
Sharing information across community portals with FOAFRealm

**Slawomir Grzonkowski*, Adam Gzella and
Sebastian Ryszard Kruk**

Digital Enterprise Research Institute
National University of Ireland
IDA Business Park
Lower Dangan, Galway, Ireland
E-mail: Slawomir.Grzonkowski@deri.org
E-mail: Adam.Gzella@deri.org
E-mail: Sebastian.Kruk@deri.org
*Corresponding author

John G. Breslin

Department of Electronic Engineering
and
Digital Enterprise Research Institute
National University of Ireland
IDA Business Park
Lower Dangan, Galway, Ireland
E-mail: John.Breslin@deri.org

Tomasz Woroniecki

Digital Enterprise Research Institute
National University of Ireland
IDA Business Park
Lower Dangan, Galway, Ireland
E-mail: Tomasz.Woroniecki@deri.org

Jarosław Dobrzański

Digital Enterprise Research Institute
National University of Ireland
IDA Business Park
Lower Dangan, Galway, Ireland
and
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology, Poland
E-mail: Jaroslaw.Dobrzanski@deri.org

Abstract: Community portals such as blogs, wikis and photo-sharing sites have become the new channels for information dissemination on the web. They contain a huge amount of valuable information that is often voluntarily delivered by experts. The most useful results of searching the web very often come from some sort of a community site. We present a method of sharing information across multiple community portals through a Social Semantic Collaborative Filtering (SSCF) system. It utilises FOAFRealm, a user profile management system which extends the popular 'Friend of a Friend' (FOAF) metadata and enables users to share the bookmarks and community documents that they create. Moreover, the proposed solution allows both a seamless connection of different portals and the easy identification of contributors.

We describe the required infrastructure, including the components that enable content sharing and browsing. Finally, we demonstrate a verification of our idea.

Keywords: online communities; information sharing; collaborative filtering; FOAFRealm; social semantic collaborative filtering; SSCF.

Reference to this paper should be made as follows: Grzonkowski, S., Gzella, A., Kruk, S.R., Breslin, J.G., Woroniecki, T. and Dobrzański, J. (2009) 'Sharing information across community portals with FOAFRealm', *Int. J. Web Based Communities*, Vol. 5, No. 3, pp.351–370.

Biographical notes: Slawomir Grzonkowski is a Researcher at Digital Enterprise Research Institute (DERI), National University of Ireland (NUI), Galway, Ireland. He is the leading person of the DRM laboratory within the applied innovation cluster in DERI Galway. He is the Creator and Main Developer of the SEcure DIgital Credentials (SeDiCi) and Fair Rights Management (FRM) projects. His research interests span many domains, mainly semantic web, security, digital rights and social networks. He has published over 20 scientific articles at international conferences, workshops and journals. He is also actively involved in the international community's research efforts, often serving as a Reviewer, PC member and Workshop Organiser.

Adam Gzella is a PhD Researcher/Research Assistant at the DERI, NUI, Galway, Ireland. He obtained a Master's in Computer Science at the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Poland (awarded with distinction) in 2006. He is a member of the Applied Innovation and Information Mining and Retrieval Units within DERI. He is experienced in both web and SemanticWeb technologies. He was developing the D-FOAF (distributed identity management) and SOA layer for this system and Social Semantic Collaborative Filtering (SSCF) component. Currently, he works in the Social Semantic Search and Browsing project, which is mainly on the notitio.us system. His research is focused on finding ways to utilise social networks and Web 2.0 in the SemanticWeb domain.

Sebastian Ryszard Kruk is the co-founder and CEO of KnowledgeHives.com, a Web 3.0 startup. Previously, he was a PhD student, Lead Researcher (Semantic Infrastructure Laboratory at the e-Learning Cluster of DERI, NUI, Galway), and Project Manager (Corrib.org) affiliated with DERI, NUI, Galway and the Gdansk University of Technology (GUT). His main areas of interest cover the semantic web and social networking technologies, digital libraries, knowledge management, information retrieval, security and distributed computing. He has published at international conferences, given tutorials and written a book on semantic digital libraries. He remains actively involved in the open source community.

John Breslin is currently a lecturer at the Department of Electronic Engineering of the NUI, Galway, Ireland. He is also an Associate Researcher and Leader of the Social Software Unit at the DERI of NUI, Galway, researching semantically enabled social networks and community portals. He is the founder of the SIOC project, which aims to semantically interlink online communities. He has received a number of awards for website design, including two Net Visionary awards from the Irish Internet Association for the Irish community website boards.ie, which he co-founded in 2000. He is a member of the IEI, IET and IEEE.

Tomasz Woroniecki received his MSc in Computer Science from Gdansk University of Technology, Poland in 2006. For two years, he was actively involved in research projects at DERI, NUI, Galway, Ireland. He co-authored a number of research papers related to social and semantic web technologies. Currently, he builds systems for the finance, telecom and insurance industries.

Jarosław Dobrzański is a Research Intern at DERI, NUI, Galway, Ireland. He is a member of the e-Learning Cluster. His research is related to the semantic web, e-learning and social networks. Currently, he is involved in the Didaskon project – the Automated Curriculum Composition based on the Work-flow Scheduling of the Semantically Annotated Learning Object Services. He is also the co-founder of the Informal Knowledge Harvester (IKHarvester), a tool for harvesting metadata for online communities' resources and exposing them for e-learning frameworks. More information about him and his research is available at <http://dobrzanski.net>. He remains actively involved in web and database technologies.

1 Introduction

The semantic web (Berners-Lee *et al.*, 2001) is increasingly aimed at new applications areas. Community portals and social networks (Adamic *et al.*, 2003) are highly visible targets for its research. Due to the popularity of community-based applications such as blogs and wikis, they have experienced rapid dynamic growth utilising interconnected information spaces through the *blogosphere* and *interwiki* links. In addition, more and more content is being created by their users, ranging from pictures on photo album-sharing sites to bookmarks on topics of interest. At the same time, these applications are experiencing boundaries in terms of information dissemination and users' profile automation.

A very simple example of a query that cannot be easily answered by today's search engines is 'show me all the content created by Mick and all his close friends in the past week'. A lot of documents created on community portals can be found with such search engines. However, the obtained results lack the level of provenance typical for information acquired using filters based on a user's social network. Community portals have much more to offer than just document publishing systems and what can be found with search engines is just the tip of the iceberg. Because of semantic technologies, we are able now to benefit from connections between such online information sources.

