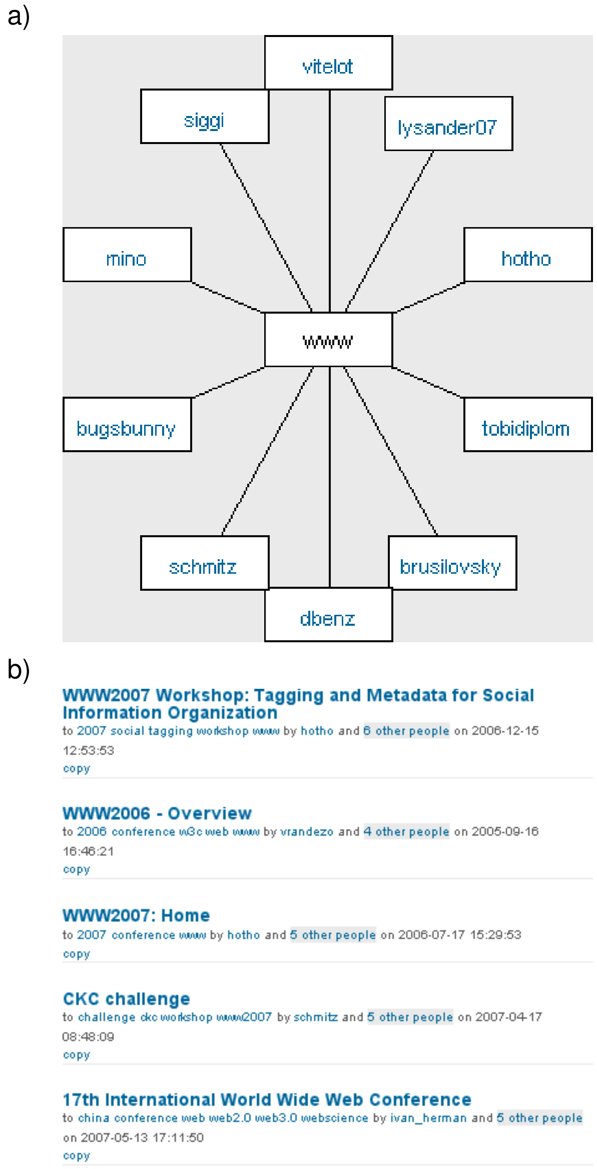
Publication Some new features of BibSonomy have been presented at the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007 (Jäschke et al., 2007).

[13] http://www.bibsonomy.org/events/statphys23

[14] http://www.bibsonomy.org/events/iswc2007

[15] http://www.bibsonomy.org/events/eswc2008

[16] http://www.kde.cs.uni-kassel.de/ws/rsdc08/



WWW2007 Workshop: Tagging and Metadata for Social Information Organization
to 2007 social tagging workshop www by hotho and 6 other people on 2006-12-15 12:53:53
copy

WWW2006 - Overview
to 2006 conference w3c web www by vrandezo and 4 other people on 2005-09-16 16:46:21
copy

WWW2007: Home
to 2007 conference www by hotho and 5 other people on 2006-07-17 15:29:53
copy

CKC challenge
to challenge ckc workshop www2007 by schmitz and 5 other people on 2007-04-17 08:48:09
copy

17th International World Wide Web Conference
to china conference web web2.0 web3.0 webscience by ivan_herman and 5 other people on 2007-05-13 17:11:50
copy

Figure 1.7: a) Ranked users of FolkRank related to the tag ‘www’. b) Ranked websites according to this tag.

Chapter 2

Tagster – Folksonomy Peer-to-Peer System for Sharing Multimedia Data

Tagster is a peer-to-peer tagging application. Very much like Flickr, Del.icio.us, etc. it allows to tag and share personal data. But instead of uploading the data to such an internet service, Tagster organizes and stores everything on the local computer. Tagster is based on a modular architecture, formerly known as the Semantic Exchange architecture (SEA) (Franz et al., 2006) which provides the basic functionality for organizing and sharing annotated information resources in a decentralized scenario. Additionally, a mechanism for managing distributed tagging statistics is integrated and the application provides different data views for easily navigating the annotated multimedia data.

2.1 Status

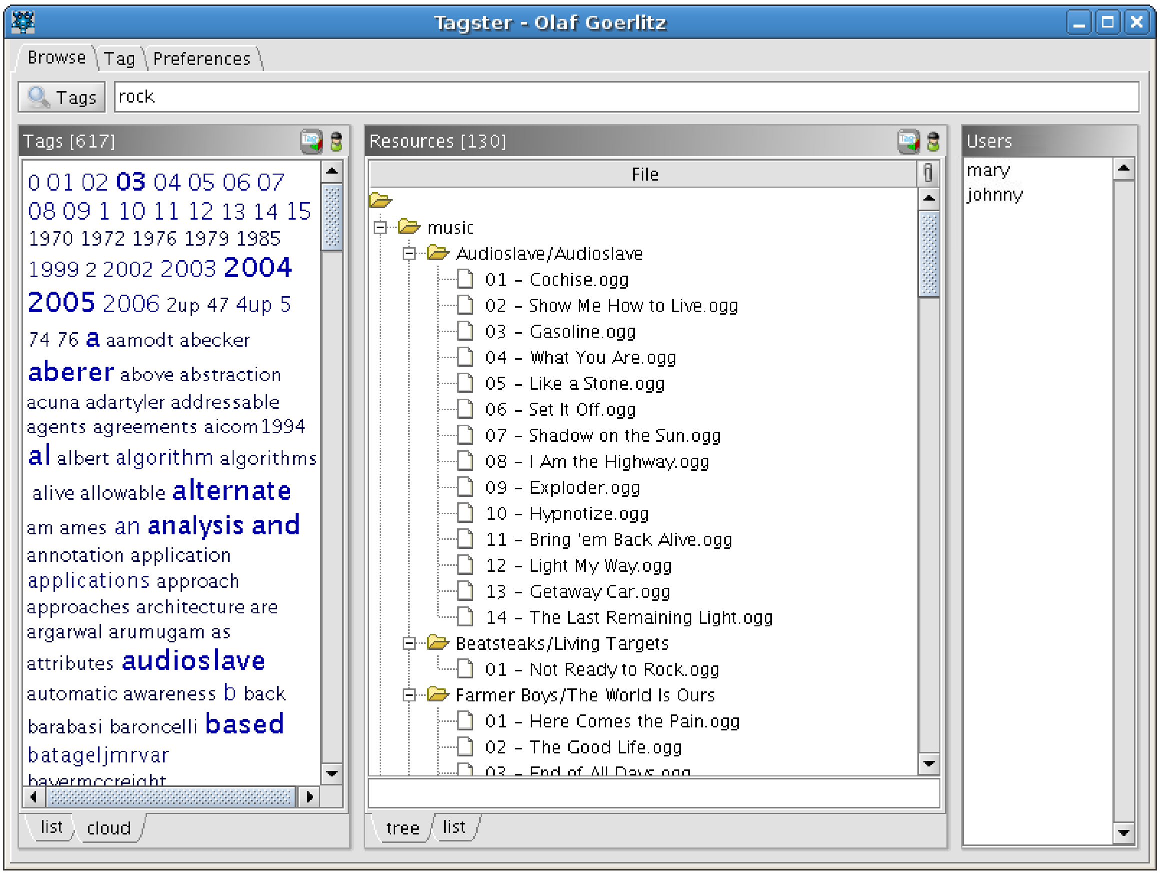
In the first year, a first prototype of Tagster had been developed which provided basic functionalities for tagging multimedia data in a decentralized fashion. During the second year the application was extended in two directions: the management of distributed tagging statistics and improvements concerning the user interaction and usability. However, technical problems like firewall tunneling and the integration of a robust distributed index required much more effort than expected. This led to a delay with respect to the original schedule. Moreover, they problems could only be solved partially with the available resources. Hence the attraction of a critical mass of users until the end of the project was questionable. Consequently, the development of Tagster was stopped to divert the effort to other tasks. Nevertheless, the application was published as opensource on Launchpad (https://launchpad.net/tagster) to give other interested people access to the application beyond the lifetime of the project.

2.2 Features

Distributed Statistics

To make tagging meta data available to all users in the network, Tagster uses a global index structure. That means each peer in the network stores a fraction of the globally available meta data and the underlying index implementation (we use Bamboo[1]) assures that the stored data can be accessed in a very efficient way. However, the index only stores pairwise relations between users, tags, and resources. Handling more complex information retrieval tasks like similarity computation of users and result ranking would require contacting many peers to gather the necessary informa-

[1] http://bamboo-dht.org



rock

0 01 02 03 04 05 06 07 08 09 1 10 11 12 13 14 15 1970 1972 1976 1979 1985 1999 2 2002 2003 2004 2005 2006 2up 47 4up 5 74 76 a aamodt abecker aberer above abstraction acuna adartyler addressable agents agreements aicom1994 al albert algorithm algorithms alive allowable alternate am ames an analysis and annotation application applications approach approaches architecture are argarwal arumugam as attributes audioslave automatic awareness b back barabasi baroncelli based batageljmrvar bayermccreight

- music
  - Audioslave/Audioslave
    - 01 - Cochise.ogg
    - 02 - Show Me How to Live.ogg
    - 03 - Gasoline.ogg
    - 04 - What You Are.ogg
    - 05 - Like a Stone.ogg
    - 06 - Set It Off.ogg
    - 07 - Shadow on the Sun.ogg
    - 08 - I Am the Highway.ogg
    - 09 - Exploder.ogg
    - 10 - Hypnotize.ogg
    - 11 - Bring 'em Back Alive.ogg
    - 12 - Light My Way.ogg
    - 13 - Getaway Car.ogg
    - 14 - The Last Remaining Light.ogg
  - Beatsteaks/Living Targets
    - 01 - Not Ready to Rock.ogg
  - Farmer Boys/The World Is Ours
    - 01 - Here Comes the Pain.ogg
    - 02 - The Good Life.ogg
    - 03 - End of All Days.ogg

mary
johnny

Figure 2.1: Tagster's main view showing the user's tag cloud and all resources tagged with 'rock'.

tion which is apparently highly inefficient. Therefore, we have developed a novel mechanisms for managing distributed statistics, called PINTS (Görlitz et al., 2008).

The basic idea for distributed meta data management is that each peer in the network is maintaining a fraction of the global meta data. With each new tag assignment the responsible index peer updates the respective index information and notifies other peer about the changes if necessary. The similarity computations are based on the cosine similarity of feature vectors, as for example tag clouds. We adapted the well known TF-IDF measure from information retrieval such that each feature combines local and global data like a user's tag frequency and the tag's popularity in the whole network. The problem, however, is to keep the statistics accurate since the propagation of every change of the global index data would cause a high message complexity. Therefore, the PINTS algorithm only propagates data updates if the estimation of the change in the depending statistics is higher than a certain threshold. That allows us to maintain accurate distributed statistical information while keeping the message complexity low in the network. PINTS is implemented in Tagster and used to display statistical information like tag clouds.

User Interface

The design of the user interface plays a major role for the usability of a software. In the case of Tagster it is important to have an interface that provides the same functionalities like the centralized folksonomy systems but also includes an intuitive way of navigating though both the personal data on the local machine and the information retrieved from the network.

However, the application's appearance is not the only aspect we consider for a good user interface but also the ease of use, i.e. the simplicity of configuration and setup/joining of the peer-to-peer network.

Navigation elements The adaption of typical navigation elements like tag clouds from the centralized systems is strongly motivated by the fact that user of such systems are already very accustomed to that type of navigation support. Therefore, one goal is to integrate the same or similar data representations such that users can get familiar with Tagster really quick. This includes, for example, the display of related information for the currently selected data items and contextual tag clouds which are a very typical navigation element.

Additionally, we have integrated resource-specific type views, i.e. the user can browse the resources by their associated Mime type. Thus, it is possible to filter a search results such that only images or documents are displayed.

Tagster's local resource organizations allows the user to tag any file on the local harddisk. Thus, also a file's path information is preserved and displayed in the resource view. To better visualize the local resources we have implemented a tree view that orders all local file in their actual folder hierarchy. Resources from the network are displayed without a hierarchy since that information is not returned for privacy reasons.

Additional functionality Since browsing of resources in the network is not enough, we have also implemented a download protocol for directly retrieving files from other users. To download a resource from the network the user just has to click on the download button next to the resources displayed tagging data. Then the file will be retrieved from the owner and saved in the local download folder. All tags already assigned to the resource will be automatically applied to the downloaded file, too.

Tagster's resource view only diplays the typical file information like name and path. To actually see the content of the files we integrated a function to start the appropriate external application that is associated with the respective file. The intention is that the user does not need to switch manually to another application to view his resources. Currently, this function is supported on Windows and Gnome-based Linux systems.

