

Towards federated Web2.0 sites: the TAGMAS approach

Jon Iturrioz
ONEKIN Research Group
University of the Basque Country
San Sebastián (Spain)
jon.iturrioz@ehu.es

Oscar Díaz
ONEKIN Research Group
University of the Basque Country
San Sebastián (Spain)
oscar.diaz@ehu.es

Cristóbal Arellano
ONEKIN Research Group
University of the Basque Country
San Sebastián (Spain)
carellano001@ikasle.ehu.es

ABSTRACT

The success of Web2.0 is draining user's resources from the desktop to the Web. An increasing number of users are keeping their pictures at *flickr*, their bookmarks at *del.icio.us*, their documents at *googleDocs* and so on. There are important advantages to be gained, but this dissemination of user's resources should go hand-by-hand with tooling that permits users to keep a global view of their resources regardless of where they are kept. Unfortunately, heterogeneity on APIs, tag conventions and message protocols hinders interoperability. Consequently, this work promotes a loosely-coupled federated view of Web2.0 sites which powers traditional desktops with tagging and searching capabilities that expand over the desktop folders to transparently account for Web2.0 sites. This federation is achieved in a user basis: the Web2.0 sites to be integrated are those that keep resources of the user at hand. The paper introduces the current status of TAGMAS, a TAG Management System that provides an interface to deal with multiple, autonomous Web2.0 sites from the desktop.

Keywords

Semantic Desktop, Folksonomies, Web 2.0

1. INTRODUCTION

One of the effects of Web2.0 is the dissemination of user resources. To benefit from the "social lever", users are moving their resources from the desktop to the "folk servers", i.e. Web-based servers that assist a given community in storing, sharing and tagging its resources. Examples are well-known: *delicio.us* [2] manages bookmarks, *google Docs* [7] manages documents, *flickr* [3] manages photos, *citeUlike* [1] manages bibliographical entries and so on.

From a user perspective, this gradual migration of resources from the desktop to folkserver evidences the important advantages brought by Web2.0 in general, and social tagging in particular [8]. However, this situation leads to resources from a given user being disseminated all along the "folksphere" including the user's desktop itself. This hinders the vision of the Semantic Web as a holistic, semantically-related network of resources. Current situation looks like more as a folkserver archipelago where each folkserver has its own folksonomy and semantic concepts. This dispersion forces a user interested in looking for her own resources being tagged, for instance, as "*Banff*", to iterate along distinct folkservers to retrieve the pictures (at *flickr*), tourist information (at *del.icio.us*),

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

conference-related papers (at *citeUlike*), and her Powerpoint presentation (at her desktop).

Although the feasibility (and convenience) of this holistic view has yet to be demonstrated, a more affordable endeavor is to attempt such a holistic view but on a user basis. That is, our vision is that folkserver should behave more like a kind of federated databases where integration is on user basis. In such a vision, each folkserver is an independent and autonomous entity that has its own local users, proprietary folksonomy, specific functionalities (e.g. wrapping such as in *citeUlike*) and resources. On top of the folkserver net however, distributed tagging and searching facilities should be built to isolate the user from the repositories keeping the resources.

Following the database similitude, a federated approach is characterized by the existence of a schema available for all users. This somehow resembles the Semantic Web vision. However, the existence of this global schema seems to be contra-intuitive in a tagging setting. Unlike database where the administrator sets this schema, tagging is a social, collaborative activity. Tags emerge from within the communities, and blending folksonomies from different communities resembles to be as difficult as in a database setting [11].

Other option is that of multidatabase systems. These systems lack a global schema, and interactively construct one as needed by the application. For our case, this implies each folkserver keeping its total autonomy, and the desktop acting as a federated wrapper. This approach seems to be better aligned with our purposes.

This paper introduces a multidatabase approach to folkserver: TAGMAS (TAG Management System). TAGMAS¹ follows a loosely-coupled federated approach where the user is the administrator of his own federated schema. We abstract from tag, envelope or site dependencies by using a layered architecture that permits the user a tag-based search of resources using SPARQL [4]. This declarative query is then gradually transformed into a set of distinct invocations where the specificities of each folkserver (data model, location or envelop protocol) is considered. The results are then transported back where details about the envelop protocol or location are gradually removed till "raw resources" matching the query are rendered to the user which ignore where the resource is located.

The rest of this paper introduces the architecture and gives a detailed account of the distinct use cases that TAGMAS permits. Some conclusions end the paper.

