SHARING INFORMATION ACROSS COMMUNITY PORTALS WITH FOAFREALM

John G. Breslin, Slawomir Grzonkowski, Adam Gzella, Sebastian R. Kruk, Tomasz Woroniecki
Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
IDA Business Park, Lower Dangan, Galway, Ireland
firstname.lastname@deri.org

ABSTRACT

Community portals such as blogs, wikis and photo sharing sites have become the new channels for information dissemination on the Web. When searching for information, many results end up in some type of community site. However, one cannot make use of the wealth of information that is available through the preferences of one's thematic social network, or through bookmarks or other documents, only those accessible through a social network. Also, due to multiple accounts being registered on a variety of community portals, there is a lack of semantics regarding the information that a particular user has created or bookmarked across this set of portals. This paper presents a method for sharing information across multiple community portals through a social semantic collaborative filtering system; this collaborative filtering extends a popular "Friend Of A Friend" (FOAF) network; it enables users to share the bookmarks and community documents that they create.

1. INTRODUCTION

Community-based applications such as blogging and wikis have become very popular and at the same time have created an interconnected information space (through the "blogosphere" and inter-wiki links). More and more content is being created by users on "Web 2.0" community portals ranging from pictures on photo album sharing sites to community event information to bookmarks on topics of interest. At the same time, these applications are experiencing boundaries in terms of information dissemination and user profile automation. A very simple example of a question that cannot be easily answered at the moment is "show me all the content created by Mick and all his close friends in the past week".

We will now introduce what a community portal is, detail some of the problems with such portals and then give an overview of the proposed solution to the problems via an example scenario.

1.1 Community Portals

Community portals are online community-specific websites that provide improved communication and contact links for a community online (e.g. one that is providing local or interest-based information). They are the most widespread platform used by communities to stay informed electronically. Members can find relevant information and may contribute any required shared information to others via the portal.

By having an online collaboration space for a community of a certain interest, community portals provide an awareness and interaction amongst a set of people whether for profit or non-profit. These portals are replacing the traditional means of information exchange. They help to provide an online global communication agora, and to strengthen the communities themselves by informing them and by providing an open place for interaction and exchange of information and ideas.

1.2 Problems with Existing Portals

Community portals have many significant advantages over traditional community collaboration methods (e.g. newsletters), and can be very useful and helpful for users of the portals. Unfortunately users of classical community portals are facing many potential problems.

Each portal usually has its own user management system. Users are forced to create an account and then remember a set of credentials. Moreover, users need to present almost exactly the same information for each portal they register to (such as their personal details, domains of interest, etc.). Finally, a community member may store different information relating to a particular domain of interest on each community portal. This information needs to be copied into one place, and then merged, which is of course time consuming; not every user would decide to perform such an operation.

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

1.3 Solution Scenario

Figure 1 illustrates a scenario which requires a solution to the problems described above. Let us take the 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.

We will begin by looking at John, and the content cloud that represents his membership of various online community portals. John is interested in digital photography, and he is a member of a weblog site and a photo sharing community. John also has three friends, Mick, Mike and Sheila, some of whom are registered to 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 also to annotating digital photos with metadata, a new interest. He also 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 the accounts on the weblog 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 weblog or photo sharing communities, he will need to copy-and-paste links for all the relevant discussions to the new bookmarking site as there is no common exchange mechanism for resources or resource links between the various communities.

In the solution proposed in this paper, 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 et al, 2006c) 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 weblog). 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.

Also, if a user specifies their 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 is registered on a bulletin board site (for Semantic Web developers) and says that she has an interest in resources tagged or categorised with "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 resources linked to those topics. One of the videos is about using Semantic Web technologies to provide an enhanced program guide for television over data networks, and is tagged as being related to "Semantic TV", and she marks this as a topic of interest. Then on the original bulletin board site, more resources (matching "Semantic TV") are presented.

1.4 Outline of the Paper

The Semantic Web (Berners-Lee et al, 2001) is increasingly aiming at applications areas. The aforementioned areas of community portals and the social networks (Adamic et al., 2003) formed therein is one of the obvious targets for Semantic Web research, due mainly to the recent explosion in the number of online social network site users, the growing popularity of other community portal sites such as blogs and media-sharing sites, and the potential benefits of connections between these community networks using semantic technologies. Quite a number of Semantic Web approaches have recently appeared to overcome the boundaries being encountered in these application areas, e.g. SIOC (semantically-interlinked online communities) (Breslin et al., 2005), structured blogging, semantic wikis, etc.