Configuration and Network Setup

Tagster requires some complex configuration settings due to its distributed nature. Especially, the network setup and peer lookup requires specific port settings etc. that the user ideally should not be bothered with. Therefore, we tried to simplify the configuration and setup process as much as possible. The user is only obliged to enter some personal information at the first start of Tagster. All other system settings are initialized to default values and only need to be changed by the user in some rare cases.

Joining a network is the most complicated part of the initialization process since some bootstrapping information is required. To simplify that process we are using a central user registry that keeps track of all active users in the network. During the network initialization phase, which is completely transparent for the user, the Tagster client first retrieves a list of peers that are used for the network rendezvouz. Afterwards, the client sends its own peer information to the registry.

One important problem of the decentralized approach is the uptime of users. Although data replication is used not all of the distributed index information may be available all the time if only a small number of users is online. To compensate this effect we have set up one peer on a server to run 24/7.

Data Gathering

In order to analyse the users tagging behavior, Tagster includes some data gathering mechanism. The interesting data is theoretically available from the distributed index or can be reconstructed from it. However, crawling the data from the index is not very efficient and can have a negative influence on the system's perfomance. Moreover, the index stores no timestamp information for the tag assignments. Instead, downloading the tagging information directly from the peers is the preferable method but not all users are online at all time.

Currently, we are relying on the cooperation of the initial test user group to extract the desired data directly from their local repository. A more sophisticated data gathering mechanism that can incrementally retrieve the data from the online peers in the network without even bothering them would be desirable but has not yet been implemented.

In the collected dataset we store for each tag assignment the timestamp, the user's ID, the tagged resource's ID, and all assigned tags. Addionally, we log all data queries resulting from the users search and browsing activity. The query logs contain a timestamp, the type of the query (i.e. query for tag, user, or resource) and the queried data item.

2.3 Installing and using Tagster

The Tagster prototype can be downloaded at http://isweb.uni-koblenz.de/Research/tagster. The source code is available at https://launchpad.net/tagster. The installation is very simple as the application comes as an executable Java archive. The initial setup is guided by a setup dialog and Tagster automatically joins the peer-to-peer network.

2.4 Experiences

We have experienced a strong interest in Tagster. Many people like the idea of sharing their resources in a peer-to-peer fashion, especially with colleagues. However, the attraction of a sufficiently large user group with a long-term interest to use the software is critical. A user's client should be running for several hours each day. Otherwise, the interest will quickly diminish because nobody can find other peoples resources.

Another severe problem are firewalls. The libraries integrated in Tagster to maintain the distributed index are all coming form research work at universities and unfortunately they are generally lacking support for firewall handling. An extension into that direction requires some significant effort since the firewall tunneling techniques are rather complicated and the integration in the libraries is hampered by poor documentation.

Chapter 3

Ikoru – A Test-bed for Collaborative Tagging and Content-Based Analysis

3.1 Motivation

The Ikoru system, developed at Sony CSL, is primarily used to experiment with collaborative tagging and content-based analysis. The project consists of a server-side component and a Web interface, which can be viewed at http://www.ikoru.net. Our motivations to develop this project included the following:

Content-based tools: Ikoru evolved from a research project that explored the combination of content-based analysis and collaborative tagging.

Data gathering: Running our own servers allows us to gather detailed user data and explore how the analysis of this data can improve tagging systems.

Extendible research platform: Ikoru aims to be an open platform that can be extended with new analysis and visualisation tools.

Multimedia: It was initially developed for images but has been extended to handle music files.

Small, reusable server: On the technological side, we have designed the Ikoru server as a small and stand-alone web component that can be easily reused and integrated in third-party projects.

We have made the first version of Ikoru available at the end of the first project year. In the second year we have extended the similarity search to audio. We have kept the Web site up and running since last year but we have been focusing more on targeted tagging experiments than on the growth of the Web site.

3.2 Similarity search for audio

In deliverable D3.3 (Section 3.3) we gave a very brief overview of a thorough study on automatic tag suggestion for music, based on the analysis of the audio signal. In that work, state-of-the-art classification techniques, developed at Sony CSL, were evaluated using a large music database that had been tagged consistently by a group of professionals using a well-defined taxonomy. Although the proposed classification technique perform better than existing techniques, the study also shows that the semantic inference remains extremely difficult. A direct translation of the used method from a clean-room database to an online tagging Web site is likely to yield unsatisfactory results.

However, content-analysis was the main reason for Ikoru's existence, and semantic inference one of it's aims. We showed last year that the image similarity search was a powerful tool, in particular when it is used with tags. Using tags, a visitor can narrow down the number of images that are displayed. At this point of the navigation in the archive, the simple content-based search becomes a useful tool to select a subset of the images. It can disambiguate, for example, between images of *apple fruit* and *Apple computers*, both tagged with *apple*.

This year we integrated the *contextual similarity search* for music into Ikoru and we decided to start with the simple approach that had been successful for images. The features we used are mostly those that are defined in the MPEG-7 audio standard[1]. To test it, we ran the feature extraction algorithms on the Last.fm data set, consisting of more than 18000 tagged music snippets of 26 seconds each. The similarity search can be tested on the Ikoru demo site (http://demo.ikoru.net).

3.3 Innovation through targeted experiments

At the outset of the project, we nourished the hope that Ikoru could grow into an active Web site. Despite the strengths of the system, this was somewhat wishful thinking. The reality is that in the last two years many sites have integrated tagging and that these sites can rely on considerable resources and infrastructure to continuously improve their offering. Technology transfers within Sony have been in principle possible and Ikoru has been presented to many product division within the group. However, the collaborations have been not trivial to set up because of the current tendency of Sony to outsource Web services.

As a results Sony CSL doesn't have a precise planning to promote Ikoru to a large audience. Instead, the current strategy is to continue to increase the usability and reliability of the software through its use in small but concrete projects. These focused projects can be managed much more easily and allow us to concentrate on innovative applications of tagging. In the future, we see Ikoru evolve as a generic back-end to store the information about resources, people, and tag assignments. We also see the focus of the tagging applications move away from purely Web-based applications towards real-world applications.

One such project is the artistic installation "Phenotypes/Limited Forms" that was exhibited at the Zentrum für Kunst und Medien (ZKM) in Karlsruhe, Germany, the Bienal de Sao Paulo in São Paulo, Brazil and the "Selective Knowledge" exhibition in Athens, Greece and still is on display in the Museum of Contemporary Art in Siegen, Germany. Although this installation – a joint project with photographer Armin Linke – is a very particular use of Ikoru, it has allowed us to gather a fair amount of data. More than 8000 visitors picked a selection of eight photos and tagged it using a special “editing table” designed for this purpose. The photos, printed in high-quality on solid boards, are taken from an archive of one thousand photos that are displayed on shelves in the exhibition space. Once the visitors tagged their selection, the editing table prints out a small booklet that they can take home.

Another interesting development, that has recently started is the use of Ikoru to store musical melodies (Pachet, 2008). Compared to photos or audio files, melodies can be analysed and generated at a semantically higher level. It has also the potential to reach a small but passionate community.

To facilitate such small tagging projects by other researchers and developers, and to let Ikoru evolve accordingly, we made the source code available under the GNU Library General Public License (LGPL). It can be found at http://sourceforge.net/projects/ikoru.

[1] The feature extraction works as follows. The signal is split in overlapping chunks of 2048 samples (approximately. 46 msec long and 23 msec overlap). Each chunk is weighted by a Hanning window. For each chunk we apply a DSP operator such as, for example, the root mean square (RMS, related to energy level), the zero crossing rate (ZCR). We then calculate the first two statistical moments (mean and variance) to aggregate the values of each chunk into a single global value. Most operators return scalar values (RMS, ZCR, ...) except for Mel-frequency cepstrum (MFCC, 20-dimensional) and the Chroma analysis (measures the 12-tone distribution, 12-dimensional). The complete list of operators include: harmonic spectral centroid, harmonic spectral deviation, harmonic spectral spread, pitch, spectral centroid, spectral flatness, spectral spread, spectral kurtosis, spectral skewness, spectral roll-off, ZCR, RMS, RHF, HFC, IQR, centroid, harmonic spectral variation, MFCC, Chroma.

Chapter 4

Zexe.net – A Community Memory for Representing Daily Experiences using Folksonomies

The Zexe.net system consists of a set of online applications and tools that allow small-scale communities to represent and communicate their views and daily lives on the web. Through the use of smart phones, communities in different cities around the world have published images, videos and sound recordings in Zexe.net for the last five years: taxi drivers in Mexico City, gypsies in Lleida and León (Spain), prostitutes in Madrid, handicapped people in Barcelona and Geneva, and motorcycle messengers (called motoboys) in São Paulo, Brazil. We call these web-based tools Community Memories, as they help communities represent and raise awareness about a commons (Steels and Tisselli, 2008).

Collaborative tagging has become a crucial tool in Zexe.net. Participants not only publish their daily experiences in the form of multimedia files, but they also tag them. Thus, Zexe.net proposes a novel usage for tags by letting users assign them to what we could call "slices of life". In the following sections, the basics of the Zexe.net system will be described.

4.1 Concept & Implementation

For each deployment of the Zexe.net platform participants were given smart phones and were coached to learn how to use them to report about issues they run into in their daily lives. These reports are aggregated on a web portal – the community memory – where they can be explored by other participants.

The reports (containing textual tags and multimedia content) were directly sent from the phones to the Zexe.net web server through the MMS (Multimedia Messaging System) service. This service enables mobile phone subscribers to send a multimedia message directly to an e-mail address. All published resources in the system are unique and they are associated to only one user. Tag assignments occur at the level of a message, meaning that all the multimedia files included in one message will share the same tag assignment. The tags assigned to a message can be entered as a comma-separated list directly on the phone, or by using an online editing application after the message has been received.

On the server side the Zexe.net system comprises a set of PHP scripts and has a MySQL database as its backbone. The data structure used to store and organize the contents sent in by the participants is very straightforward:

- Multimedia elements, called attachments, are bundled together in small packages, called *messages*.

- Messages, in turn, are bundled together in larger containers, called *channels*. A channel can belong to a single participant or to various ones. Collective channels are created in order to aggregate specific shared topics.

4.2 Deployments

Two deployments of the Zexe.net system are of particular interest for TAGora: the *canal*MOTOBOY project* and *GENEVE*accessible*.

4.2.1 canal*MOTOBOY

In *canal*MOTOBOY* (http://www.zexe.net/SAOPAULO; 2007) we have implemented an automatic detection of the singular and plural forms of the tags. These rules were written specifically for the Portuguese language. Thus, tags which exist in both of these forms are bundled together, publicly displaying only the most popular form. Each channel in *canal*MOTOBOY* features its own tag cloud. Tag clouds can be customized so that they emphasize either the frequency or the popularity of tags. Frequency refers to the number of times a tag has been used, either by a single user (in the case of individual channels) or by the whole group (in the collective channel). Popular tags can only be viewed in the tag cloud of the collective channel; their size is proportional to the number of participants who have used them.

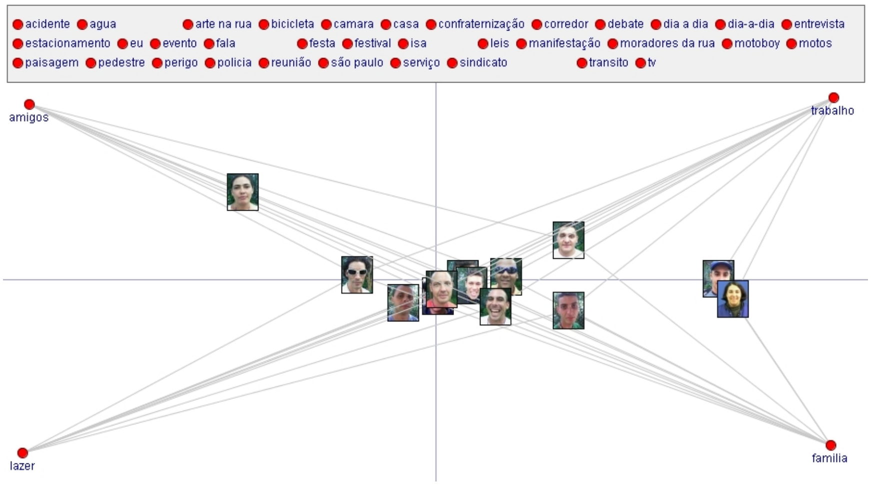

● acidente ● agua ● arte na rua ● bicicleta ● camara ● casa ● confraternização ● corredor ● debate ● dia a dia ● dia-a-dia ● entrevista
● estacionamento ● eu ● evento ● fala ● festa ● festival ● isa ● leis ● manifestação ● moradores da rua ● motoboy ● motos
● paisagem ● pedestre ● perigo ● policia ● reunião ● são paulo ● serviço ● sindicato ● transito ● tv
amigos
trabalho
lazer
familia

Figure 4.1: Tag-Participant Network in *canal*MOTOBOY*.