We begin the paper with an explanation of what a community portal is; we also detail some of the problems that such portals have. We then give an overview of the components that are necessary to solve them. Our main solution uses a standards-aware

technology called FOAFRealm, a user profile and relationships management system; it operates over D-FOAF, a distributed Peer-to-Peer (P2P) authentication and trust infrastructure. D-FOAF is designed to operate without a centralised authority. We also describe the related work in this area. We then present the profile management and social bookmarking systems in detail, followed by a description of how documents and resources can be classified and exchanged amongst the members of a distributed network community. Then, we present the informal knowledge harvester and faceted navigation solutions. They can be used to gather the content of the bookmarked resources and easily browse through the harvested and manually acquired resources. Finally, we conclude our research and outline some plans for future work based on our results so far in connecting the content of users and their social networks across multiple community sites.

2 Community portals

Community portals are websites that provide improved communication and contact links for its users (*e.g.*, one that provides local or interest-based information). Such portals are the most widespread platforms used by communities to stay informed electronically. The members can find relevant information and may contribute any required shared information to others via the portal. By having online collaboration spaces for a community of a certain interest, community portals provide an awareness and interaction amongst a set of people, whether for profit or nonprofit. These portals are replacing the traditional means of information exchange. They help provide an online global communication agora and strengthen the communities themselves by keeping them informed and providing an open place for interaction and the exchange of information and ideas.

2.1 Problems with existing portals

Each portal usually has its own user management system. Therefore, users are forced to create accounts and then remember sets of credentials, although they present almost exactly the same information at each of the portals that they register to (such as their personal details, domains of interest, *etc.*). At different community portals, a community member may store different information related to a particular domain of interest. This information could be copied into one place and then merged. It is time consuming and not every user would decide to perform such an operation.

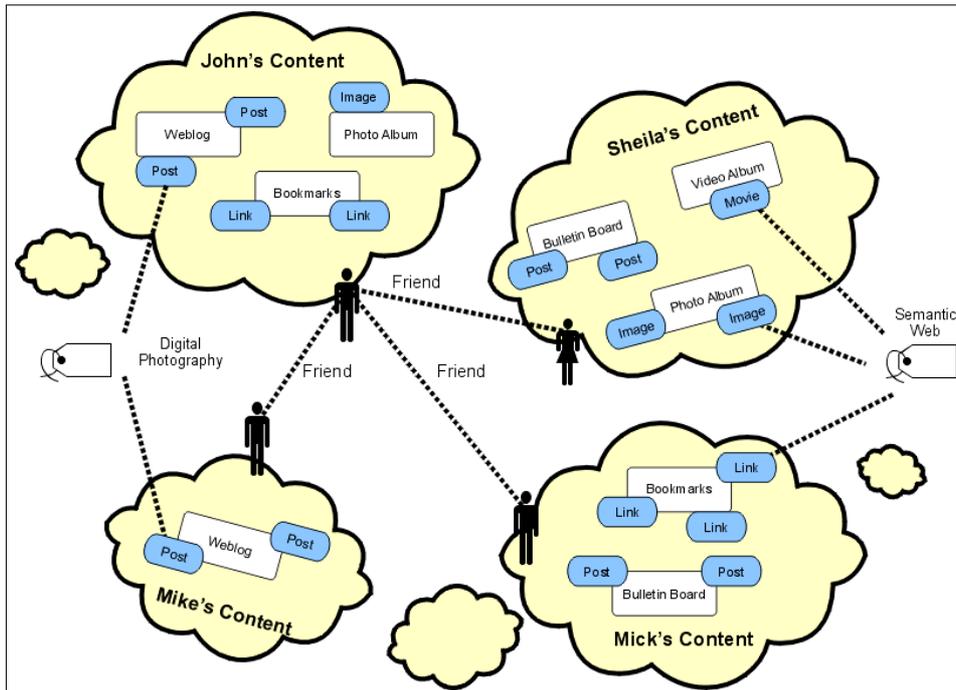
Another problem occurs when a user wants to gain some knowledge gathered by an expert or a friend. Unfortunately, the resources maintained by those people can be located in many different portals; hence, the users often waste their time on importing these resources manually; they may even abandon the operation in favour of using other, less competent knowledge.

To conclude, we note that community portals have many significant advantages over traditional online collaboration methods (*e.g.*, newsletters). They can be very useful and helpful for the users of the portals. Moreover, the users of classical community portals are facing many potential problems.

2.2 Solution scenario

Figure 1 illustrates a scenario which requires a solution to the problems stated in Section 2.1. Let us consider a case where we have a number of people who know each other and are interested in more than just one topic, *e.g.*, the semantic web and digital photography.

Figure 1 Sharing content between disparate community sites using a social network and topics of interest (see online version for colours)



We will begin by looking at John and the content that could represent his membership in various online community portals. John is interested in digital photography and he is a member of a blog site and a photo-sharing community. John also has three friends (Mick, Mike and Sheila); some of them are registered in the same communities as he is. He wants to sign up for a collaborative bookmarking site so that he can start bookmarking some of his favourite links relating to photos and annotating digital photos with metadata, a new interest. Moreover, he knows that his friend Mick is a member of the bookmarking site; John hopes to use some of Mick's expertise in the semantic web and metadata to help with his search for useful resources.

Unfortunately, he cannot use either of his accounts on the blog or photo-sharing communities to login to the new bookmarking site, so he will need to register a new account for that portal.

What is more, if there are any interesting posts on digital photography and annotations in the blog or photo-sharing communities, he will need to copy and paste the links to all the relevant discussions to the new bookmarking site because there is no common exchange mechanism for resources or resource links between the various communities.

John has also noticed that some of the community portals allows the extensive annotation of provided information using tags, controlled vocabulary or relations between resources; the annotations, however, are useless since the community portals offer only simple search or tag-based (tagcloud) browsing. Therefore, he is unable to make use of these annotations.

We claim that if the aforementioned sites were connected using a P2P-based distributed user profile and relationship management system (FOAFRealm (Kruk, 2004) via D-FOAF (Kruk *et al.*, 2006b)), then John could use Social Semantic Collaborative Filtering (SSCF) (Kruk and Decker, 2005) to pass links to items of interest from the photo-sharing site to the collaborative bookmarking site (or to any of his friends, *e.g.*, for Mike to use on his blog). He could also simply use SSCF to bookmark any interesting items under a category folder called, for example, *Digital Photo Semantic Annotations*, and then refer to this folder from any of the communities to which he is registered. Since the underlying information is heavily interconnected, John can use faceted navigation techniques to browse through the data provided by various members of the community.