2. TAGMAS ARCHITECTURE

This paper argues that a loosely coupled architecture may be better suited for integrating heterogeneous and autonomous folk-

¹The term "*tagma*" stems from invertebrate biology. A tagma is a specialized grouping of arthropodan segments, such as the head, the thorax, and the abdomen with a common function.

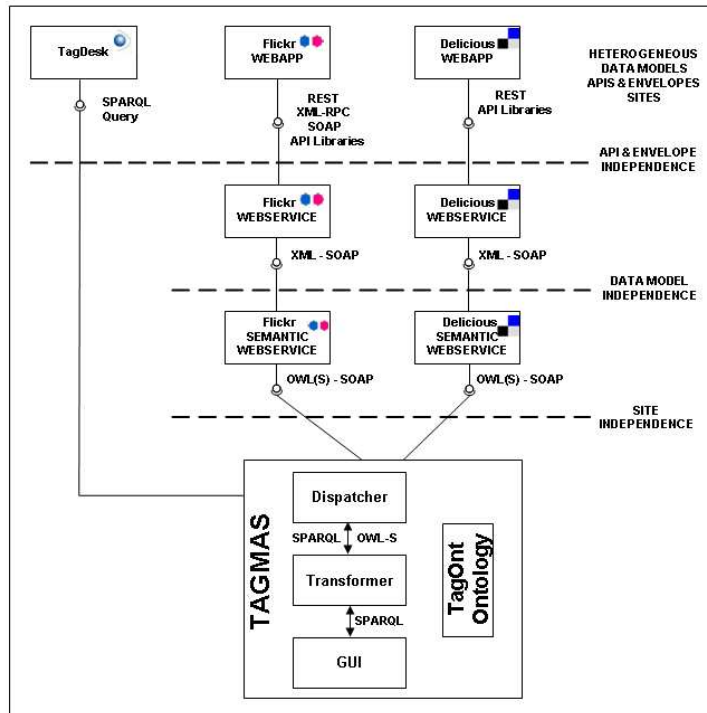


Figure 1: The TAGMA architecture.

servers. This heterogeneity stems from four main sources, mainly, the data model, the API model, the enveloped model and the site place. Figure 1 shows the distinct layers that gradually abstract away from this heterogeneity so that the desktop can handle resources without being involved with the data model, API, envelope, or site where the resources are located. Next paragraphs discuss each of these layers.

API and Envelope independence. The disparate protocols to communicate with actual folkserver (REST, XML-RPC or SOAP) difficult server communication and integration. The goal on this layer is to offer a common way to route petitions (i.e. the envelope) offering a homogeneous syntactic interface using web service's SOAP.

Data model independence. A key point to integrate different data sources is a formal description of each data source that permits its automatic integration by machines. Tims Berners-Lee said that “*Semantic web is an extension of the current one, in which information is given well defined meaning, better enabling computers and people to work in cooperation*” [5]. Ontologies realize this vision, where terminology of heterogeneous systems is specified using ontologies, and ontology mapping and reasoning can resolve the mismatches between the systems thereby realizing semantic integration. For this purpose, the tagging ontology described at [10] is used. The access to data sources is realized through Semantic web services, where service parameters are described in terms of the underlying ontology. Each web service has a semantic-web-service counterpart that offers a machine-understandable interface so integration is facilitated.

Site independence. This layer incorporates the source of the

tag (i.e. the folkserver) as part of the tagging ontology. By turning the folkserver location into tag metadata, the location is easily available to agents to route queries to appropriate locations. For instance, the query “get all tags used by a tagger X” outcomes a mash-up of several tags and resources obtained from different data sources that have been ontologically aligned. Additionally, end-users can now put additional restrictions on their queries based on the source of the tag without leaving their favorite query language (e.g. SPARQL). As an example, users can now “get all resources being tagged as T1 in source S1, and T2 in source S2”. This somehow permits users to manually bridge tagging conventions used in different folkserver, should ontology integration fail to achieve this mapping.

It is important to notice that the desktop is a main source of resources. Consequently, the desktop itself has been wrapper as a local Web2.0 server (since TAGMA is deployed at the desktop). This implies that desktop resources can also be tagged. However, the desktop is local to TAGMA and this makes futile the use of Web Services to communicate between these two local processes (i.e. the resource manager of the desktop and TAGMA). Hence, the wrapper of the resource manager has been made SPARQL aware, i.e. able to understand SPARQL queries.