Figure 4.1 shows the Tag-Participant network, a tool that shows the participants' position in relation to tags on a 2D plane. Tags are attractors, which means that the closer a participant is to a tag, the more he or she has used it. The objective of this tool is to reflect and compare the participants' tagging activities visually.

4.2.2 GENEVE*accessible

Tag cloud for frequent and popular tags can also be found in *GENEVE*accessible* (http://www.zexe.net/GENEVE; 2008) the latest deployment of Zexe.net, which involves handicapped people who portray the state of the urban accessibility in Geneva, Switzerland. However, there are a number of significant innovations:

1. Every participant has a GPS-enabled mobile phone. Whenever the GPS is active, the corresponding geographical information (latitude and longitude) will be embedded in every pho-

tograph they take. Thus, this information is associated to individual multimedia elements. Google Maps is used for on-line GIS support.

2. Tag clustering was improved by introducing the possibility of manually creating groups of synonyms, which can include any number of tags. For example, the group "marches" can contain its singular and plural forms, "marche" and "marches", and also typographic errors, such as "marhe". A second level of clustering is allowed by the possibility of creating channels from tags. Thus, for example, the channel "obstacles" includes the tag groups "dangers", "déviations", "entrées", "escaliers", etc.

3. Keyword searches for tags have been implemented.

4. A special application installed on the mobile phones provides the users of *GENEVE*accessible* with a list of the 10 most popular tags to choose from when tagging a photograph. This application is constantly updated with data coming from the Zexe.net database. The application also allows users to add a tag which is not found on the list.

Because of the specific goals of *GENEVE*accessible*, an initial list of tags to be used for urban obstacles was convened with the participants at the start of the project. Needless to say, these tags quickly became the most popular ones.

4.3 The role of tagging

In both projects, tag clouds act as search interfaces. The way in which tag clouds allow searches can be explained as follows:

1. A user select a desired tag *t1* from the tag cloud. The tag is highlighted, and the results are presented.

2. Only co-occurring tags are enabled in the tag cloud for further selection. If the user selects one of these tags, the search is refined by executing a database query that includes all the selected tags: *t1* AND *t2* AND *t3* AND ... AND *tn*.

3. The user can deselect any of the previously selected tags at any time.

4. To deselect all tags, the user can press the "reset" button at any time.

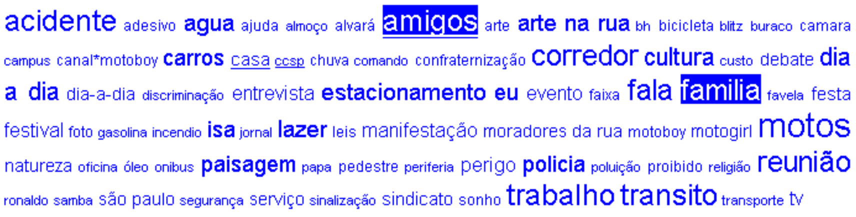

acidente adesivo agua ajuda almoço alvará amigos arte arte na rua bh bicicleta blitz buraco camara
campus canal*motoboy carros casa ccsp chuva comando confraternização corredor cultura custo debate dia
a dia dia-a-dia discriminação entrevista estacionamento eu evento faixa fala familia favela festa
festival foto gasolina incendio isa jornal lazer leis manifestação moradores da rua motoboy motogirl motos
natureza oficina óleo onibus paisagem papa pedestre periferia perigo policia poluição proibido religião reunião
ronaldo samba são paulo segurança serviço sinalização sindicato sonho trabalho transito transporte tv

Figure 4.2: Tag cloud showing highlighted (amigos, familia) and co-occurring (casa, ccsp) tags in *canal*MOTOBOY*

In *GENEVE*accessible*, tag searches not only produce a set of resulting photographs, but are also reflected in the corresponding map. The map itself becomes also an interface for navigation through the images, and can be used interactively, together with the tag cloud. In fact, selected tags or tag groups can be associated to markers of a specific color on the map.

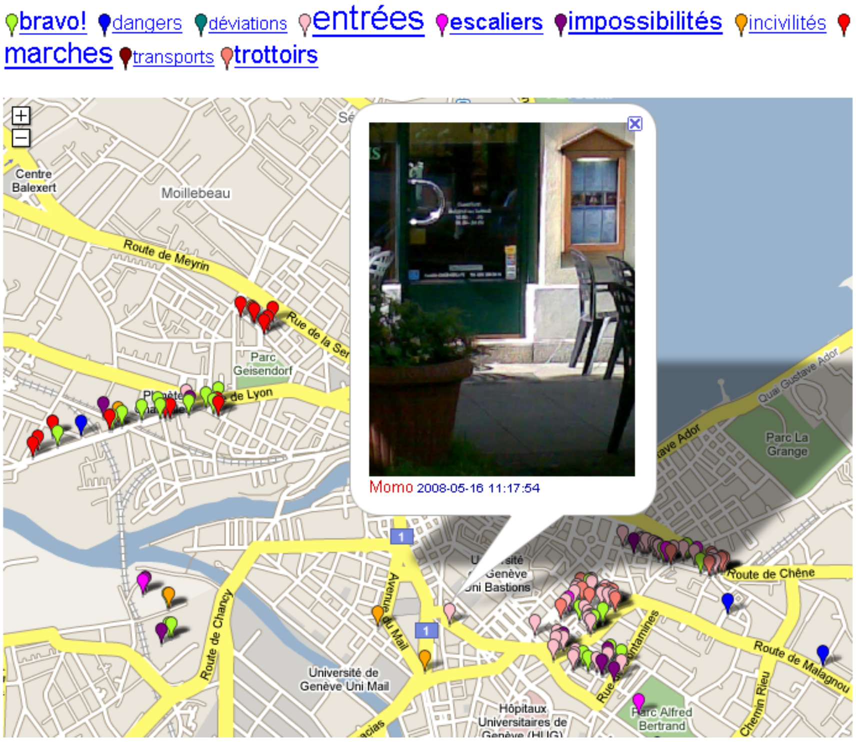


Figure 4.3: The “obstacles” channel in *GENEVE*accessible*, with the tag groups representing the different types of obstacles in the tag cloud. Each tag group has its corresponding marker color on the map.

Besides allowing publication and navigation, the Zexe.net system also offers an online editing tool which participants use to manage their own channels and perform tasks on individual multimedia elements, such as editing their associated tags or their geographical location. The system also provides a control panel for the system administrator, in which different housekeeping tasks can be performed. Among these tasks are the creation of tag groups, the creation of tag-based channels and the choice of the specific tags to be highlighted on the map.

4.4 Analysis of tagging activity

The following table represents the projects' tagging activity figures, as measured on 20/05/2008:

Project	Duration (months)	Users	Tags	Messages	Tag Assignments
*canal*MOTOBOY*	13	15	712	7.975	8.079
*GENEVE*accessible*	3	16	107	2.039	3.188

Vocabulary growth

These graphs show the vocabulary growth of both projects after a first three-month period. Each curve represents the growth of an individual vocabulary.

The following table shows the quantities of messages and tags existing in each project after these initial periods:

Project	Messages	Tags
*canal*MOTOBOY*	2.687	472
*GENEVE*accessible*	2.039	107

It can be observed that, while the number of messages sent by these two groups are quite

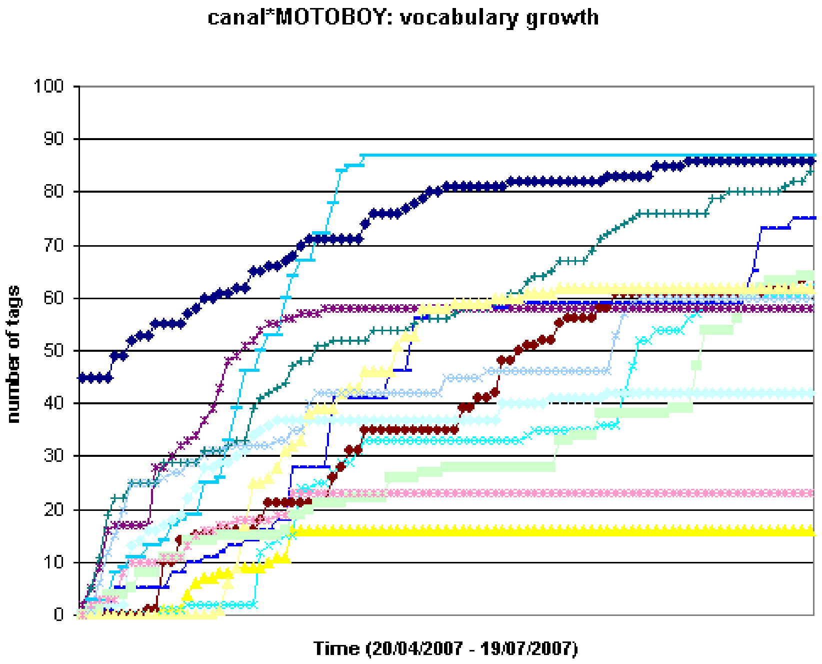


Figure 4.4: Vocabulary growth in the first three months of *canal*MOTOBOY*

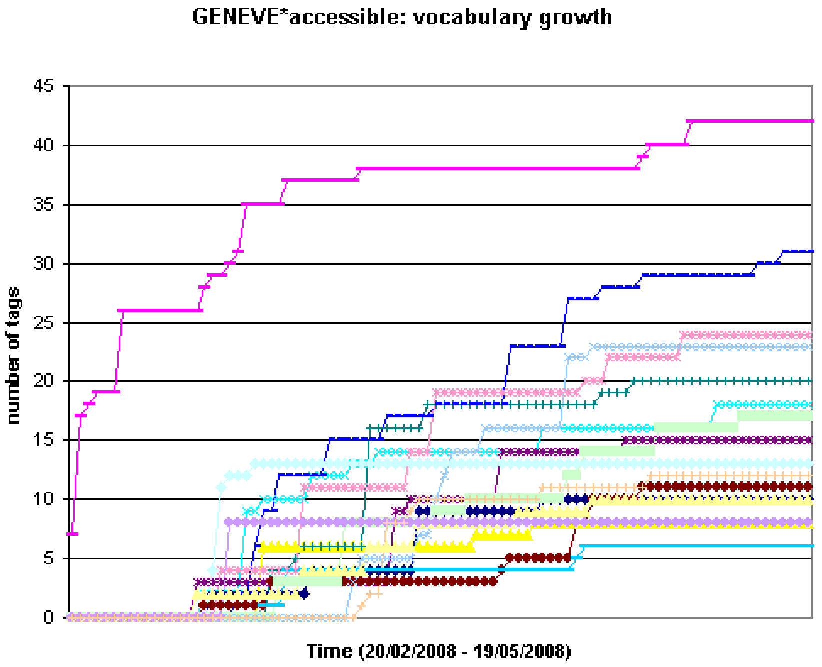


Figure 4.5: Vocabulary growth in the first three months of *GENEVE*accessible*

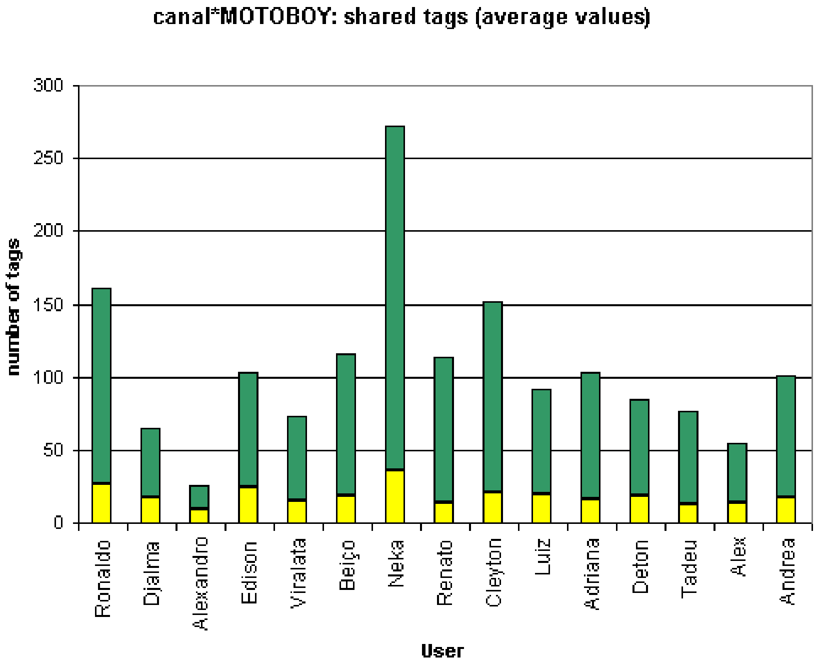


Figure 4.6: Number of distinct tags per participant (green) with average number of shared tags (yellow). Shared tags are tags which are also used by at least another participant.

close, the number of tags in *canal*MOTOBOY* is more than 4 times higher than that of *GENEVE*accessible*. This can be due to the combination of two causes:

1. The *canal*MOTOBOY* project has open-ended goals, while *GENEVE*accessible* has very specific ones. The reduction in the scope of topics to be treated in *GENEVE*accessible* is reflected directly in the number of tags used by the participants.