Furthermore, if a user specifies his/her topics of interest on one site, then these can be used to match other resources (discussions, pages, *etc.*) matching those topics of interest on any other site they register for.

For example, Sheila, who is registered on a bulletin board site (for semantic web developers), says that she has interested in resources tagged or categorised as *Semantic Web* and *IPTV*. She registers (via FOAFRealm) on another site (a video-sharing site), the site picks up information from her profile that says she is interested in *Semantic Web* and *IPTV* and presents her with the resources linked to those topics. One of the videos is about using semantic web technologies to provide an enhanced programme guide for television over data networks and is tagged as being related to *Semantic TV*. She marks it as a topic of her interest. Then, on the original bulletin board site, more resources (matching *Semantic TV*) are presented.

3 Background-related work

3.1 Online communities

Online communities have become more and more popular and they can no longer be considered niche systems. Each country and business trade has at least several popular portals. Hildreth and Kimble (2000) proposed the following definition for a community:

“It has a common set of interests to do something in common, is concerned with motivation, is self-generating, is self-selecting, is not necessarily co-located, and has a common set of interests motivated to a pattern of work not directed to it.”

Additionally, Kondratova and Goldfarb (2003) distinguished the three main objectives of online communities:

- 1 to supply content to the users
- 2 to encourage members to participate in the community by contributing knowledge
- 3 to facilitate communication and interaction between the members.

The key features that differentiate online communities are the various kinds of forums, wikis and chat rooms, as well as the online and offline events that they may have (*e.g.*, the semantic web community portal¹ uses a mailing list as its primary communication medium). It is difficult to say which type of community portal is the most popular because there are as many portals as interests. One potential candidate could be Wikipedia,² since it gathers people regardless of interests.

3.1.1 Semantically interlinked online communities

Semantically Interlinked Online Communities (SIOC)³ is an initiative whose goal is to interconnect online communities (Breslin *et al.*, 2006). SIOC can be used in published or subscribed mechanisms, as it stores community-like metadata such as information about the post's author, enclosed links, the creation time and connection with other webpages. Thanks to SIOC, each part of the community portal, blog, forum, *etc.*, can be semantically described using a unified vocabulary. Thus, the applications that utilise SIOC could easily search, exchange and link the content created by communities.

The core of the SIOC framework is the SIOC ontology, which consists of a set of classes and properties which link them:

- Site – is the location of an online community or set of communities.
- Forum – is a discussion area housed in a site.
- Post – can be formed as an article, a message or an audio- or videoclip. A post is written by an author, has a topic, content, external links, *etc.*
- User – represents an account held by an online community member.
- Usergroup – is a set of accounts of users interested in a common subject matter.

3.2 Social networks

The aim of the Friend of a Friend (FOAF) (Brickley and Miller, 2005) standard is to utilise machine-readable homepages for describing people. Moreover, the idea also proposes storing links between people and the activities that they take part in (*i.e.*, by specifying their topics of interest). The FOAF vocabulary⁴ was introduced by Brickley and Miller. It is strongly dependent on World Wide Web Consortium (W3C)⁵ standards, especially Resource Description Framework (RDF) and eXtensible Markup Language (XML). A number of applications have been developed that make use of the metadata provided using this vocabulary: FOAF-A-Matic⁶ allows those not familiar with XML to easily create people descriptions and FOAFNaut⁷ provides a visualisation of any social network formed using FOAF user profiles.

Additionally, many online social network sites have taken advantage of Milgram's (1967) famous 'six degrees of separation' experiment. Friendster (Boyd, 2004) was initiated in 2002 and has received some patents in this domain. Furthermore, there has been a large growth of business-oriented networks. LinkedIn⁸ and Ryze⁹ manage

professional contacts and enable users to find an employer or an employee. A special issue of *Complexity* published in August 2002 considered the role of networks and social network dynamics (Skvoretz, 2002). The aims of the issue were to show the complexity for different levels of network architecture and help with the comprehension of network-based analyses and explanations.

More and more people are interested in joining online social networks; for example, WLSpaces¹⁰ has already gathered more than 120 million people. The portal offers a very simple social network model that allows specifying only one type of relationship between users: friend. WLSpaces owes its huge popularity to strong connections with Windows and the well-designed user interface.

Another large portal, Xanga,¹¹ has gathered 40 million users. The offered functionality is similar to WLSpaces. The applied social network model is also limited to the 'friend' relationship.

In the given examples, the number of users is comparable to the number of people in large countries and, thus, we note the importance of online social network implementations.

3.3 *User profile management systems*

Many research projects, ranging from open source to commercial, have been proposed to deal with user profile management. Some of them, like Drupal¹² and XML User Profiles (XUP)¹³ (the latter being similar to the W3C FOAF metadata recommendation), offer sophisticated features such as distributed profiles and Single Sign-On functionality.

The Identity 2.0¹⁴ protocol was proposed for the exchange of digital identity information. The general idea entails that users be supported with enhanced control over the information entrusted to the other members of the community. The authors of Sxip 2.0 (Hardt, 2004) announced that their system implements the protocol. In addition, it makes it possible to adjust security needs to a specific site.

Probably the most famous profile management system is Microsoft Passport¹⁵ and its predecessor, CardSpace.¹⁶ Both systems support the reuse of profile information across different services. Although it is an interesting idea, frequent bug reports and the centralised topology mean that the system has not yet been commonly accepted by other sites.

Another example of a decentralised digital identity system is OpenID,¹⁷ in which users' online identities are given either by Uniform Resource Identifiers (URI) (such as for their blog or a homepage) or OASIS Extensible Resource Identifier (XRIs).¹⁸

3.4 *Faceted navigation*

Managing unstructured information, like that generated by online communities, can pose many problems, which arise either from the lack of predefined controlled vocabulary or from relations between pieces of information that are hard to query using keyword-based techniques.

In the first case, solutions like folksonomies (Mika, 2005) or SSCF (Kruk and Decker, 2005) can give a hint on the relations between different concepts in the community-driven vocabulary. The information about relations between concepts delivered by the community can be represented as a taxonomy; such a taxonomy is either given explicitly for a certain user, *e.g.*, in SSCF, or is automatically created with

clustering algorithms, *e.g.*, in folksonomies. Yee *et al.* (2003) has constructed a user interface paradigm called ‘faceted navigation’ to browse the information categorised with concepts with taxonomies.