TAGMA can be seen at work in figure 2. Using TAGMA GUI (see figure 6) the user is firstly prompted to insert the desired tags which is interpreted as “get all resources tagged as Banff”. The system then creates a SPARQL query based on TagOnt ontology [10] that matches this request as shown in figure 2. TAGMA analyzes the semantics of the query and rewrites the query in term of semantic APIs. Next, each semantic web service reinterprets the query in

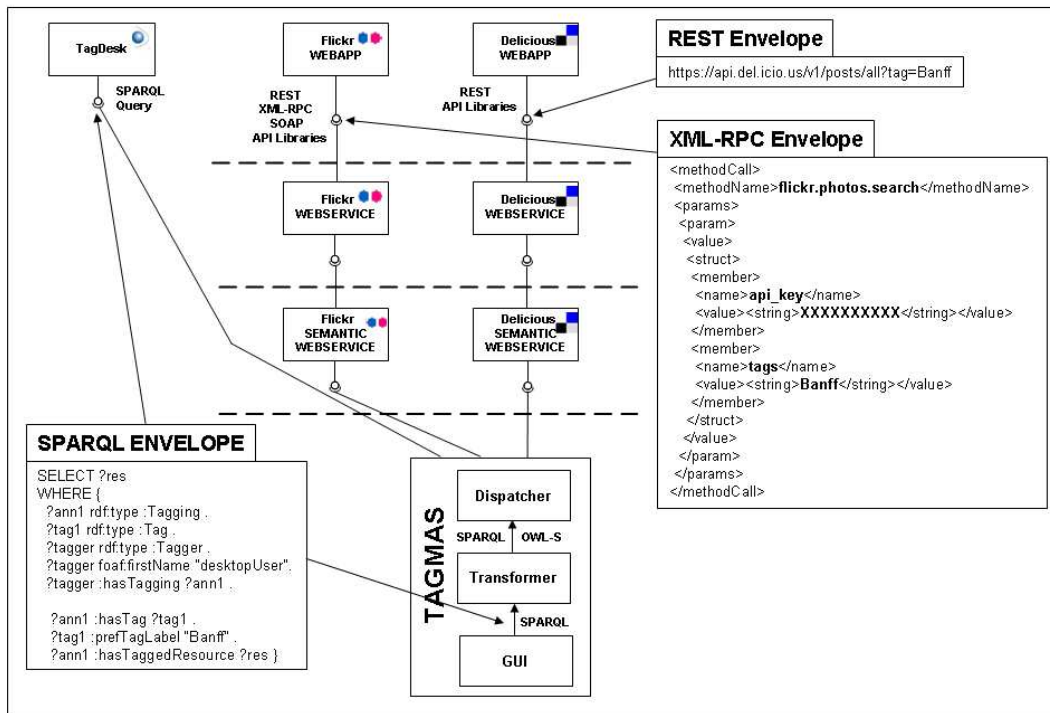


Figure 2: From SPARQL query to specific folksonomy queries.

SOAP terms, and invokes the corresponding web service. Finally, each requested web service is rewritten using the envelop format specific to each folkserver. An exception to this process is the resource manager at the desktop. Being local and SPARQL aware, the query can go straight to this resource manager with no transformation. In figure 2, the user query results into three requests: one to *delicious* using *REST* notation; a second one directed to *flickr* using *XML-RPC* envelopes; and finally, a third one that goes to the local resource manager with a direct petition.

3. USE CASES

This section outlines the main facilities currently provided by *TAGMA*, namely,

1. automatic tag creation, which permit to create desktop-specific tags,
2. folksonomy loading, which permit to import a folksonomy from a folkserver,
3. resource annotation, where a resource can be annotated along loaded folksonomies,
4. resource searching, where tag-based filtering is used to locate resources regardless of where the resource is hold. This facility is parameterized for the folkservers whose folksonomy has been downloaded into the desktop.