2. The mobile application that lets users choose tags from a list is only available to the participants of *GENEVE*accessible*. Tag suggestion is very likely dampening the growth of individual vocabularies.

Alignment of tags in canal*MOTOBOY

These graphs represent the extent to which tags are shared in *canal*MOTOBOY*:

It must be noted that the motoboys form a very close group, and meet very regularly to discuss the project.

Alignment of tags in GENEVE*accessible

Individual vocabularies in the *GENEVE*accessible* project show a high degree of alignment, which can be due to the reduced scope of the project itself and the usage of the phone-based application which suggests a list of pre-existing tags.

4.4.1 Self-consistency in tagging behavior

The percentages of non-unique tags in the individual vocabularies of participants in both projects are illustrated in the following graphs:

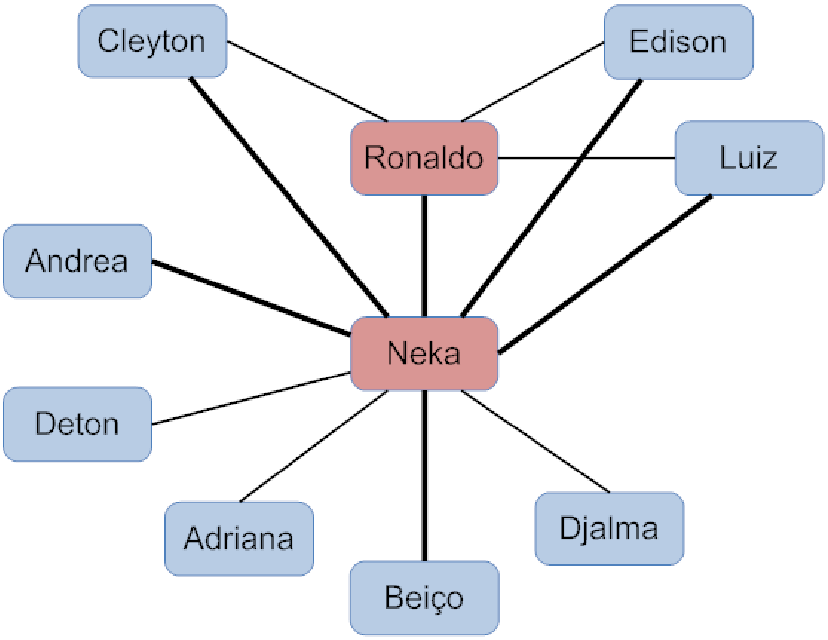


Figure 4.7: Network of participants who share more than 30 tags. Thicker lines represent more than 40 shared tags. Red nodes represent participants who have a degree > 4.

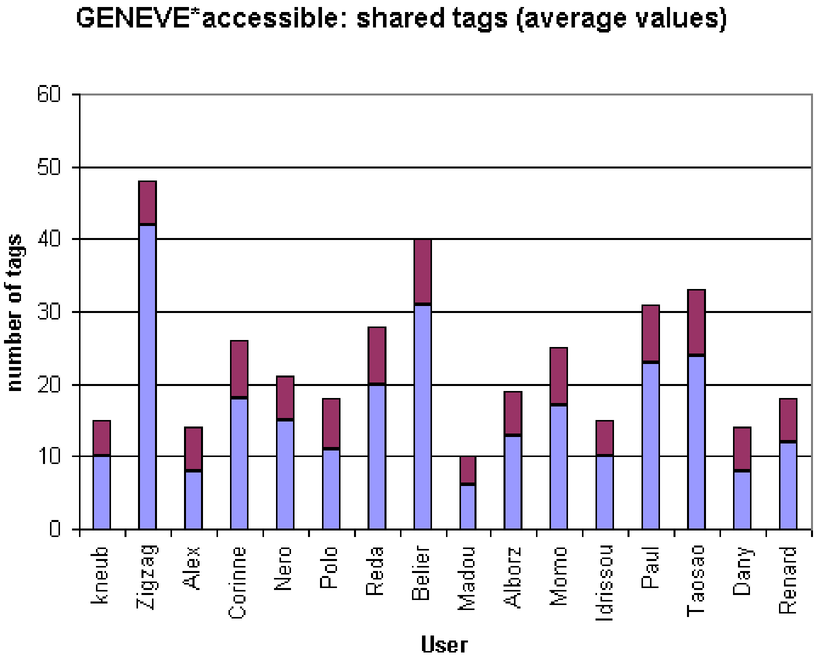


Figure 4.8: Number of distinct tags per participant (purple) with average number of shared tags (light blue). Shared tags are tags which are also used by at least another participant.

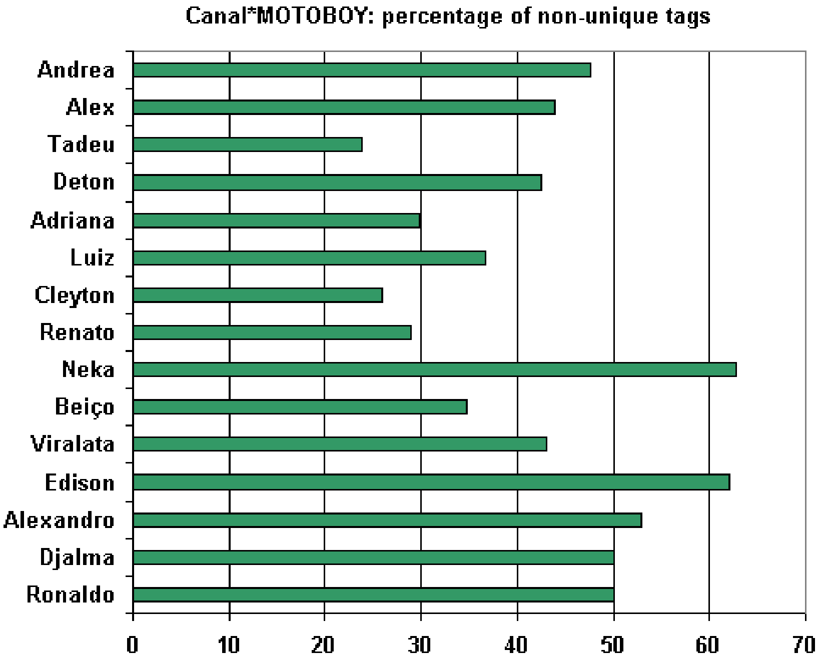


Figure 4.9: Self-consistency of tagging behaviour in *canal*MOTOBOY*

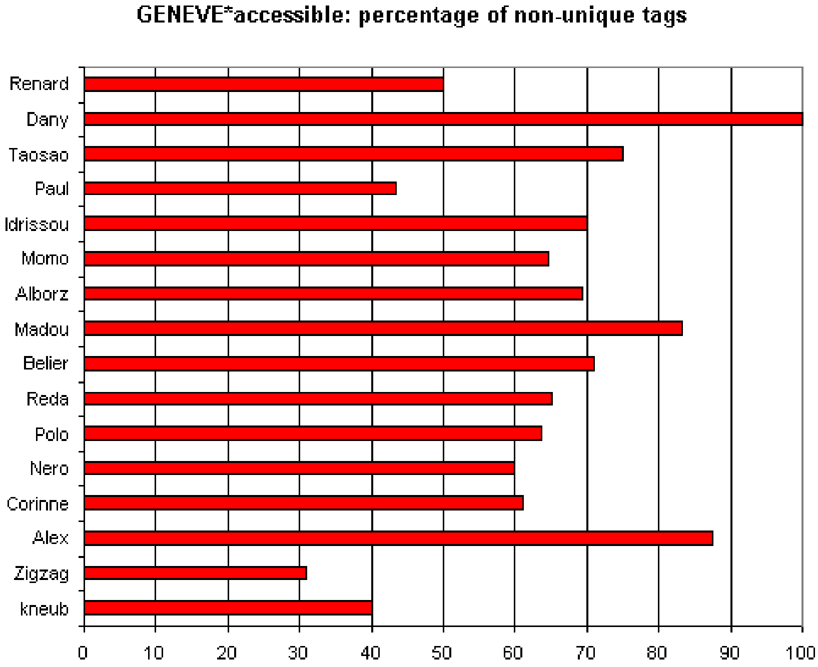


Figure 4.10: Self-consistency of tagging behaviour if *GENEVE*accessible*

The self-consistency measure represents the extent to which individuals are consistent with their own tags: a high percentage means that the participant regularly reuses her tags, while a lower one indicates that the participant normally uses unique tags. The causes which affect vocabulary growth in GENEVE*accessible also determine the comparatively high self-consistency rate of the individual users' tagging behaviour in that project.

4.5 Conclusions

Although the Zexe.net project already existed in some form since 2003, the applications for the two deployments mentioned here were totally re-written in order to support folksonomies. We found that the concept and the mechanics of tagging were understood very quickly by the participants of these projects, even by those who were not technically literate. The inclusion of folksonomies in these projects greatly improved the way in which the participants dealt with emerging topics, and provided a bottom-up way for representing the issues and views of the involved groups in a much more accurate and fine-grained way. By analysing and comparing the tagging activity in *canal*MOTOBOY* and *GENEVE*accessible*, we also show how the scope of a project's focus and tag suggestion can influence the growth and diversity of folksonomies. The Zexe.net system includes the basic functionalities of folksonomies: tagging with or without suggestions, tag clouds which are viewable using different criteria (frequency or popularity), filtering searches through tags and grouping.

Within the scope of TAGora, the ideas behind the Zexe.net platform have lived on in its successor NoiseTube (cfr. Chapter 5) which extends its concepts to create a collaborative community platform to sense and tag exposure to pollution.

Chapter 5

NoiseTube – Pollution tagging

NoiseTube is a participative sensing (Burke et al., 2006) and tagging platform that aims to enable citizens to gather, manage, visualise and distribute data on urban noise pollution. This innovative application of collaborative tagging extends and builds upon the concepts pioneered in the Zexe.net project (cfr. Chapter 4).

The Zexe.net project focussed on exploring new applications of tagging for the benefit of off-line communities facing a variety of issues related to the sustainable exploitation of a commons. Through the use of smart phones, communities in different cities around the world have published content on the Zexe.net platform, resulting in so-called Community Memories (Steels and Tisselli, 2008) which help these communities represent and raise awareness about shared concerns. Starting from the 3rd year of the TAGora project we began to extend this idea by applying collaborative tagging to a new type of resource: the exposure of individual people to pollution.

The challenge we set ourselves was to find ways in which collaborative tagging and data collection through mobile phones can augment the practice of pollution monitoring in cities.

5.1 Context: The case of noise pollution

As a concrete case we studied the problem of noise pollution, resulting in the NoiseTube platform (Maisonneuve et al., 2009). We chose the case of noise pollution because it is a major problem in urban environments, affecting human behaviour, well-being, productivity and health and because we believed it would be possible to measure noise via the microphones of mobile phones.

According to the EU Green Paper (EUC, 1996), "Environmental noise, caused by traffic, industrial and recreational activities is one of the main local environmental problems in Europe and the source of an increasing number of complaints from the public". EU experts estimate that 80 million people suffer from noise levels considered as unacceptable, and 170 million people experience serious annoyance during daytime in the European Union. Recognising this as a prime issue, the European Commission adopted the European Noise Directive (European Parliament and Council, 2002) requiring major cities to establish a noise management policy. The first step is to assess the current noise conditions in cities by gathering real world data and building noise maps in order to better understand the problem and support the creation of local action plans.

While many large cities are already investing in the creation of such maps to comply with the EU directive, these maps generally only show long term averages over large areas. In fact, much of the data on different types of environmental pollution collected by government officials, NGO's and the private sector are incomplete, out of date or largely based on estimations due to a lack of measurements in the field. The assessment of exposure of individuals to pollution is a real problem due the complexity of measuring it on a wide scale. The gathering of personal exposure data representative for a city-wide population is considered to be impractical using traditional approaches. Therefore current monitoring of environmental pollution relies heavily on simulations based on computer models. This comes with a considerable, but hard to estimate, error margin and lacks the spatio-temporal granularity to produce meaningful information on the level of individual citizens and their daily environment.

The NoiseTube project started out from the assumption that it should be possible to improve upon this situation – for the case of noise pollution – if, potentially, every citizen would carry around a personal measurement device that is connected to a citywide network of such sensors.

5.2 Approach

We extended and modified the Zexe.net platform to create a new community memory system called NoiseTube (as in "a YouTube" for noise). The principal goal was to make it possible for citizens to turn their cell phone into a personal, mobile noise sensor. The device monitors the level of environmental noise and feeds this measurement data, along with collaborative tagging data, into a centralised web-based database accessible for everyone.