In the latter case, when community annotations involve relations between information sources, a sole faceted navigation is not enough; usually, the community delivers the relations, so it is difficult to manually create a navigation interface in advance. Furthermore, to fully exploit the relations between information sources, actions such as finding similar resources or browsing through the graph of interconnected information are required. An example of the first one would be to find other posts with tags similar to the post currently being viewed; the latter one would be to browse from a list of posts to a list of their authors to the list of sites they post to. Solutions like BrowseRDF (Oren *et al.*, 2006) or MultiBeeBrowse (Kruk *et al.*, 2007a) have already been proposed to allow navigation on interconnected information sources.

4 Design components

In this section, we describe the components that are crucial from the perspective of satisfying the requirements of the given scenario (see Section 2.1).

4.1 Distributed user profile management

The transparent distribution of a user’s profile offered by FOAFRealm fits very well with the requirement for interportal cooperation. Information about a user’s preferences may be collected from all connected sites and the modifications made in any site are visible across the others. For example, when the user subscribes to the ‘Annotating Images’ category on a visited portal, a new fragment of his personal profile is created and propagated to other sites. Eventually, every site sees the user’s profile as a union of all fragments on the rest of sites. FOAFRealm hides the complexity of managing distributed data and offers a clean interface for querying and storing users’ profiles.

4.2 D-FOAF: architecture of a distributed user profile management

FOAFRealm implements and extends the metastandard proposed by FOAF. The extension made it possible to use the standard as a user profile management system. The prototype implementation is delivered for the Tomcat JSP container.¹⁹

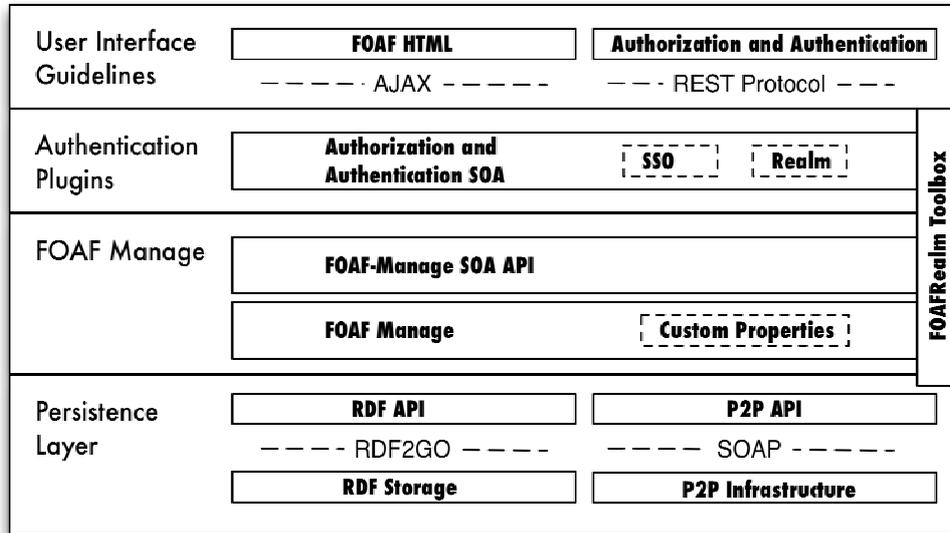
The D-FOAF project is a distributed implementation of the idea introduced by the FOAFRealm project. Therefore, the project is crucial from the perspective of sharing profile information across multiple community portals.

The current implementation comprises four layers (see Figure 2):

- 1 user interface guidelines
- 2 authentication plugins
- 3 the main component that delivers the implementation of the Dijkstra algorithm for quantifying distance and friendship and manages the lower layers
- 4 the bottom layer that provides access to the P2P infrastructure and an RDF repository

In addition, there is also the toolbox module that spans all the layers.

Figure 2 The FOAFRealm architecture



4.2.1 User interface guidelines

This layer is responsible for providing user interface solutions. The current implementation delivers an asynchronous solution based on Asynchronous JavaScript and XML (AJAX).

The FOAF Hypertext Markup Language (HTML) component delivers a set of files responsible for delivering end-user functionality. The component consists of servlets, Java beans, server-side configuration files, JSP files, tags, JavaScript codes and styles. Authorisation and authentication operations are performed every time a user tries to access any of the protected contents or if a user logs in to the system. The operations are requested by an authentication plugin and they are passed to the FOAF Manage module. The required computations are performed on a social network digraph.

4.2.2 Authentication plugins

Currently, there are several plugins that allow using the FOAFRealm in existing web applications. The plugin that offers the richest functionality was designed for Tomcat. The architecture was designed in the Service-Oriented Architecture (SOA) manner and, thus, developers are able to easily adapt the system to their needs.

4.2.3 FOAF Manage

The middleware of the system is provided by the FOAF Manage module. The layer provides both the functionality to the upper SOA component and to two bottom layers: the P2P network (see Section 4.2.5) and system persistence (see Section 4.2.4). Most functionalities of the system is configured at this layer.

Furthermore, the FOAF Manage module delivers an implementation of a distributed Dijkstra algorithm that is crucial to calculate the friendship values in the provided social network.

4.2.4 Persistence layer

The persistence layer is based on RDF triples. The content stored in the system consists of users' profiles and their social relationships. The RDF Application Programming Interface (API) delivers the means to manipulate users' profiles, including their social relationships. This module does not use any particular RDF storage, but invokes the RDF2GO²⁰ library. Therefore, access to the storage is transparent from the system perspective and that storage can be easily replaced with another one. The RDF Storage module is the repository that the system uses. At the moment, Sesame²¹ is used, which was chosen before applying the RDF2GO library. Currently, the repository can be replaced with another one by means of the FOAFRealm configuration.

4.2.5 P2P network

The system is distributed and uses a highly effective hypercube-based social network. Different instances of the system can be connected by means of the HyperCuP (Schlosser *et al.*, 2003) protocol. The P2P API takes advantage of the HyperCuP Lightweight Implementation (Grzonkowski and Kruk, 2006). The main benefit from this network is the efficient broadcast implementation; it is extensively used by single registration and Single Sign-On features.

4.2.6 FOAFRealm toolbox

The toolbox module is a helper component that provides common functionalities, for example, profile registration and secure connections. The component is accessible both by authentication plugins, FOAF Manage and the bottom layer.