This paper will describe a combination of efforts to share and access the information related to networks of people across various community sites. The 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 authentication and trust infrastructure. D-FOAF is designed to operate without a centralised authority. We will begin by describing related work in this area, and then we will describe the profile management and social bookmarking systems in some detail, followed by details of how documents and
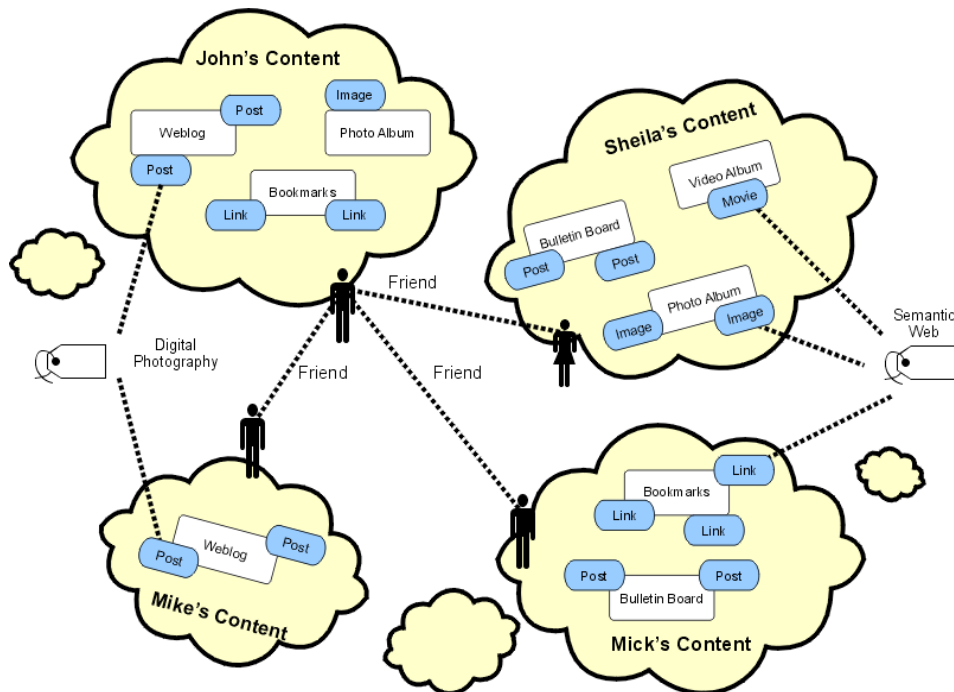

Figure 1. Sharing content between disparate community sites using a social network and topics of interest.

resources can be classified and exchanged amongst members of a distributed community network. Finally, we will outline some plans for future work based on our results so far in connecting the content of a user and their social network across multiple community sites.

2. BACKGROUND RELATED WORK

2.1 Online Communities

Online communities have become more and more popular, and they can no longer be considered as niche systems. Every country and every business trade has at least several popular portals. Hildreth at al. (2000)

propose 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) distinguish between three main objectives of online communities. Firstly, to supply content to the users. Secondly, to encourage members to participate in the community by contributing. Finally, it has to facilitate communication and interaction between the members. The key features that differentiate online communities are the various kinds of forums, wikis, chat rooms, as well as online and offline events that they may have (e.g. the Semantic Web community portal[1] uses a mailing list as its primary communication medium). It is difficult to say which type of community portal is the most popular, because as many interests as there are portals exist. One potential candidate could be Wikipedia, since it gathers people regardless of interests.

2.2 Social Networks

The aim of the "Friend Of A Friend" or "FOAF" standard is to utilise machine-readable homepages for describing people. Moreover, the idea also proposes storing links between people and activities that they take part in (i.e. by specifying their topics of interest). In order to achieve this, the "FOAF-vocabulary"[2] was introduced by Brickley and Miller. The vocabulary is strongly dependent on W3C[3] standards, especially RDF and XML. A number of applications have been developed that make use of metadata provided using this vocabulary: FOAF-A-Matic[4] allows those not familiar with XML to easily create people descriptions, and FOAFNaut[5] provides a visualization of any social networks formed using FOAF user profiles.

Additionally, many online social network sites have taken advantage of Milgram's (1967) "six degrees of separation" observation, especially Friendster (Berners-Lee et al, 2001) which was initiated in 2002 and has received some patents in this domain. Furthermore, there has been a large growth of business-oriented networks. LinkedIn[6] and Ryze[7] 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 aim of the issue was to show the complexity for different levels of network architecture, and to help with the comprehension of network-based analyses and explanations.