The advantages of such an approach are multiple. At the individual level, we empower citizens to measure their personal exposure to noise in their daily environment, and to also annotate, geo-localize (manually or through GPS) and publish this data through the NoiseTube community memory website to inform others about their situation. At a collective level, by involving many citizens we are effectively building a sensor network which:

1. is capable of monitoring the dynamics of urban noise pollution with much more spatio-temporal detail than traditional simulation-based approaches;

2. matches the rich diversity of lifestyles, activities of people at a limited cost;

3. supplies annotated data on real individual exposure for future environmental/health studies;

4. builds a collaborative exposure map with a semantic layer facilitating local decisions and collective actions;

5. and, more generally, facilitates the engagement of the citizens in a new environmental role.

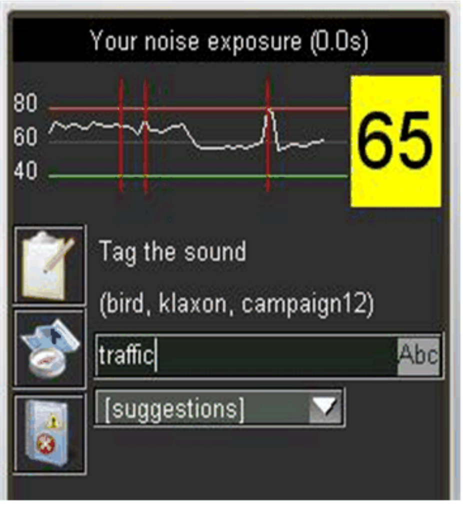
5.3 Features

The NoiseTube platform consists of an application for mobile phones and a web application. The mobile application (see 5.3.1) acts an actual noise level meter which uses the microphone of the phone (or an external one) to measure exposure to noise in real time (in 1 second intervals). This information is shown on the screen of the device and is enriched with metadata such as a timestamp, GPS coordinates and manual and automatic tags, before being sent to the central web application. This web application (see 5.3.2), which is accessible through the NoiseTube website (http://www.noisetube.net), collects and aggregates the measurements from the distributed network of phone-based noise sensors and provides features to navigate, download and visualise the data in different formats.

5.3.1 Mobile application

The mobile application we have developed generates two, equally important, streams of data. First of all it turns the phone in a noise meter which, when used properly, measures the level of environmental noise with reasonable accuracy. Secondly the mobile application complements the raw measurement data with a stream of metadata tags such as timestamps, geo-tags, human annotation tags and automatic exposure pattern tags.

(a) A NoiseTube contributor measuring the level of environmental noise at a construction site using her mobile phone as a personal noise sensor instrument. The data is directly sent to the NoiseTube web application to update the exposure map of the city



(b) The user interface of the mobile application. On top, the visualization of the exposure. On the right, the current L_{eq} in dB(A) with a colour representing the degree of risk (green, yellow or red). At the top left, the log the exposures with red vertical lines representing tag assignments. In the centre, a free text field to tag the exposure with a dynamic suggestion list)

Measuring personal noise exposure

To turn off-the-shelf smart phones into noise sensors we implemented a signal processing algorithm which computes – in real-time – accurate noise level measurements based on an audio signal. The accuracy of the measurement was evaluated with laboratory tests which were carried out in a sound-proof audio studio. In a later stage we conducted a series of real world tests in the streets of Paris, in collaboration with BruitParif[1], the official observatory of noise pollution for the Parisian region. We concluded this evaluation and calibration process with an observed average error of $\pm$ 2.5 dB(A), which we consider accurate enough for our goals.

Social tagging to contextualise pollution and identify causes

Measuring noise exposure is not enough. We also need to identify the causes of the pollution to react on it. As people are excellent at recognising noise sources, they can annotate the measures regarding the cause or context of their exposures such as cars, aircraft or neighbours via the mobile application to inform the community about it. In fact public noise maps often provide only a very limited information regarding the source or context of noise. This sort of semantic information is collected through social collaborative tagging. This type of metadata is vital to build meaningful noise maps for both citizens and decision makers.

Geo-tagging

Because indoor positioning is almost impossible using GPS, we let users indicate their location using a free text (e.g. an address) or by referring to an entry in a list of, predefined, "favourite" places (e.g. “home”, “office”) that will be transformed into GPS coordinates using geocoding. For example, by specifying subway stations a path followed in the subway can be reconstructed afterwards (see Fig. 5.1).

[1] http://www.bruitparif.fr

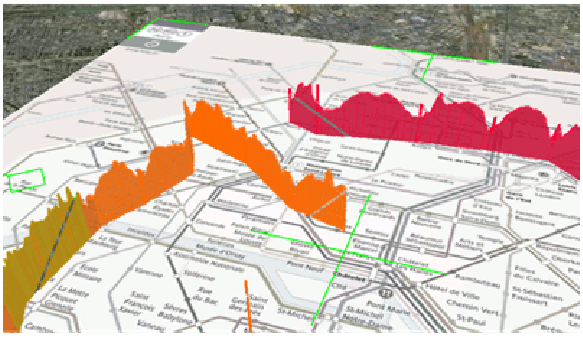


Figure 5.1: Noise map of two subway lines (indoor location) reconstructed using the geo-tagging feature in the mobile application

Automatic tagging of exposure patterns

Using techniques to extract semantic descriptions of low level features – similar to ideas from the Ikoru project; the mobile application not only acts as a sensor but also does post-processing to detect exposure patterns and tag them automatically. For the moment two basic patterns are supported:

> **sudden high variation**: when there is a sudden high variation (+ 15 dB(A)) in a short time (< 3 seconds) the application automatically adds the tag *"sudden peak"* to the last measure.
>
> **long and risky exposure**: When the user is exposed to a high level of noise (> 80 dB(A) for a long time (> 20 seconds) the mobile application automatically adds the tag *"risky exposure"*.

Extracting basic patterns to generate tags allows adding a semantic description of these levels of exposure and the tags can be used to power the same navigation and visualization features as the other (human) tags. In our case we can discover clusters where people had sudden high variations or long dangerous exposure without further inferences. Furthermore using the phones to distribute the computational effort allows a better scalability of the general system.

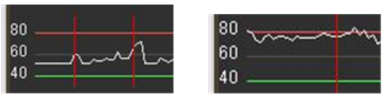


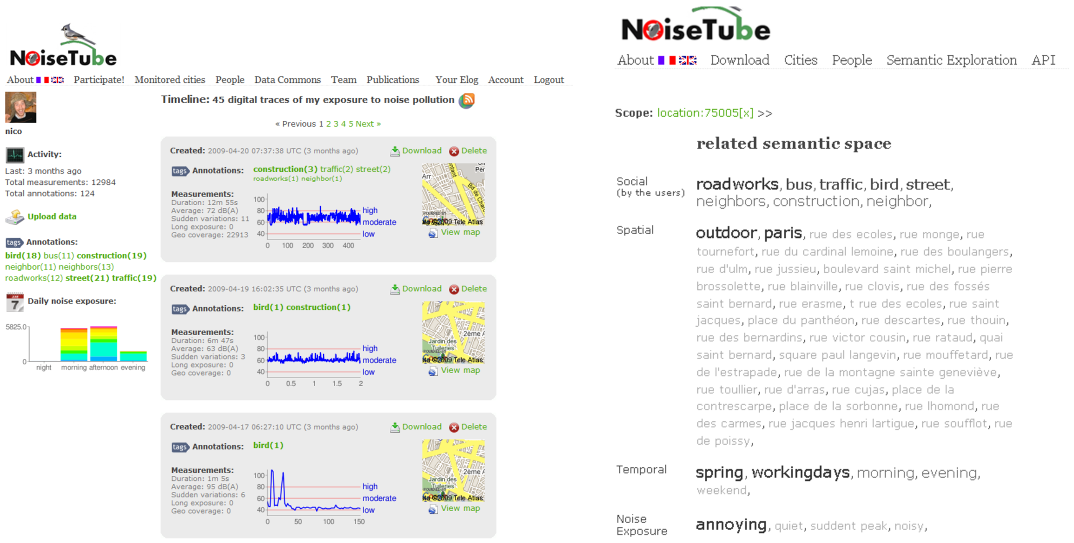
Figure 5.2: Screenshots of the noise exposure histogram in the phone UI, showing tagged events (red lines). The application can automatically tag different basic patterns of exposure such as sudden large variations in loudness or long lasting exposures to riskful loudness levels

5.3.2 Web application

The web application provides several ways to navigate and visualise the exposure to noise of people in cities. Once the measured data are sent the server, any user can see his own contributions or exposures by going to the NoiseTube website (http://www.noisetube.net). Fig. 5.3(a) and 5.3(b) show screenshots of principal parts of the web application.

Semantic Exploration

Navigating or searching meaningful information through a large amount of numerical data captured by sensors is a difficult task. People think and reason with concepts, not numbers, requiring


NoiseTube
About Participate! Monitored cities People Data Commons Team Publications Your Elog Account Logout
nico
Upload data
5825.0
0
night
morning
afternoon
evening
100
80
60
40
0
100
200
300
400
high
moderate
low
Download
Delete
View map
0.5
1
1.5
2
50
150
NoiseTube
About Download Cities People Semantic Exploration API

Timeline: 45 digital traces of my exposure to noise pollution

« Previous 1 2 3 4 5 Next »

Activity:
Last: 3 months ago
Total measurements: 12984
Total annotations: 124

Annotations:
bird(18) bus(11) construction(19) neighbor(11) neighbors(13) roadworks(12) street(21) traffic(19)

Daily noise exposure:

Created: 2009-04-20 07:37:38 UTC (3 months ago)
Annotations: construction(3) traffic(2) street(2) roadworks(1) neighbor(1)
Measurements:
Duration: 12m 55s
Average: 72 dB(A)
Sudden variations: 11
Long exposure: 0
Geo coverage: 22913

Created: 2009-04-19 16:02:35 UTC (3 months ago)
Annotations: bird(1) construction(1)
Measurements:
Duration: 6m 47s
Average: 63 dB(A)
Sudden variations: 3
Long exposure: 0
Geo coverage: 0

Created: 2009-04-17 06:27:10 UTC (3 months ago)
Annotations: bird(1)
Measurements:
Duration: 1m 5s
Average: 95 dB(A)
Sudden variations: 6
Long exposure: 0
Geo coverage: 0

(a) The *eLog* (environmental log or exposure log) page showing a list of recorded digital traces of the exposure of a user

Scope: location:75005[x] >>

related semantic space

Social (by the users) roadworks, bus, traffic, bird, street, neighbors, construction, neighbor,

Spatial outdoor, paris, rue des ecoles, rue monge, rue tournefort, rue du cardinal lemoine, rue des boulangers, rue d'ulm, rue jussieu, boulevard saint michel, rue pierre brossolette, rue blainville, rue clovis, rue des fossés saint bernard, rue erasme, t rue des ecoles, rue saint jacques, place du panthéon, rue descartes, rue thouin, rue des bernardins, rue victor cousin, rue rataud, quai saint bernard, square paul langevin, rue mouffetard, rue de l'estrapade, rue de la montagne sainte geneviève, rue toullier, rue d'arras, rue cujas, place de la contrescarpe, place de la sorbonne, rue lhomond, rue des carmes, rue jacques henri lartigue, rue soufflot, rue de poissy,

Temporal spring, workingdays, morning, evening, weekend,

Noise Exposure annoying, quiet, suddent peak, noisy,

(b) The "‘semantic exploration"’ feature enables users to explore data by drilling down on a dataset by iteratively selecting tags (some automatically generated) from different semantic dimensions

Figure 5.3: NoiseTube web application

expertise and effort to translate these numerical values into a higher level of meaning. In this context we developed a new feature in Noisetube called "Semantic Exploration" (see Fig. 5.3(b)) to facilitate the exploration of annotated (i.e. tagged) noise exposure data.

The system works by generating and associating tags, for different dimensions such as "time" and "location", for measurement data to allow interpretation at a higher semantic level. This projection of numerical data into a semantic space should allow an easier way to explore the data for any user, especially the non-expert ones.

These tags are automatically generated by a set of "‘interpreters"’, each responsible to assign tags for a specific dimension of the semantic space. Currently we have defined four such semantic dimensions: time, space, social and environmental. For instance the interpreter "‘Location"’ will take as input measurements with its associated geographical coordinates (e.g. "latitude=48.85;longitude=2.35") and will use reverse geocoding to assign additional tags such as street name, zip code and city, and the tag "‘outdoor"’ because this interpreter considers any measurement with GPS-coordinates to have been captured outside. The interpreter "‘Time"’ will for example assign tags such as "‘morning"’, "’winter"’, "‘workingday"’ to the measure dating from "‘12/02/08 10:02:02"’ (using also the location of the measure to determinate the season).