4.3 Social semantic collaborative filtering

SSCF (Kruk *et al.*, 2007b) is based on two concepts: distributed collections and the annotations of resources. Each user classifies only a small subset of knowledge based on the level of expertise they have on a specific topic. This knowledge is later shared across the social network.

4.3.1 Classifying community portal documents

During their online activity, users can bookmark some resources. Unfortunately, such information needs to be properly classified to be used by the system; the SSCF module allows its users to classify their bookmarks as *domains of interest*, which are represented by semantically annotated catalogues. Domains contain bookmarks and may also include other domains. This structure needs to be well classified; the user's taxonomy of catalogues needs to refer to other knowledge organisation systems. SSCF can utilise well-known classification systems with the Java Binding for Ontologies, Taxonomies and Thesauri (JOnto)²² plugin; a user can annotate a catalogue's content using, *e.g.*, the Dewey Decimal Classification (DDC) system, WordNet and the Open Directory Project (dmoz):

- DDC²³ is a general knowledge organisation tool that is continuously revised to keep up with knowledge. DDC is currently the world's most widely used library classification system. Libraries in more than 135 countries use DDC to organise and

provide access to their collections and DDC numbers are featured in the national bibliographies of more than 60 countries. DDC provides a structural hierarchy, which means that all topics (aside from the ten main classes) are part of the broader topics above them. The class of a resource is shown by a decimal number with at least three digits. The first digit is the main class (for example, '500' represents science). The second digit indicates the division (for example, '500' is used for general works on the sciences, '510' for mathematics). The third digit indicates the section ('530' is used for general works on physics, '531' for classical mechanics). A dot follows the third digit in a class number, after which division by ten continues to the specific degree of classification needed.

- WordNet²⁴ is an online lexical reference system whose design is inspired by the current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organised into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets. Currently, the WordNet database consists of over 200 000 word-sense pairs (over 150 000 unique strings).
- dmoz²⁵ is the most widely distributed database of web content classified by humans. Its editorial standards body of netizens provides the collective brain behind resource discovery on the web. dmoz powers the core directory services portal for the web's largest and most popular search engines. All dmoz resources (structure and content) are freely available for use.

With SSCF, different parts of community portals can be classified using the methods described above. A user can easily assign a class to discussions, wiki pages, blogs and photo albums, as well as normal pages on the web. Moreover, the JOnto plugin was developed in a very flexible manner; hence, it will also support taxonomies that are not available yet.

4.3.2 Mechanism for exchanging documents between people

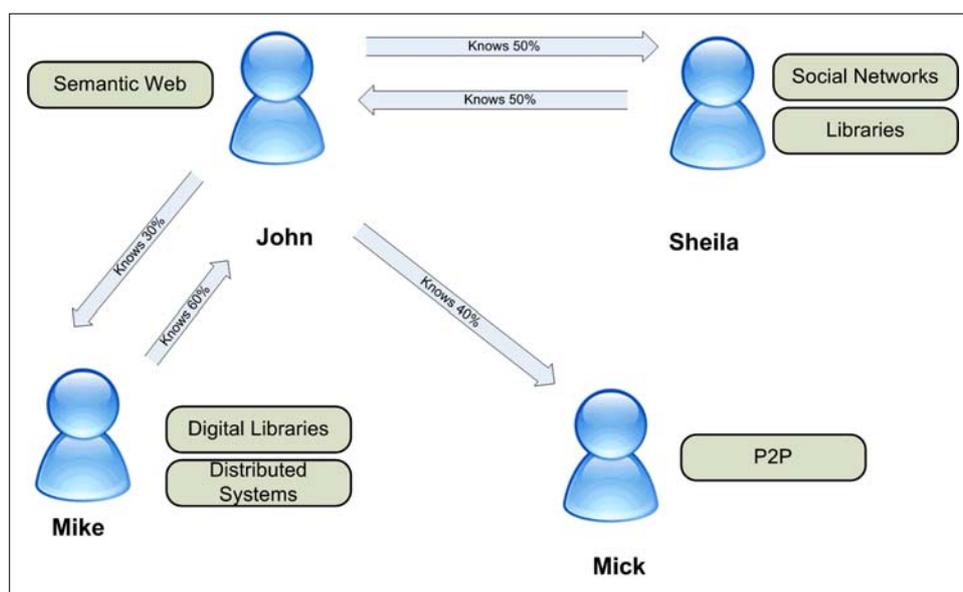
SSCF is strongly dependent on the social network, which can be stored as a directed graph. Nodes describe users, whereas edges represent the relationships between them (see Figure 3). Additionally, to overcome security problems, each link between two people can also have an assigned trust level that decides whether access should be granted or denied. Users can have collections of bookmarks (*i.e.*, a private bookshelf as described by Kruk *et al.* (2005)) which represent their knowledge; later, they can render this knowledge accessible to their friends. Resources are collected in the private bookshelf according to the user's point of view, as expressed by their categories taxonomy.

Each collection can be ranked with the quality metrics assigned to it. These quality metrics express the expertise level of the owner on the particular topic. This information can be computed with the PageRank algorithm (Brin and Page, 1998) applied to graphs of collection inclusions and the social network. Moreover, users are aware of the expertise level of some of their friends; this information can be used while looking for resources. Usually, the resources that belong to close friends, who are experts on a given topic, are potentially useful and reliable.

In sum, such an infrastructure provides an excellent environment for obtaining shared documents. The presented approach differs in many ways from present trends; sharing files via current P2P standards usually only depends on a number of free slots

or the quantity of shared files. Furthermore, SSCF allows users to specify access control policies to each catalogue; they can restrict access to a certain subgraph of a social network.

Figure 3 The accessibility of knowledge in SSCF (see online version for colours)



4.4 Capturing metadata for community documents

Informal Knowledge Harvester (IKHarvester) is an online service that harvests metadata for online communities and delivers them in a formal way. SSCF employs IKHarvester in such a way that every time a user bookmarks a webpage with SSCF, IKHarvester tries to capture some data. If it succeeds, it saves the data to a shared repository.