3.1 Automatic tag creation

TAGMA permits to tag resources based on the folder directory that keeps these resources. The user spends considerable time to

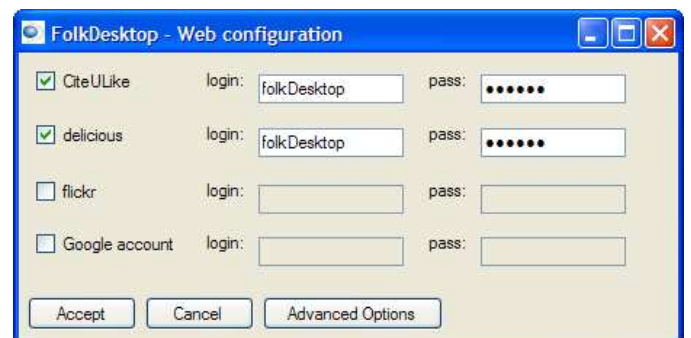


Figure 3: Configuring folk server parameters.

create sophisticated hierarchies to match his mental classification model that are barely utilized in the search process [6]. Usually the name of the folders gives semantic information about the files stored in, and the path of the file that can be used to tag the contained files. As an example, suppose "Conference\WWW\Photos" folder, where several digital pictures are stored as "DSC000*.JPG" files. If the user forgets the directory name, there is no way to retrieve pictures. Using automatically tagged tool, the user selects the desired folder, selects a *folkScan* option, and all the photos stored in this folder will be tagged with "Conference", "WWW" and "Photos" tags. In the future, if more pictures are stored in the folder, the system will automatically tag them with the folder path tags.

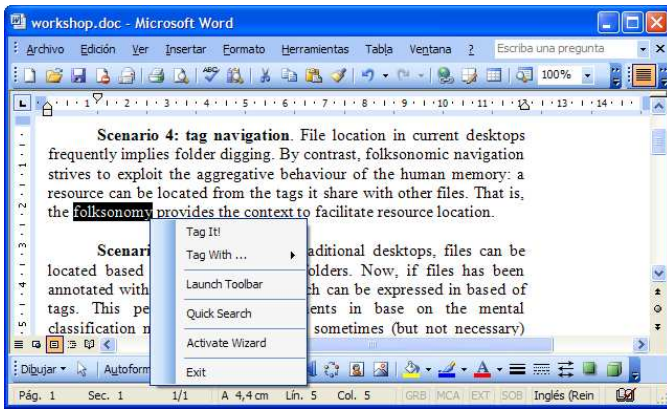


Figure 4: Annotating resources with content words.

3.2 Folksonomy loading

Besides local tags, *TAGMA* allows to import tags from folk-servers where the user has an account. In the example, the tags of *flickr* and *del.icio.us* are downloaded into the desktop. This permits to use these tags for annotating resources at the desktop. At annotation time, the user selects the tag set to be used (see figure 3).

To prevent the folksonomy stored at the desktop from becoming outdated with the one at the folkserver, in advanced options a refresh period can be specified whereby a downloading process is launched with the specified frequency.

3.3 Resource annotation

Desktop resources can be annotated with the tags imported from the folk-servers or automatically obtained from the folder structure. However, it is very common to create “spontaneous” tags as the resource is annotated. Traditional desktop offers this possibility where keywords can be associated with the resource. However, now these keywords belong to the desktop as a whole so that a tag created when annotating a resource, is available for annotating other resource’s.

TAGMA offers two ways for spontaneous tagging. The former case is based on the content of the resource itself (see Figure 4). The words of the file content are the first candidates to be tags of the resource. For a word to become a tag, the user selects the word, clicks on the middle button, and selects the “*Tag it*” action. The system also stores the most recent used tags in a cache which are displayed in a “*Tag with*” option of the menu.

The second option is to introduce manually the tag (see figure 5). Through the tag menu, the user selects the “*Launch toolbar*” option. The tags appearing in this window are associated with the resource which is loaded in the active application of the system. If the user moves to another window, the information of the tag window changes automatically. This window shows both the user defined tags (marked as user tags) and the system tag (i.e. those smartly inferred by the system). The bottom input box, permits to add new tags to the active file. To avoid create similar tags (e.g. plurals), when the user is editing the tag, the system displays all the existing tags that match the edited word. In the example of figure 5, the tags of the drop-down menu has been obtained from *del.icio.us* and *citeULike* folksonomies.