Once tagged by people or by the machine, the data can be explored by any user via a section in the website called "‘Semantic Exploration"’. In this section, the dataset of measurements is represented semantically using a set of tag clouds, one for each dimension. To explore the dataset one can play with the different dimensions and filter recursively by clicking on the tags. Each time a tag is selected the user effectively drills down or zooms in on the data. In each such iteration the semantic representation (i.e. the tag clouds) of the results is recomputed to get a clear semantic picture of the selected subset. The results can also be projected in the geographical space by downloading a dynamically generated map (as a KML file for Google Earth) representing the current subset.

We believe that the combination of automatically generated tags, along with human assigned tags, organised in different semantic dimensions and the iterative drill-down exploration feature provides users with a powerful and innovative way to navigate in large environmental datasets such as offered by NoiseTube.

Visualising exposure map with a semantic layer

For each city an noise map can be downloaded as a KML file, which can be visualised using the Google Earth[2] application. Each map consists of multiple layers, such as the noise exposure layer, consisting of all measurements and a semantic layer with all tags. This map is constructed by aggregating all the shared contributions. The layer of tags adds a context and meaning to the physical measurements of noise pollution allowing, for instance, to identify the sources of noise (see Fig. 5.4).

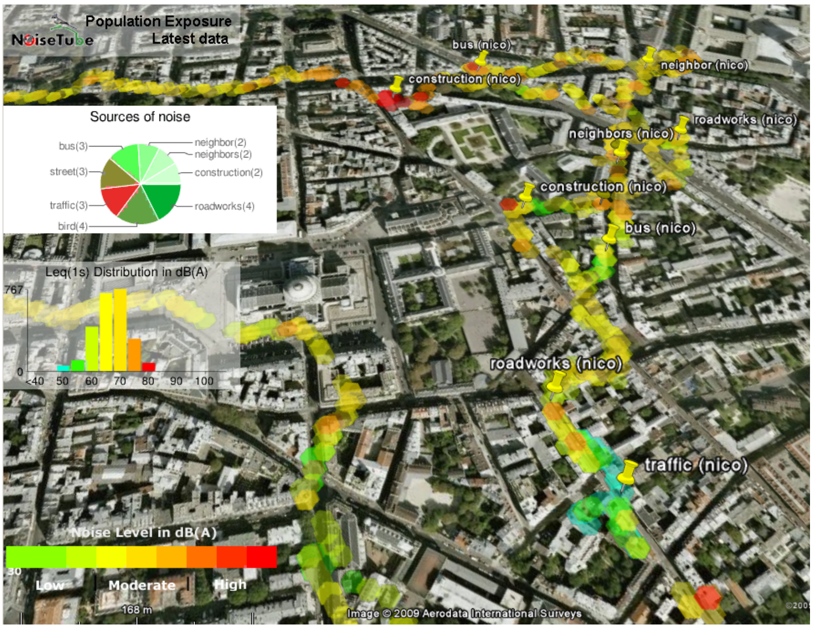


Figure 5.4: Visualisation using Google Earth representing the collective exposure to noise pollution generated by all the measurements and a tag cloud (shown as pie chart) with sources of noise

The concept of a tag cloud is represented on a map by a pie showing the distribution of the different sources of a given area. We created a feature allowing to contextualise this tag cloud according to the geographical area displayed: if the user moves or zooms, the tag cloud is recomputed in real time.

Web API to access public data

Data in environmental pollution and exposure to it are generally not directly accessible for the public or scientists, limiting their exploitation by third parties. the NoiseTube platform provides a simple web API for publishing or accessing raw data and tags. Using this API, external users can access annotated, individual or collective noise exposure data, for example to create web mash-ups or to analyse data.

5.4 Implementation

The current version of the mobile application was written in Java and is aimed primarily at smart phones running the Symbian/S60 operating system. The program was mainly tested on Nokia smart phones such as the N95 8GB. Although untested, many other phone brands and models are supported as well, as long as the device supports the Java J2ME platform, with multimedia

[2] http://earth.google.com

and localisation extensions. A GPS receiver (built in or an external unit that is connected via Bluetooth) is needed to localize measurements. The web application is implemented using Ruby on Rails, Postgres, PostGIS, Google Maps and Google Earth.

5.5 Conclusion

The participatory approach to noise monitoring using mobile phone, as pioneered by NoiseTube, fits within a larger trend of using mobile phones for environmental purposes (Patel-Predd, 2009; The Economist, 2009).

The use of collaborative tagging and other related features such as the automatic tagging of pollution data to better understand their contexts and bootstrap their interpretation are innovations directly coming from the research done within TAGora. The inclusion of folksonomies greatly improved the usability of the system by providing the users with a bottom-up way for representing the issues related to their exposure to pollution in a much more accurate way and at a semantic level, improving navigation and search though large amounts of environmental data.

Despite its being a relatively young project, NoiseTube has already been presented at several workshops, conferences and covered by the press (refer to final PDK document for a full list). Furthermore several collaborations with NGO's (e.g. Awaaz foundation[3], India) and environmental agencies (e.g. BruitParif, the official observatory of Paris) are currently being set up to continue this research track beyond the TAGora project.

[3] http://www.awaaz.org

Chapter 6

TAGnet – a tool for awareness and management of personal metadata

6.1 Concept

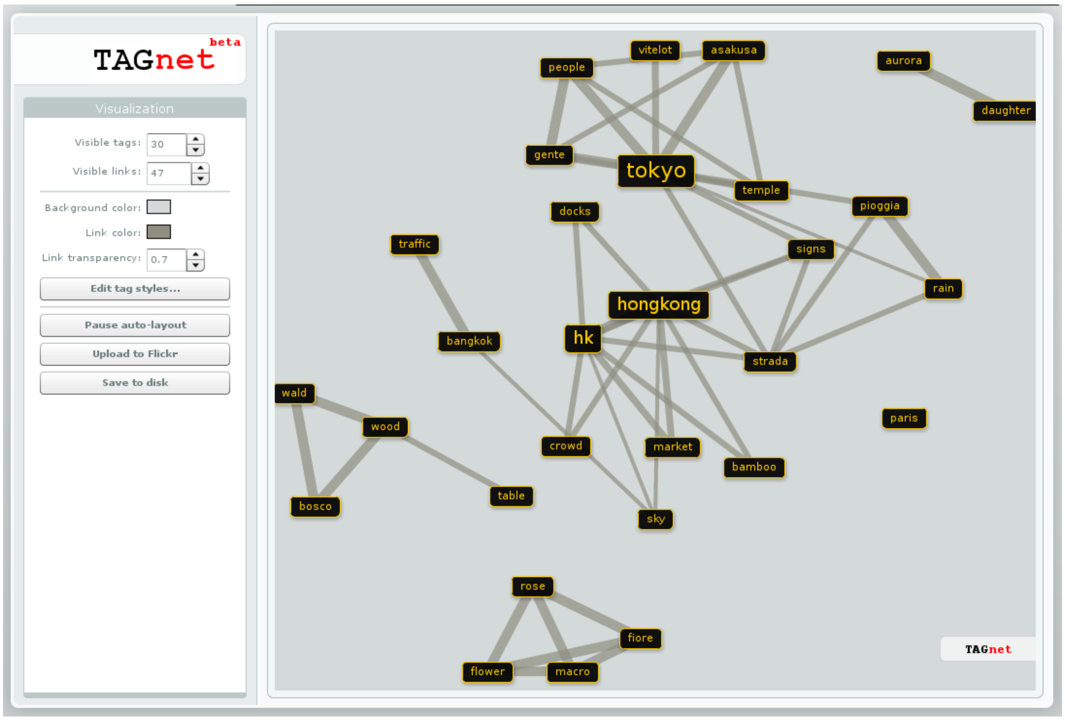
People are not fully aware of the metadata they use to annotate resources. Tagging requires little effort and implies no strong committment to consistency: this fosters the externalization of large bodies of metadata, and at the same time makes the structure of the metadata rather unpredictable, even for the user performing the annotation. In the context of a single user, the tag co-occurrence network exposes many of the semantic relations among tags, and between tags and the broader context defined by user's interests and experiences. Visualizing the tag co-occurrence network and allowing the user to manipulate it provides her with a sort of "semantic mirror" that can be used for awareness, for navigation, and for re-organization of metadata. To this end, an application initially not envisioned in the work-plan, *TAGnet*, was designed to exploit the results of WP3 and WP4. The concept was developed by the PHYS-SAPIENZA team in collaboration with the ISI Foundation in Turin. We were able to leverage resources that allowed us to develop this application at no cost for the Consortium. On achieving maturity the application was renamed from *TAGnet* to Netr, because the domain tagnet.org was unavailable for registration. The application was publicly deployed as *Netr* at the URL http://www.netr.it. In the following we will interchangeably refer to this application using any of the two names.

TAGnet is a prototype web-based application (htp://www.netr.it) designed to provide users with a reflexive tool to expose regularities and patterns in their own tag-based annotations. Tagging patterns can reveal a lot about a user's experience, her interests and her emergent conceptualizations, but users are not aware of these patterns until these regularities are made explicit by means of data analysis and state-of-the-art visualization. TAGnet currently targets Flickr users and provides them with a view on their annotations (tags) that exposes actionable information.

6.2 Implementation

The development of TAGnet takes the move from an early experiment in this direction (http://www.visualcomplexity.com/vc/project.cfm?id=231), coded in Python and using Python bindings for the Flickr APIs (http://flickrapi.sourceforge.net/).

The very same logic and code were used to develop an interactive visualization by using state-of-the-state technology. In particular, the visualization client (Fig. 6.1) is a web-based application built on top of the Adobe Air framework (http://www.adobe.com/products/air/) and developed in Flex (http://flex.org/). The visualization and the force-based layout system use a customized version of the Flare (http://flare.prefuse.org) visualization toolkit. The Flickr APIs (http://www.flickr.com/services/api/) are used to fetch data from Flickr.



Visible tags: 30
Visible links: 47
Background color: ▨
Link color: ▨
Link transparency: 0.7

Figure 6.1: TAGnet at work.

A Flickr user can connect to the *TAGnet* web site and insert her username. The system fetches from Flickr a list of the tags associated with each annotated photo and computes a co-occurrence network, which is subsequently visualized by using the above force-based layout engine. The user interface allows users to tune the number of tags displayed by the interface, and the threshold of co-occurrence controlling whether a link is drawn or not between two tags. The user can also dynamically exclude tags, mark two tags as “synonyms”, mark a tag as a lexical variation of another tag, mark a few tags as “important”, and so on. By means of these actions, supported by the user interface, users can further structure the metadata and drive the visualization towards what they think is a better representation of the categories and conceptual structures they consider relevant. The user can switch among several different layout schemes and even freeze the layout engine to arrange tags manually according to her will. The resulting (user-manipulated) visualization of the tag co-occurrence network can be uploaded to Flickr and shared by clicking on a button.

6.3 Perspectives and road-map

In perspective *TAGnet* will be also used as a tool to explore emergent conceptualizations and tag ranking strategies for social annotations. To this end, extensive logging of interface events has been foreseen so that one can compare the measures of node importance and link strength computed by our system with the same notions as explicitly expressed by the user by means of the interface controls. This will yield insights into node ranking and similarity (Cattuto et al., 2008a,b; Markines et al., 2009) in folksonomies, as well as a better understanding of what constitutes a better graphical layout (from the perspective of the end-user) in visualizing tag metadata.

In the long term, the user interface of the system will be cloned and customized to set up user studies targeting specific questions on user behavior, emergent categorization and conceptual structures, as exposed by the annotations of a given user. These experiments will be kept separated from the main system not to impair the applicative goal of *TAGnet*, which will be improved and kept focused as a tool for reflexive exploration of tagging patterns in the context of a single user.

Chapter 7

MyTag – Personalized Search and Exploration

7.1 Concept

Nowadays Web 2.0 platforms like YouTube, Flickr and del.icio.us provide large amounts of resources such as videos, photographs and social bookmarks. Common to the platforms is the classification by so called tags that can be used for organization and retrieval. A current limitation of tagging platforms is their confinement to a single media type. Furthermore, a magnitude of platforms exists for each media type. Thus, in both cases of searching resources of either the same or of different media type, a user has to search multiple platforms. For example, a user needs to search on del.icio.us, RawSugar and Bibsonomy to find bookmarks or on Flickr and YouTube to find media related to e.g. an artist. Another limitation results from the ranking of resources as implemented by platforms such as YouTube, Flickr, and del.icio.us. Usually, the overall popularity of a resource is used for ranking search results. A personalized search is currently missing that takes the interests of a user into account.

MyTag aims at solving the previously described limitations of current tagging platforms by searching different content types like photos, videos and social bookmarks from different sources in parallel. The search is transparently executed via the public APIs of the different tagging systems and the retrieved results are presented in separate columns for each content type (see Fig. 7.1). Besides, MyTag collects the search interests of registered users and offers them a personalized ranking of results. Additionally, an intelligent search assistant that helps in disambiguating the current search terms by grounding them to possibly relevant articles found in DBPedia. The architecture of MyTag ensures its extensibility towards further tagging platforms.