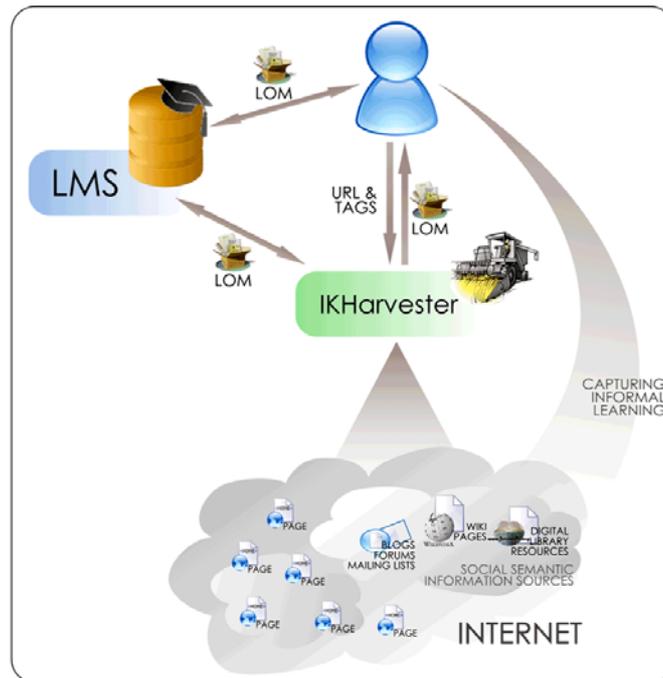
The current version of IKHarvester tracks (semantic) blogs, (semantic) wikis and semantic digital libraries. In general, IKHarvester looks for RDF documents related to a webpage with a specified URL. This relation is described in the HTML source code of such a page. Besides reading pure RDF data, IKHarvester uses microformats which allow embedding RDF into HTML documents. Moreover, IKHarvester is capable of creating RDF descriptions for nonsemantic information sources like Wikipedia. For that reason, it parses the HTML code of an article to collect some data from it, including a title and external links (see References and Notes).

To describe blogs and wikis, IKHarvester utilises SIOC. Semantic blogs allow us to acquire metadata in SIOC by using SIOC exporters.²⁶ For nonsemantic blogs and wikis, IKHarvester creates semantic annotations in SIOC from the captured information about the resources.

Beside the raw RDF, IKHarvester can provide the gathered information in various formats. An example has been presented in Figure 4. It shows how the system can be used as an informal knowledge provider for e-learning frameworks. The collected data

can describe the learning material captured from various online communities. Hence, semantic annotations can be easily transformed into Learning Object Metadata (LOM) (Brase *et al.*, 2003) that are used by many Learning Management Systems (LMSs).

Figure 4 The IKHarvester (see online version for colours)



4.5 Faceted navigation on unstructured metadata

Another important component for accessing interconnected information from various community portals is MultiBeeBrowse, a faceted navigation on unstructured metadata (which could be delivered by IKHarvester); this service delivers a solution to both problems stated in Section 2.4. It consists of two components: SOA based on Representational State Transfer (REST) services²⁷ and an AJAX user interface.

- 1 *REST-based SOA* – Simple Object Access Protocol (SOAP) is inadequate for implementing web services for semantic web applications; an argument for REST solutions, in the context of the MultiBeeBrowse service, is that the GET action defines an idempotent request, *i.e.*, the subsequent calls of the same Uniform Resource Locator (URL) should return the same results. Therefore, SOA for MultiBeeBrowse (MBB) is based on REST; a chain of browsing actions, as well as browsing history, is represented with a unique URL. This URL can be exchanged later between users; hence, MBB allows browsing collaboratively.
- 2 *Context-zoomable interface* – MBB conveys end queries with hexagon lozenges; a history of user actions is no longer only linear and a user can choose from up to five paths to continue or replay his/her browsing. Four user interface views support users in browsing: the classic view (a list of results), hexagon view (access to

the multidimensional history), honeycomb view (a 3D visualisation of lozenges representing end queries) and context view (access to the browsing history of each user). Switching from the classic view to the context view, a user zooms out from the end query to the history of all end queries previously performed.

In addition to the browsing features, MultiBeeBrowse allows users to share their browsing experience with their friends; the SSCF component, built on the top of FOAFRealm, has been extended to deliver, in combination with MultiBeeBrowse, a *collaborative browsing* experience. Users can bookmark their browsing paths, share them with their friends and refine each other's queries; these all leads to new experience in browsing supported by the community.