3.4 Resource searching

Once the files have been annotated, *TAGMA* can be used to find

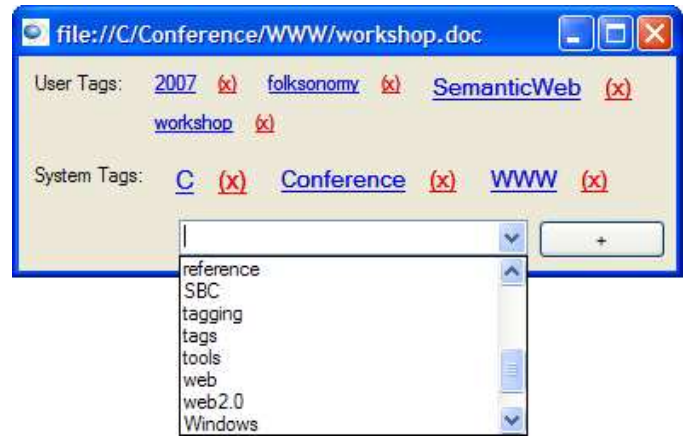


Figure 5: Resource Tag Toolbar.

documents along the folksonomy. The next two scenarios show how the user can exploit the semantic information of the resources through tag search and tag navigation.

Scenario 4: tag search (see figure 6). In traditional desktops, files can be located based on their names and folders. Now, if files have been annotated with a folksonomy the search can be expressed based on tags [9]. This permits to find a documents being based on the mental classification model of the user, that sometimes (but not necessarily) matches with folder tag classification hierarchy, and alleviates the problem of hierarchic classification of actual file systems, where documents can only be stored and located in one place. Figure 6 gives an example of this situation. Suppose that the user wants to find all resources related to *Bariff*. The localized resources has been obtained by different ways: (a) following its resource path (e.g. *c:\trips\bariff\hotelRegistration.pdf*), (b) through user tag annotation on local resources (e.g. *c:\Papers\www\2007\folkDesk.ppt*), (c) uploaded pictures to *flickr* (e.g. *DSC0001.jpg*), (d) annotated URLs in *del.icio.us* (e.g. *www2007.org*) or (e) edited documents in *google Docs* (*folkDesk.doc*). Note that the user can specify the folk-servers where the search will be done.

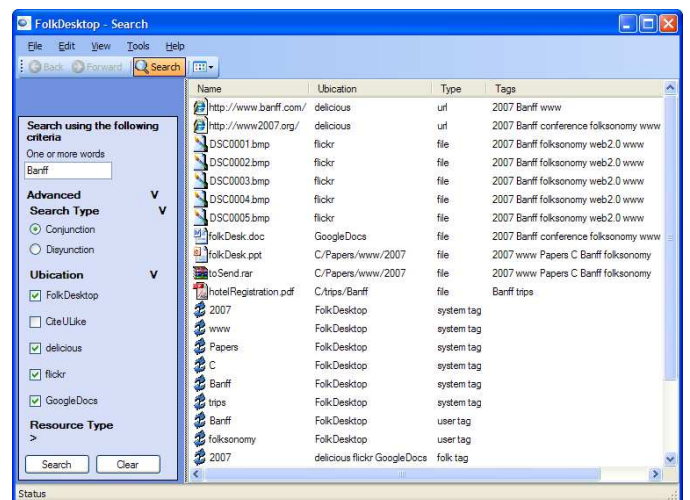


Figure 6: Tag search. The user searches resources using folksonomy tags.