2.3 User Profile Management Systems

The issue of user profile management systems is a ubiquitous matter. Many research project concepts have been proposed, ranging from open source to commercial. Some are offering sophisticated features such as distributed profiles and single sign-on functionality. Examples of such systems are Drupal[8] and XUP[9] (the latter being similar to the W3C FOAF metadata recommendation).

Another idea called Protocol Identity 2.0[10] has been proposed for the exchange of digital identity information. The general idea entails that users be supported with enhanced control over the information entrusted to other members of the community. The authors of Sxip 2.0 (Hardt, 2004) have announced that their system will provide this feature. In addition, it will be possible to adjust security needs to a specific site.

Probably the most famous profile management system is Microsoft Passport[11] which support the reuse of profile information across different services. Although an interesting idea, frequent bug reports and the centralised topology mean that the system has not yet been commonly accepted by other sites.

---

[1] Semantic Web Community Portal: http://www.semanticweb.org/
[2] FOAF: http://xmlns.com/foaf/0.1/
[3] W3C: http://w3c.org
[4] FOAF-A-Matic: http://www.ldodds.com/foaf/foaf-a-matic.html
[5] FOAFNaut: http://www.foafnaut.org/
[6] LinkedIn: http://www.linkedin.com/
[7] Ryze: http://ryze.com/
[8] Drupal: http://drupal.org/
[9] XML User Profiles: http://xprofile.berlios.de/
[10] Identity 2.0: http://www.identity20.com/
[11] Microsoft Passport: http://www.passport.net/

An example of a decentralised digital identity system is OpenID[12], in which a user's online identity is given either by a URI (such as for their blog or a home page) or an XRI (OASIS Extensible Resource Identifier)[13].

3. SOLUTION USING FOAFREALM AND D-FOAF

3.1 Distributed User Profile Management (A)

The transparent distribution of a user's profile offered by FOAFRealm fits very well with the requirement for inter-portal cooperation. Information about a user's preferences may be collected from all connected sites and 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.

3.2 Social Semantic Collaborative Filtering (B)

Social semantic collaborative filtering (SSCF) is based on two concepts: distributed collections and annotations of resources. Each user classifies only a small subset of the knowledge, based on the level of expertise they have on a specific topic. This knowledge is later shared across the social network.

3.2.1 Classifying Community Portal Documents (B.1)

During their online activity users can bookmark some resource. Unfortunately such information needs to be properly classified to be used by the system; the SSCF module allows users to classify their bookmarks with "domains of interest", which are represented by semantically annotated catalogs. Domains contain bookmarks and may also include other domains. This structure needs to be well classified; user's taxonomy of catalogs needs to refer to other knowledge organization systems. SSCF can utilise well known classification systems with JOnto[14] plugin; a user can annotate catalog's content using e.g. DDC, WordNet and dmoz:

- The Dewey Decimal Classification (DDC)[15] is a general knowledge organisation tool that is continuously revised to keep pace 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 sixty 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 3 digits. The first digit 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.
- WordNet[16] is an online lexical reference system whose design is inspired by 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 consist of over 200k word-sense pairs (over 150k unique strings).

[12] OpenID: http://openid.net/
[13] OASIS Extensible Resource Identifier (XRI) TC: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xri
[14] JOnto – Java Binding for Ontologies, Taxonomies and Thesauri: http://sf.net/projects/jonto/
[15] The Dewey Decimal Classification (DDC): http://www.oclc.org/dewey/
[16] WordNet: http://wordnet.princeton.edu/

- dmoz is the Open Directory Project[17], 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. The Open Directory powers the core directory services portal for the Web's largest and most popular search engines. All Open Directory resources (structure and content) are freely available to use.

With FOAFRealm-SSCF different parts of community portals can be classified using methods described above. A user can easily assign a class to discussions, wiki pages, blogs, photo albums, as well as normal pages on the Web.

3.2.2 Mechanism for Exchanging Documents Between People (B.2)

Social semantic collaborative filtering is strongly dependant on the social network, which can be stored as a directed graph. Nodes describe users whereas edges represent the relationships between them. Additionally, to overcome problems with security, each link between two people can also have an assigned trust level that decides whether access should be granted or denied. Users can have a collections of bookmarks (i.e. a private bookshelf as described by Kruk at al. (2005)), which represent their knowledge; later they can renders 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 quality metrics assigned to it; therefore the owner is able to specify their expertise level on a particular topic that can be computed with the PageRank algorithm (Breslin et al., 2005) 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 given topic, are potentially useful and reliable.