5 Implementation and evaluation

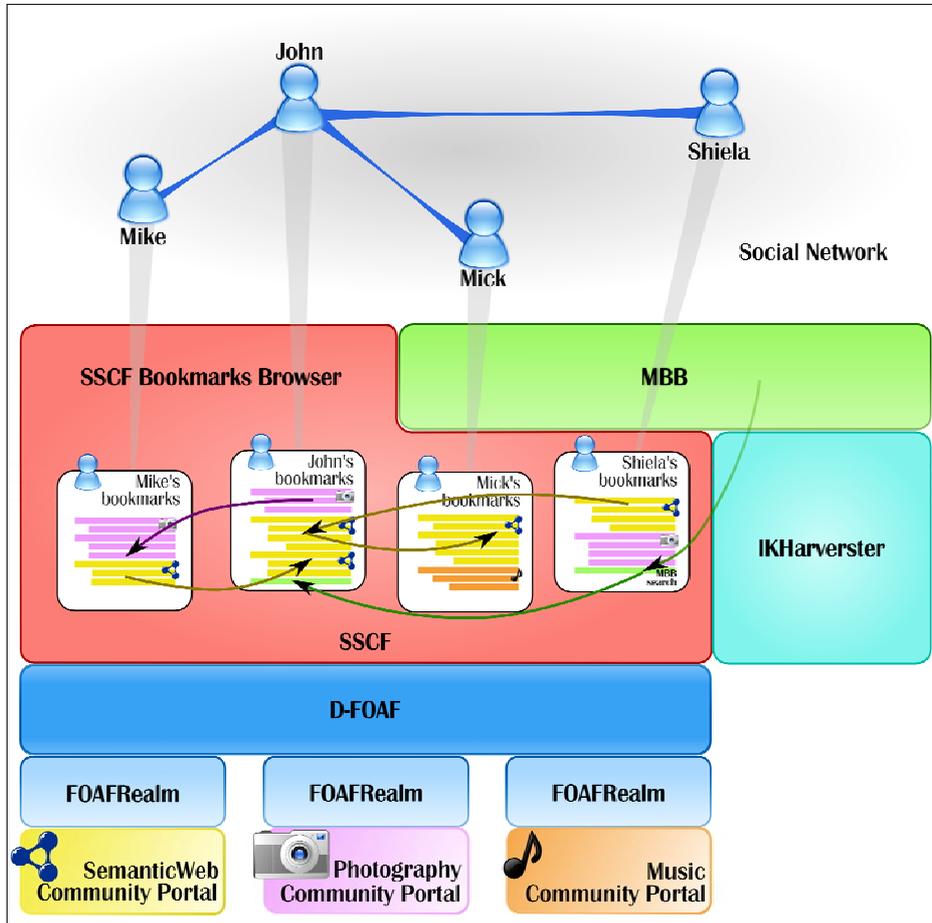
5.1 All the parts coming together

To solve the problems stated in the paper, we have combined all the presented components. To better explain how our solution works, we show an example (see Figure 5) based on our scenario (see Section 2.2). At the top, we can see the social network. John is a friend of Mike, Mick and Sheila. This information comes from the FOAFRealm/D-FOAF layers. At the bottom of the figure, there are three community portals (SemanticWeb, Photography and Music portals). Each of them is using the FOAFRealm/D-FOAF solution.

Each person has his/her own collection of bookmarks (see the middle portion of the figure). According to our scenario, John is the member of the Photography portal. But he knows Mike and he wants to use some of his expertise in the topic of *SemanticWeb*. Since the SemanticWeb portal (which Mike is using) is FOAFRealm-enabled, he does not need to create a new account there to utilise its features. Moreover, with SSCF, he simply browses the interesting and valuable pages bookmarked by Mike and copies some of the content into his own bookmarks structure. Such a process is continued between the other friends on other community portals; knowledge is shared between them. For example, Sheila has some resources about SemanticWeb saved and John imported this information into his own set. Then Mick, who was browsing John's bookmarks, found the same collection interesting and imported it into his own structure. As a result, despite the fact that Mick is not connected to Sheila, he was able to use her knowledge in the *SemanticWeb* domain.

When John and his friends were browsing the web and adding new bookmarks from the community portals into SSCF, IKHarvester gathered all the information from those pages in the background. The MBB interface, which uses the information from SSCF and IKHarvester, can now be used by each person to browse the data and find valuable information. The users have two options for browsing and discovering resources. They can either browse their friends' bookmarks hierarchies or browse the data with the MBB interface. The element that joins those two interfaces is the aforementioned *collaborative browsing*. Sheila constructed some complicated multifaceted query in the MBB system, which returns an interesting set of resources. She is able to save this process as an SSCF bookmark in her SSCF collection. John imports such a resource and does the same query again later. He can even refine and save it as a new bookmark, which can again be spread throughout the social network.

Figure 5 All the parts coming together (see online version for colours)



5.2 Scenario validation

In Section 2, we have stated several problems that are important for existing community portals:

- Most of the portals have their own user management systems.
- The users have to create multiple accounts and credentials.
- Although we trust our friends and experts, we cannot take advantage of their experience.
- Knowledge is distributed among many independent portals.
- Gathering knowledge is time-consuming.

We claim that by using FOAFRealm, we solve the first two problems. With its advanced distributed model and collaboration features, the system aims to be a complex solution for managing identities and preferences. Single Sign-On and single registration

let users comfortably use multiple services and helps them start up new sites by connecting them with the existing popular ones. The FOAFRealm library provides the authentication and social network features that can be easily incorporated into a portal or any other application.

SSCF is an application that utilises FOAFRealm. With fine-grained access control lists, it is easy to share resources among friends and spread knowledge in a community. Browsing others' bookmarks hierarchy gives users the benefit of using valuable resources collected by other possible domain experts. Therefore, SSCF partially solves the third and fourth problem.

MBB-faceted navigation, together with data gathered by informal knowledge harvesting (performed by IKHarvester) incorporated into SSCF, offers even more features to its users. They could not only browse through bookmarks, but also through the content of bookmarked resources. *Collaborative browsing* (SSCF bookmarks from MBB browsing paths) is yet another way to utilise expertise in a social network. Hence, those two components are also crucial from the perspective of the third and fourth problem.

Finally, IKHarvester is a service that aims to make knowledge gathering less time-consuming.

6 Conclusions

We detailed many issues with community portals that are experiencing boundaries in terms of content dissemination and profile automation; for example, users have to repeatedly sign up for various community sites and they cannot make use of their stored resource links or annotations between sites. Similarly, users cannot easily make use of their social networks between sites, for example, to leverage the skills of a friend who may be an expert in one domain on a different community site. We have described the FOAFRealm and D-FOAF implementations that are low-level infrastructures for our solutions. They have overcome many boundaries by providing a distributed user profile management system along with SSCF. We have described three other components based on the FOAFRealm infrastructure: SSCF, IKHarvester and MultiBeeBrowse. IKHarvester allows a community of users to easily discover knowledge from other community portals, such as wikis, blogs and fora. SSCF provides an excellent method of sharing resources between friends or associates by defining the level of expertise that a person has in a certain domain (according to those they are connected to via a social network) and by suggesting various resources based on these expertise levels. Finally, MultiBeeBrowse allows navigating, with the help of other community members, across the information gathered from various community sites.

7 Future work

As a next step of the evolution, we have already initiated a research project called DigiMe (Kruk *et al.*, 2006a). The aim of DigiMe is to deliver SSCF features that will support mobile devices and provide users with better control over their profile information. Thus, users would be able to store collaborative resources and profile

information on their mobile devices; the DigiMe system uses this information to explore the *ad hoc* social networks paradigm. Also, further research on the accessibility of the browsing component will be conducted to target a broader group of community members.

Moreover, one of the places where the presented solutions have already come together is the *notitio.us*²⁸ service. It uses FOAFRealm (for user management), SSCF (bookmarking web resources and MBB browser paths), IKHarvester and MBB.

8 Acknowledgements

This material is based upon works supported by Enterprise Ireland under Grant No. *ILP/05/203* and by Science Foundation Ireland (SFI) under the DERI-Lion project (SFI/02/CE1/1131). Three first authors are listed in alphabetical order. The authors would like to acknowledge Siobhán Dervan, S.P. Williams and the corrib.org working group for their help and fruitful discussions.