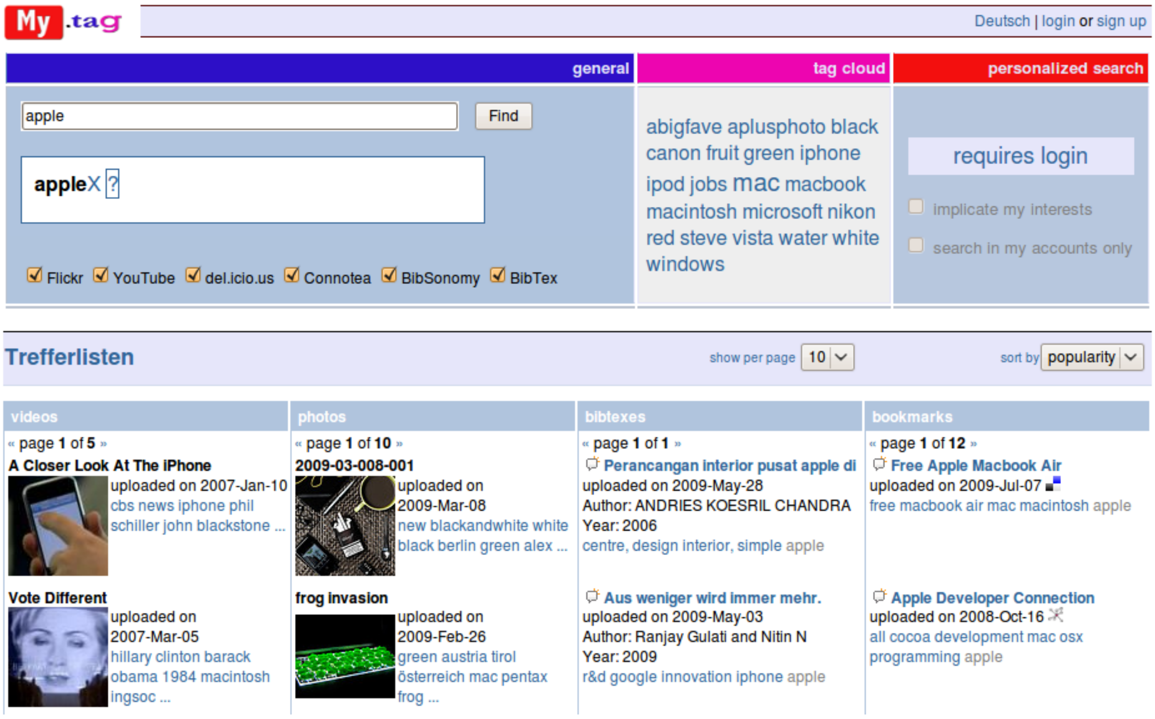
7.2 Features

- Incorporating further platforms and content types: With Bibsonomy and Connotea, two additional platforms were introduced into the system. By incorporating Bibsonomy, MyTag now also supports the search in bibliographic references.

- Merging search result lists for the same content type: By incorporating Bibsonomy and Connotea, we now retrieve search results for bookmarks from three different platforms. This required to introduce an algorithm into Mytag that merges the bookmarks coming from Delicious, Connotea and Bibsonomy into a single result list. The technical details about the merging algorithm are available in (Grabs, 2009).

- Intelligent search assistant I: In a collaborative effort with the Southampton team, we introduced an intelligent search assistant into MyTag. It automatically analyzes the current search terms of the users and sends it to a disambiguation web service offered by the Southampton University. This web service grounds the search tags in articles from DBPedia and returns possible related terms. MyTag then analyzes the returned list of possible meanings of the search terms and filters those meanings which are not represented in the search results retrieved from the tagging platforms. The remaining grounded terms are then presented to the user. The user can then select the intended meaning of the search term and re-rank the current list of results so that resources corresponding to the intended meaning are ranked higher.

- Intelligent search assistant II: In (Abbasi and Staab, 2009) a method is proposed how to identify generalized and specialized tags. This analysis was applied on the collected Flickr data set and the resulting lists of generalized and specialized tags were also used for suggesting tags to a user during his search. But the evaluation in (Scharek, 2009) showed that users only seldomly found the suggested tags useful for refining their search. Because of these mixed evaluation results, this search assistant was never included into the publicly available version of MyTag.



apple

appleX ?

☑ Flickr ☑ YouTube ☑ del.icio.us ☑ Connotea ☑ BibSonomy ☑ BibTex

abigfave aplusphoto black canon fruit green iphone ipod jobs mac macbook macintosh microsoft nikon red steve vista water white windows

□ implicate my interests
□ search in my accounts only

show per page 10 sort by popularity

« page 1 of 5 »
A Closer Look At The iPhone
uploaded on 2007-Jan-10
cbs news iphone phil schiller john blackstone ...

Vote Different
uploaded on 2007-Mar-05
hillary clinton barack obama 1984 macintosh ingsoc ...

« page 1 of 10 »
2009-03-008-001
uploaded on 2009-Mar-08
new blackandwhite white black berlin green alex ...

frog invasion
uploaded on 2009-Feb-26
green austria tirol österreich mac pentax frog ...

« page 1 of 1 »
Perancangan interior pusat apple di
uploaded on 2009-May-28
Author: ANDRIES KOESRIL CHANDRA
Year: 2006
centre, design interior, simple apple

Aus weniger wird immer mehr.
uploaded on 2009-May-03
Author: Ranjay Gulati and Nitin N
Year: 2009
r&d google innovation iphone apple

« page 1 of 12 »
Free Apple Macbook Air
uploaded on 2009-Jul-07
free macbook air mac macintosh apple

Apple Developer Connection
uploaded on 2008-Oct-16
all cocoa development mac osx programming apple

Figure 7.1: A screenshot from MyTag

7.3 Implementation

MyTag is implemented using the Ruby on Rails framework as it supports efficient development of web-applications. The MyTag architecture realizes the model-view-controller paradigm (MVC). A view layer at the top is responsible for the interaction with the user while the control layer in the middle processes data from the model layer, e.g. by computing personalized rankings.

Two personalization features are provided for search: First, a search can be restricted to resources uploaded by the user. This feature requires that a user enters her external account names for Flickr, del.icio.us, and/or YouTube into her profile. Searching only own resources is implemented by using the corresponding feature from the integrated tagging platforms. The second personalization feature allows for ranking search results based on the user's personomy. The personomy is automatically built based on the resources the user picks from the result set. It is modeled by a vector p of tag frequencies representing the previous search interests of the user. As it is based on the implicit feedback given by selecting from the search results, no additional user effort is required to gain personalization. Using implicit user feedback is a very promising approach to personalizing search results or web browsing in general. This feature adds an advantage compared to systems such as Flickr and del.icio.us, where personalization requires adding resources to the system, i. e. the explicit feedback of users.

Chapter 8

Live Social Semantics – A Platform for Connecting People

8.1 Concept

Most conference attendees would agree that *networking* is a crucial component of their conference activities. We strove to significantly further the state of the art by developing LSS; a Semantic Web application that integrates (a) the available wealth of linked semantic data, (b) the rich social data from *existing* major social networking systems, and (c) a physical-presence awareness infrastructure based on active radio-frequency identification (RFID). These data layers were brought together to create a rich and integrated network of information (see figure 8.1).

Acquiring and integrating these heterogeneous, but overlapping, data sources enabled us to provide a new experience and services to conference attendees. The main goal was to encourage conference participants to network, to find people with similar interests, to locate their current friends, and to make new ones.

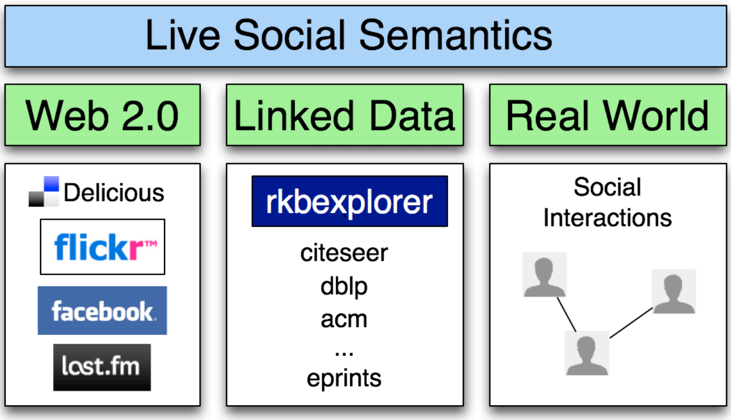


Figure 8.1: The layercake of LSS

8.2 Implementation

Data from various Web 2.0 sources were imported using APIs or screen scraping, and subsequently converted to RDF. Figure 8.2 provides a global picture of the Live Social Semantics framework. The vertical axis partitions the diagram according to two spaces: the virtual world (i.e. data about individuals held in the web), and the real world (i.e. RFID contact data). Data in the virtual world is sourced from social networking sites, to obtain social tagging data and contact networks, as well as the Semantic Web (SW), to obtain information about publications, projects, and the individuals COP (via RKBExplorer and semanticweb.org).

The Profile Builder (center, top of diagram) processes an individual's tagging activities and links them to DBpedia[1] URIs using the TAGora Sense Repository. Similarly, their favourite music artists from LastFM are linked to DBpedia URIs using DBTune[2]. In turn, the Profile Builder automatically suggests to users a list of interests that they can edit, and elect to expose to other participants.

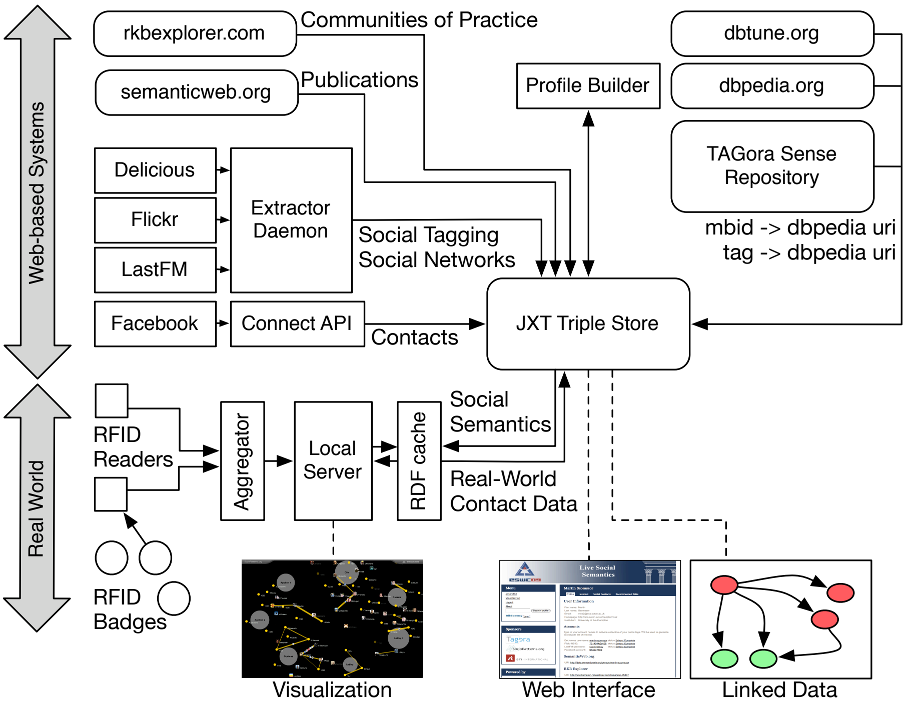


Figure 8.2: LSS system architecture

8.3 Features

LSS provided several services to the users. Figure 8.3 shows an overview of some of those services. LSS mined real-world interactions of conference attendees using hardware and software infrastructure developed by the SocioPatterns project ((Barrat et al., 2008)). The conference name badges of the users were equipped with active RFID badges. The RFID badges engage in multi-channel bi-directional radio communication, and by exchanging low-power signals which are shielded by the human body, they can reliably assess the continued face-to-face proximity of two individuals. We assume continued face-to-face proximity to be a good proxy for a social interaction between individuals.