To sum up, 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 catalog; they can restrict access to a certain sub-graph of a social network.

3.2.3 All the Parts Coming Together (A + B.1 + B.2)

With its advanced distributed model and collaboration features, FOAFRealm aims to be a complex solution for managing identities and preferences. Fine grained access control lists make it easy to share resources among friends and to spread knowledge in a community. Single sign-on and single registration lets a user comfortably use multiple services and also helps to start up new sites by connecting them with existing popular ones. Browsing others' bookmarks and annotations gives users the benefit of using valuable resources collected by experts. By simply deploying our FOAFRealm library, all of these features can be easily incorporated into a portal or any other application which wants to leverage utilisation of social networks.

4. FUTURE WORK

The next step of evolution is a system called DigiMe (Kruk et al, 2006a), a research topic that has recently been initiated. The aim of DigiMe is to deliver SSCF features that will support mobile devices and provide users with better control over their profile information. Users can store collaborative resources and profile information on their mobile device; the DigiMe system uses this information to explore the ad-hoc social networks paradigm.

5. CONCLUSIONS

In this paper, we detailed many issues with community portals that are experiencing boundaries in terms of content dissemination and profile automation. 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

---

[17] The Open Directory Project: http://dmoz.org/

easily make use of their social networks between sites, for example, to leverage the skills of a friend who may an expert in one domain on a different community site. We have described the FOAFRealm and D-FOAF implementations that can overcome these boundaries by providing a distributed user profile management system along with social semantic collaborative filtering. This system provides an excellent method of sharing resources between friends or associates by defining the level of expertise that a person has on a particular topic (according to those they are connected to via a social network) and by suggesting various resources based on these expertise levels. We finished by briefly describing future work for mobile devices.

ACKNOWLEDGEMENT

REFERENCES

Adamic et al., 2003. A Social Network Caught in the Web, First Monday, vol. 8, no. 6, http://firstmonday.org/issues/issue8_6/adamic/

Berners-Lee et al, 2001. The Semantic Web, *Scientific American*, May 2001

Boyd, 2004. Friendster and Publicly Articulated Social Networking, *Proceedings of the Conference on Human Factors and Computing Systems (CHI 2004)*

Breslin et al., 2005. An Approach to Connect Web-Based Communities, *Proceedings of the 2nd IADIS International Conference on Web Based Communities (WBC 2005)*, pp. 272-275, Carvoeiro, Portugal

Brin and Page, 1993. The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Computer Networks and ISDN Systems*, vol. 30, pp. 107-117

Dodds, 2004. An Introduction to FOAF, http://www.xml.com/pub/a/2004/02/04/foaf.html

Hildreth et al, 2000. Communities of Practice in the Distributed International Environment, *Journal of Knowledge Management*, vol. 4, no. 1, pp. 27-37.

Hardt, 2004. Personal Digital Identity Management, *Proceedings of the FOAF Galway Workshop*, September 2004.

Kondratova and Goldfarb, 2003. Design Concepts for Virtual Research and Collaborative Environments, *Proceedings of the 10th ISPE International Conference On Concurrent Engineering: Research and Applications*

Kruk, 2004. FOAF-Realm - control your friends' access to resource, *Proceedings of the FOAF Galway Workshop*, September 2004

Kruk et al, 2005. JeromeDL - Reconnecting Digital Libraries and the Semantic Web, *Proceedings of the 16th International Conference on Database and Expert Systems Applications*

Kruk et al, 2006a. DigiMe - Ubiquitous Search and Browsing for Digital Libraries, *Proceedings of the MoSO Workshop at MDM Conference*, Nara, Japan, 2006

Kruk et al, 2006b. D-FOAF: Distributed Identity Management with Access Rights Delegation, *Proceedings of the 1st Asian Semantic Web Conference*, Beijing, 2006

Kruk et al, 2006c. Social Semantic Collaborative Filtering for Digital Libraries, *Journal of Digital Information, Special Issue on Personalization*, 2006

Milgram, 1967. The Small World Problem, *Psychology Today*, pp. 60–67, May 1967.

Skvoretz, 2002. Complexity Theory and Models for Social Networks, *Complexity*, vol. 8, no. 1, pp. 47–55