References

- Adamic, L.A., Buyukkokten, O. and Adar, E. (2003) 'A social network caught in the web', Vol. 8, No. 6, http://firstmonday.org/issues/issue8_6/adamic/.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The semantic web', *Scientific American*, May.
- Boyd, D.M. (2004) 'Friendster and publicly articulated social networking', Paper presented at the *Conference on Human Factors and Computing Systems (CHI 2004)*, Vienna, Austria.
- Brase, J., Painter, M. and Neijdl, W. (2003) 'Completing LOM – how additional axioms increase the utility of learning object metadata', Paper presented at the *3rd International Conference on Advanced Learning Technologies (ICALT2003)*, 9–11 July, p.493.
- Breslin, J.G., Decker, S., Harth, A. and Bojars, U. (2006) 'SIOC: an approach to connect web-based communities', *Int. J. Web Based Communities*, June, Vol. 2, No. 2, pp.133–142.
- Brickley, D. and Miller, L. (2005) 'FOAF vocabulary specification', <http://xmlns.com/foaf/0.1>.
- Brin, S. and Page, L. (1998) 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks and ISDN Systems*, Vol. 30, pp.107–117.
- Grzonkowski, S. and Kruk, S.R. (2006) 'HyperCuP lightweight implementation: a universal solution for making application distributed', *Demo and Posters Proceedings of the 3rd European Semantic Web Conference*, Budva, Montenegro.
- Hardt, D. (2004) 'Personal digital identity management', Paper presented at the *FOAF Galway Workshop*, September.
- Hildreth, P. and Kimble, C. (2000) 'Communities of practice in the distributed international environment', *Journal of Knowledge Management*, Vol. 4, No. 1, pp.27–37.
- Kondratova, I. and Goldfarb, I. (2003) 'Design concepts for virtual research and collaborative environments', Paper presented at the *10th ISPE International Conference on Concurrent Engineering: Research and Applications*, Lisse, Portugal, 26–30 July, pp.797–803.
- Kruk, S.R. (2004) 'FOAF-Realm – control your friends' access to resource', *Proceedings of the FOAF Galway Workshop*, September.
- Kruk, S.R. and Decker, S. (2005) 'Semantic social collaborative filtering with FOAFRealm', Paper presented at the *Semantic Desktop at ISWC 2005*, Galway, Ireland.
- Kruk, S.R., Decker, S. and Zieborak, L. (2005) 'JeromeDL – reconnecting digital libraries and the semantic web', Paper presented at the *16th International Conference on Database and Expert, Systems Applications*, Copenhagen, Denmark.

- Kruk, S.R., Grzonkowski, S., Gzella, A. and Cygan, M. (2006a) 'DigiMe – ubiquitous search and browsing for digital libraries', Paper presented at the *MoSO Workshop at MDM Conference*, Nara, Japan.
- Kruk, S.R., Grzonkowski, S., Gzella, A., Woroniecki, T. and Choi, H.C. (2006b) 'D-FOAF: distributed identity management with access rights delegation', Paper presented at the *1st Asian Semantic Web Conference*, Beijing.
- Kruk, S.R., Gzella, A., Czaja, F., Bultrowicz, W. and Kruk, E. (2007a) 'MultiBeeBrowse – accessible browsing on unstructured metadata', *Proceedings of the 6th International Conference on Ontologies, DataBases, and Applications of Semantics*, Vilamoura, Portugal.
- Kruk, S.R., Gzella, A., Dobrzanski, J., McDaniel, B. and Woroniecki, T. (2007b) 'E-learning on the social semantic information sources', Paper presented at the *2nd European Conference on Technology Enhanced Learning*, Crete, Greece.
- Mika, P. (2005) 'Ontologies are us: a unified model of social networks and semantics', Paper presented at the *4th International Semantic Web Conference*, Galway, Ireland.
- Milgram, S. (1967) 'The small world problem', *Psychology Today*, May, pp.60–67.
- Oren, E., Delbru, R. and Decker, S. (2006) 'Extending faceted navigation for RDF data', Paper presented at the *ISWC'2006*, Athens, GA, USA.
- Schlosser, M., Sintek, M., Decker, S. and Nejdil, W. (2003) 'HyperCuP – hypercubes, ontologies and efficient search on P2P networks', Paper presented at the *Agents and Peer-to-Peer Computing (AAMAS 2003)* Bologna, Italy.
- Skvoretz, J. (2002) 'Complexity theory and models for social networks', *Complexity*, Vol. 8, No. 1, pp.47–55.
- Yee, K.P., *et al.* (2003) 'Faceted metadata for image search and browsing', Paper presented at the *Conference on Human Factors in Computing Systems*, Florida, USA.

Notes

- 1 SemanticWeb Community Portal, <http://www.semanticweb.org/>.
- 2 Wikipedia, <http://wikipedia.org/>.
- 3 SIOC, <http://sioc-project.org/>.
- 4 FOAF, <http://xmlns.com/foaf/0.1/>.
- 5 W3C, <http://w3c.org>.
- 6 FOAF-A-Matic, <http://www.ldodds.com/foaf/foaf-a-matic.html>.
- 7 FOAFNaut, <http://www.foafnaut.org/>.
- 8 LinkedIn, <http://www.linkedin.com/>.
- 9 Ryze, <http://ryze.com/>.
- 10 WLSpaces, <http://spaces.live.com/>.
- 11 Xanga, <http://xanga.com>.
- 12 Drupal, <http://drupal.org/>.
- 13 XML User Profiles, <http://xprofile.berlios.de/>.
- 14 Identity 2.0, <http://www.identity20.com/>.
- 15 Microsoft Passport, <http://www.passport.net/>.
- 16 Windows CardSpace, <http://msdn.microsoft.com/cardspace/>.
- 17 OpenID, <http://openid.net/>.
- 18 OASIS Extensible Resource Identifier (XRI) TC, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xri.
- 19 Tomcat, <http://jakarta.apache.org/tomcat/>.

- 20 RDF2GO, <http://rdf2go.ontoware.org/>.
- 21 Sesame, <http://openrdf.org/>.
- 22 JOnto – Java Binding for Ontologies, Taxonomies and Thesauri, <http://sf.net/projects/jonto/>.
- 23 The Dewey Decimal Classification (DDC), <http://www.oclc.org/dewey/>.
- 24 WordNet, <http://wordnet.princeton.edu/>.
- 25 The Open Directory Project, <http://dmoz.org/>.
- 26 SIOC Exporters, <http://sioc-project.org/exporters>.
- 27 SOA for MBB, <http://wiki.s3b.corrib.org/MBB/SOA>.
- 28 notitio.us, <http://notitio.us/>.