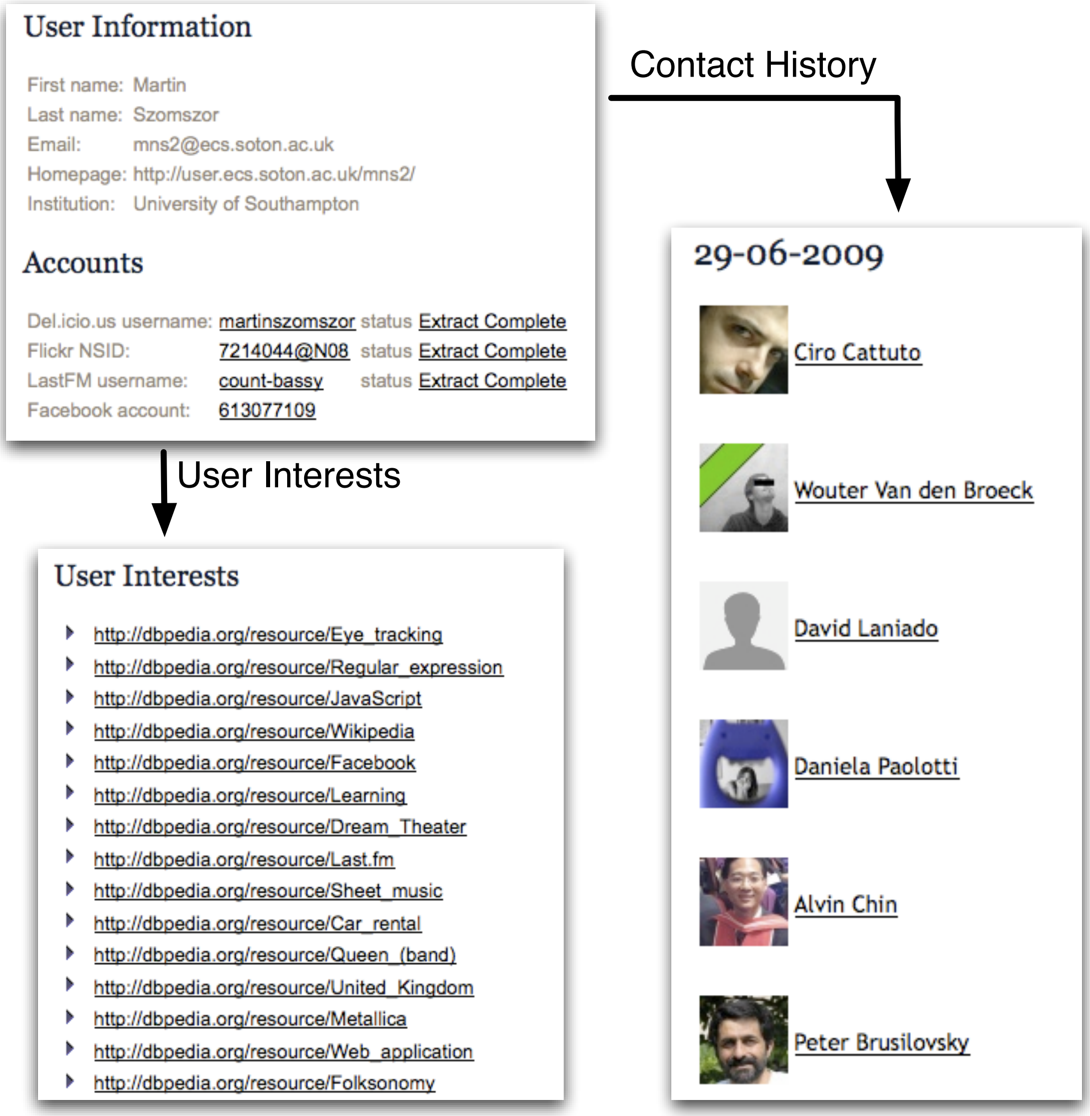
LSS provides two types of visualisations of RFID-contacts; a global visualisation, showing all users in the conference rooms and their live face-to-face contacts, and a personal visualisation, showing the accumulative view of someone's face-to-face contacts.

In previous work (Szomszor et al., 2008), we devised an architecture to automatically generate a list of DBpedia URIs to represent interests a person might have by reasoning over their social tagging activity. This was integrated into LSS, where any social tagging information from Delicious and Flickr is collected and converted to an RDF representation (according to the TAGora tagging

[1] http://dbpedia.org
[2] http://dbtune.org/

ontology[3]). This information is then used to generate a profile of interest for each user (see D4.5 for further detail).

In addition to the features above, LSS also displayed lists of online-friends that are at the conference, and recommended conference talks based on the social connection of a person, as well as on his/her community of practice.



User Information

First name:	Martin
Last name:	Szomszor
Email:	mns2@ecs.soton.ac.uk
Homepage:	http://user.ecs.soton.ac.uk/mns2/
Institution:	University of Southampton

Accounts

Del.icio.us username:	martinszomszor	status	Extract Complete
Flickr NSID:	7214044@N08	status	Extract Complete
LastFM username:	count-bassy	status	Extract Complete
Facebook account:	613077109		

User Interests

- http://dbpedia.org/resource/Eye_tracking
- http://dbpedia.org/resource/Regular_expression
- http://dbpedia.org/resource/JavaScript
- http://dbpedia.org/resource/Wikipedia
- http://dbpedia.org/resource/Facebook
- http://dbpedia.org/resource/Learning
- http://dbpedia.org/resource/Dream_Theater
- http://dbpedia.org/resource/Last.fm
- http://dbpedia.org/resource/Sheet_music
- http://dbpedia.org/resource/Car_rental
- http://dbpedia.org/resource/Queen_(band)
- http://dbpedia.org/resource/United_Kingdom
- http://dbpedia.org/resource/Metallica
- http://dbpedia.org/resource/Web_application
- http://dbpedia.org/resource/Folksonomy

29-06-2009

Ciro Cattuto

Wouter Van den Broeck

David Laniado

Daniela Paolotti

Alvin Chin

Peter Brusilovsky

Figure 8.3: Interests and social contacts information displayed by LSS

8.4 Deployment

LSS was very successfully deployed at the 2009 European Semantic Web Conference (ESWC09) and HyperText Conference (HT09).

At ESWC09, 139 of the conference attendees registered on our LSS site and together entered 246 social profiles from Delicious, Flickr, lastFM and Facebook. Out of those users, 59 entered at least one tagging account (Delicious, Flickr, or lastFM). Our policy was not to use the generated profile

[3] http://tagora.ecs.soton.ac.uk/schemas/tagging

unless it is verified and saved by the users, to avoid publishing anything that the users might not be happy with. In the end, 31 users had a non-empty profile of interest generated for them. When generating those profiles, a total of 1210 DBPedia concepts were proposed (an average of 39 per person across the 31 profiles), out of which 247 were deleted.

When comparing the results from Delicious and Flickr, we see that 17% of concepts proposed from Delicious Tags were deleted, and 32% respectively for Flickr tags. This suggests that the accuracy of topics harvested from Delicious tags was more accurate than those from Flickr. Inspection of the concepts removed shows that Flickr was likely to suggest concepts referring to years and names.

More detail about LSS can be found in (Alani et al., 2009).

Bibliography

Green Paper on Future Noise Policy. Technical Report COM(96) 540, Commission of the European Communities, November 1996. URL http://ec.europa.eu/environment/noise/greenpap.htm.

XWiki. 2004. http://www.xwiki.com.

Rabeeh Abbasi and Steffen Staab. RichVSM: enRiched Vector Space Models for Folksonomies. In *HYPERTEXT 2009, Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*, Turin, Italy, 2009. ACM.

Harith Alani, Martin Szomszor, Gianluca Correndo, Ciro Cattuto, Alain Barrat, and Wouter Van den Broeck. Live Social Semantics. In *Proceedings of the International Semantic Web Conference (ISWC)*, Westfields Conference Center near Washington, DC, 2009.

Alain Barrat, Ciro Cattuto, Vittoria Colizza, Jean-François Pinton, Wouter Van den Broeck, and Alessandro Vespignani. High resolution dynamical mapping of social interactions with active RFID, 2008. http://arxiv.org/abs/0811.4170.

Jeffrey A. Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. Participatory sensing. In *World Sensor Web Workshop (WSW'06) at ACM SenSys'06, October 31, 2006, Boulder, Colorado, USA*, October 2006. URL http://www.sensorplanet.org/wsw2006/6_Burke_wsw06_ucla_final.pdf.

Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems. In *Proceedings of the 3rd Workshop on Ontology Learning and Population (OLP3)*, pages 39–43, Patras, Greece, July 2008a. ISBN 978-960-89282-6-8. URL http://olp.dfki.de/olp3/. ISBN 978-960-89282-6-8.

Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In Amit P. Sheth, Steffen Staab, Mike Dean, Massimo Paolucci, Diana Maynard, Timothy W. Finin, and Krishnaprasad Thirunarayan, editors, *The Semantic Web – ISWC 2008, Proc.Intl. Semantic Web Conference 2008*, volume 5318 of *LNCS*, pages 615–631, Heidelberg, 2008b. Springer. URL http://dx.doi.org/10.1007/978-3-540-88564-1_39.

Martin Dougiamas. Moodle. 1999. http://www.moodle.de.

European Parliament and Council. Directive 2002/49/EC relating to the Assessment and Management of Environmental Noise. *Official Journal of the European Communities*, 18.7.2002:12–26, 2002. URL http://ec.europa.eu/environment/noise/directive.htm.

Thomas Franz, Carsten Saathoff, Olaf Görlitz, Christoph Ringelstein, and Steffen Staab. SEA: A Lightweight and Extensible Semantic Exchange Architecture. In *Proceedings of the 2nd Workshop on Innovations in Web Infrastructure. 15th International World Wide Web Conference (Edinburgh, Scotland)*, 2006.

Olaf Görlitz, Sergej Sizov, and Steffen Staab. PINTS: Peer-to-Peer Infrastructure for Tagging Systems. In *Proceedings of the Seventh International Workshop on Peer-to-Peer Systems, IPTPS08*, Tampa Bay, USA, February 2008.

Daniel Grabs. Beschreibung und Evaluation des MyTag Merge Algorithmus. Master's thesis, Universität Koblenz-Landau, 2009.

Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information Retrieval in Folksonomies: Search and Ranking. In *Proceedings of the 3rd European Semantic Web Conference*, volume 4011 of *LNCS*, pages 411–426, Budva, Montenegro, June 2006. Springer. ISBN 3-540-34544-2. URL http://www.kde.cs.uni-kassel.de/hotho/pub/2006/seach2006hotho_eswc.pdf.

Robert Jäschke, Miranda Grahl, Andreas Hotho, Beate Krause, Christoph Schmitz, and Gerd Stumme. Organizing Publications and Bookmarks in BibSonomy. In Harith Alani, Natasha Noy, Gerd Stumme, Peter Mika, York Sure, and Denny Vrandecic, editors, *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007. URL http://www2007.org/workshops/paper_25.pdf.

Beate Krause, Christoph Schmitz, Andreas Hotho, and Gerd Stumme. The Anti-Social Tagger - Detecting Spam in Social Bookmarking Systems. In *AIRWeb '08: Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, pages 61–68, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-159-0. doi: http://doi.acm.org/10.1145/1451983.1451998. URL http://airweb.cse.lehigh.edu/2008/submissions/krause_2008_anti_social_tagger.pdf.

Nicolas Maisonneuve, Matthias Stevens, Maria E. Niessen, Peter Hanappe, and Luc Steels. Citizen Noise Pollution Monitoring. In Soon Ae Chun, Rodrigo Sandoval, and Priscilla Regan, editors, *Proceedings of 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, volume 390 of *ACM International Conference Proceeding Series*, pages 96–103. Digital Government Society of North America / ACM Press, May 2009. URL http://portal.acm.org/citation.cfm?id=1556176.1556198.

Benjamin Markines, Ciro Cattuto, Filippo Menczer, Dominik Benz, Andreas Hotho, and Gerd Stumme. Evaluating Similarity Measures for Emergent Semantics of Social Tagging. In *18th International World Wide Web Conference*, pages 641–641, April 2009. URL http://www2009.eprints.org/65/.

F. Pachet. Description-Based Design of Monophonic melodies, 2008. Submitted to Computer Music Journal.

Prachi Patel-Predd. Cellphones for Science. *IEEE Spectrum*, 46(2):16, February 2009. ISSN 0018-9235. doi: 10.1109/MSPEC.2009.4770599.

Matthias Scharek. Optimierung von Suchmaschinen basierend auf dem Suchverhalten von Benutzern im Internet. Master's thesis, Universität Koblenz-Landau, 2009.

L. Steels and E. Tisselli. Social Tagging in Community Memories. In *Proceedings of the 2008 AAAI Spring Symposium*, Social Information Processing, Stanford University, California, USA, 2008.

Martin Szomszor, Harith Alani, Ivan Cantador, Kieron O'Hara, and Nigel Shadbolt. Semantic Modelling of User Interests based on Cross-Folksonomy Analysis. In *submitted to Int. Semantic Web Conf., Karlsruhe, Germany*, 2008.

The Economist. Sensors and sensitivity. *The Economist – Technology Quarterly*, 391(8634): 21–22, June 6th 2009. URL http://www.economist.com/sciencetechnology/tq/displaystory.cfm?story_id=13725679.

Zope. Kebas Data. May 2002. http://sourceforge.net/projects/kebasdata/.