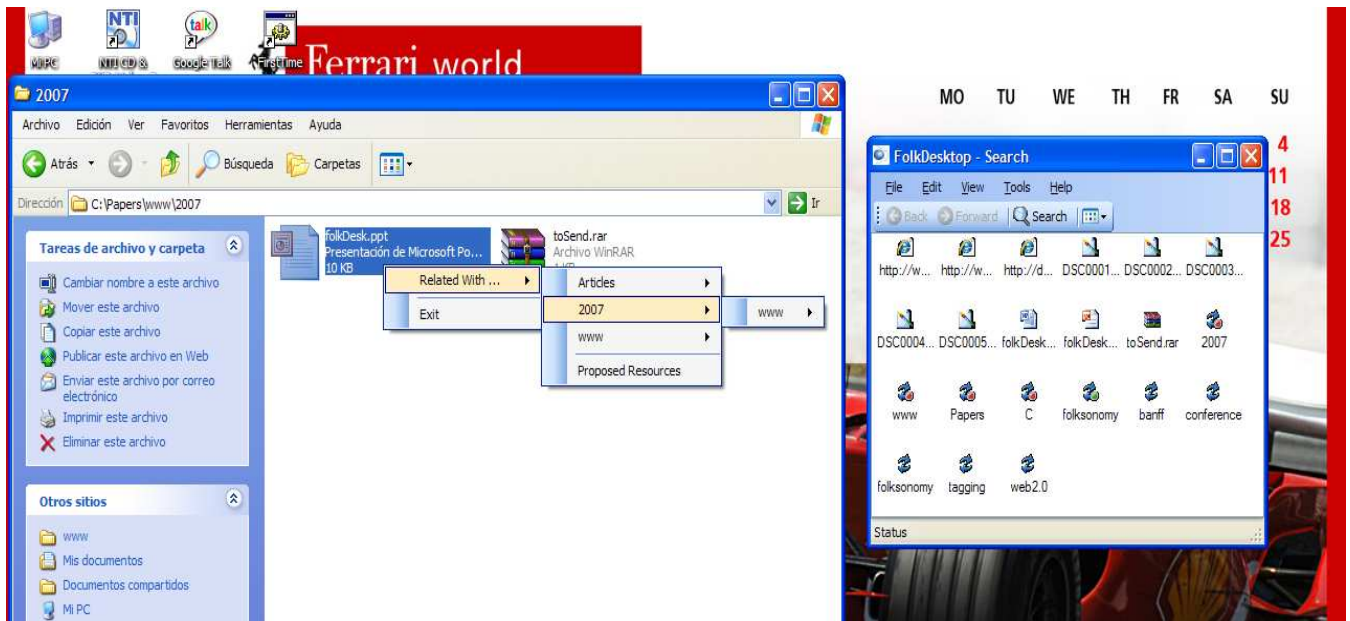


Figure 7: Tag navigation. The user navigates through folksonomy tags.

The result is the resources annotated with the queried tag. For each resource, a name, ubication (i.e. folkserver name or local path), type (i.e. file, URL or tag) and additional marked tags are displayed. Additionally to the resources, all the tags related with the localized resources are displayed, where its type indicates its origin (*user* created, *system* inferred or *folk* downloaded). Clicking on the tag, the search is redefined adding a selected tag in conjunction to the previous tags.

Scenario 5: tag navigation (see figure 7). File location in current desktops frequently implies folder digging. By contrast, folksonomic navigation strives to exploit the aggregative behavior of the human memory: a resource can be located from the tags it shares with other files. That is, the folksonomy provides the context to facilitate resource location.

As an example, consider the user is in the desktop and selects a file marked with *articles,2007,www* tags. By clicking the middle button, a menu pops up with all the existing tags ordered by the number of resources marked with this tag. If the user wants to refine the navigation, a second level set with the rest of tags appears. In the example of the figure 7, the user navigate to all the files tagged with *2007+www* tags. Note that the obtained document, must not necessarily be stored in the *2007/www* folder.

4. CONCLUSIONS

The success of Web2.0 is draining user's resources from the desktop to the Web. More resources are being uploaded to folk-servers for the sake of the community-ness. However, this leads to a fragmentary vision where resources of a given user are scattered all along the "folksphere". To overcome this situation, this paper promotes a federated vision of folkservers where the desktop plays the role of a federated system, and resource location is conducted along the integrated folkservers in a transparent way.

So far, *TAGMA* focuses on federated search and basic tagging. The next follows-on include direct uploading of resources from the desktop without moving to the folkserver, and facilities for user-based folksonomy integration.

5. REFERENCES

- [1] Citeulike home page, 2007. <http://www.citeulike.org/>.
- [2] Delicious home page, 2007. <http://del.icio.us/>.
- [3] Flickr home page, 2007. <http://www.flickr.com/>.
- [4] SPARQL Query Language for RDF, 2007. <http://www.w3.org/TR/rdf-sparql-query/>.
- [5] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
- [6] Paul Alexandru Chirita, Rita Gavriloaie, Stefania Ghita, Wolfgang Nejdl, and Raluca Paiu. Activity Based Metadata for Semantic Desktop Search. Technical report, L3S, December 2004.
- [7] Google. Google docs and spreadsheets, 2007. <http://docs.google.com>.
- [8] Tony Hammond, Timmo Hannay, Ben Lund, and Joanna Scott. Social Bookmarking Tools(i). *D-Lib Magazine*, 11(4), 2005.
- [9] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information Retrieval in Folksonomies: Search and Ranking. In *ESWC*, pages 411–426, 2006.
- [10] Torben Knerr. Tagging Ontology - Towards a Common Ontology for Folksonomies, 2007. <http://tagont.googlecode.com/files/TagOntPaper.pdf>.
- [11] Adam Mathes. Folksonomies - Cooperative Classification and Communication Through Shared Metadata, 2006. <http://www.adammathes.com/academic/computer-mediated-communication>.