Getting to Me – Exporting Semantic Social Network Information from Facebook

Matthew Rowe, Fabio Ciravegna

Web Intelligence Technologies Lab
Department of Computer Science
University of Sheffield, UK
{m.rowe,f.ciravegna}@dcs.shef.ac.uk

Abstract: Information sharing forms a large component of the Social Web as web users become citizens within the web sphere. Within 'walled garden' services they interact socially with their peers through blog posts, image sharing and writing on one another's 'walls'. Porting this data for reuse opens up the social graph associated with the user. In this paper we present work to export Semantic information from the social networking site Facebook using the FOAF ontology by mashing up several Web 2.0 services. We explain the details of our approach and how it supports the data portability movement.

Keywords: data portability, identity, social networks, social networking sites

1 Introduction

The uptake of Social Web and Web 2.0 sites and services has seen a staggering number of people organising their lives online, and moving their offline social space into the virtual realm. A blog post is the new public ego-centric diary entry, writing on someone's virtual 'wall' is the new email, and planning an event can be done within minutes, including sending all invitees their invitations. The general public has embraced the Social Web with open arms, willingly divulging their personal information, their email address, their home address, who they are friends with, and what they like to do. This abundance of information forms a rich social sphere, enabling intelligent reuse for advertisers and consumer groups. However, one may pose the question, who owns this information? A common misconception is that the

user, essentially the author, owns it. It is also a misconception that companies will tell you everything they have logged about you using their site by simply asking for it as the general public, in the UK, believe they have a right to such information based on the data protection act[1]. Therefore surely there is a need for exporting, or claiming back, our information and what belongs to us.

The social web has grown to 9 million unique Facebook users, 5 million unique MySpace users and 4.1 million unique Bebo users [18] in the UK alone, this sharp rise in users and the accompanying habit of information divulgence has begun to plateau: In February of 2008, Facebook saw it's first slump in UK numbers [18], and the movement to make data portable from the social web, namely from social networking sites (SNS), saw the foundation of the Data Portability group[2] to oversee data portability standards and roadmap work within this area. It has become apparent that web users are beginning to wonder what happens to their data.

The data portability movement focuses on opening up the social graph, enabling social information to become more accessible. In 2007 work by [17] explained a 'Bill of Rights for Users of the Social Web' detailing the role of web users as the sole owners of their information, they have the ability to control the distribution of their information, and have the freedom to allow or deny access to their information to any requesting site or service. Such rights allow for a user-centric identity allowing a web user to take their identity with them, similar to the real world, as they move through the social web. The reputation they build up on Ebay, for example, would be transferable to other similar auction sites deeming the level of trust associated with the buyer or seller. The idea that an individual is to be treated as a citizen and not as a consumer was presented in early work by [13] in which virtuality is merged with reality and people in the real world can be held accountable for their online actions.

In this paper we present our work contributing to the data portability movement. We have provided an online application called the Facebook FOAF Generator[3] allowing any user of Facebook to extract their identity and the associated relationships. We use Semantic Web technologies to formalise the information we export into a Semantic format, enabling machine readability. At present generating reusable FOAF is a laborious process, either building it by hand, or using an existing generation tool does not explicitly capture identity. Instead using existing content from a Web 2.0 site bootstraps this process, and provides a rich source of FOAF content. FOAF provides a useful specification for describing information found on SNS, which can be later reused by FOAF enabled sites and services.

Following this section we explain existing methods for formalising identity information using Semantic Web technologies. Section three explains existing techniques for the extraction of social information both from the wider Web and the Social Web. Section four then explains our approach to exporting identity information from Facebook. Section five presents the response we have had to the application, and section six describes our conclusions from the work that we have carried out.

---

[1] www.opsi.gov.uk/Acts/Acts1998/ukpga_19980029_en_1

[2] http://www.dataportability.org/

[3] http://ext.dcs.shef.ac.uk/~u0057/FoafGenerator

2 Formalising Identity

Before we move on to describing the state of the art within the sensitive area of exporting *my* data from a SNS, it is crucial that we explain existing specifications for describing who a given person is, and who they are associated with. Therefore we describe two existing semantic technologies and their application in the formalisation of personal information, we then move on to present the use of small-level semantics to express implicit knowledge within web pages.

FOAF

In order to capture knowledge depicting a given person the Friend of a Friend (FOAF) specification was established. Work in [3] describes an ontology containing classes and properties designed to encapsulate existing identity knowledge available on the Internet at the time. A person begins by describing himself or herself using the *foaf:Person* class, listing key identity attributes such as name, gender, and resources relating to them. They can also list their interests, and each person is uniquely identified by using the *foaf:mbox* property containing their email address (it is still useful to consider how this is normally the primary key for online 'signup' forms). An alternative identification property is *foaf:openid* conforming to the OpenID[4] perspective of using a unique single URI to establish the identity. The person in question then moves on to describing their friends, each friend is an instance of the *foaf:Person* class. FOAF is both machine-readable, and human-readable, and was adopted by LiveJournal[5], the blogging site, to offer the facility for each user to export their personal information. FOAF has seen a slow adoption by Web 2.0 sites and services, this could be due to the lack of interest in the exportation of social information from one rival site to another. It is, however, the most widely used specification for expressing personal and relationship information within the Semantic Web community.

SIOC

The creation and development of online communities has provided web users with a virtual realm where they can express their thoughts, gain feedback and critique, and interact with individuals harbouring similar interests and beliefs. Most modern web users participate in a web community in some form, albeit forums, chat rooms, newsgroups and SNS. Each community can be interpreted as a single island of rich knowledge, unlinked and unique. The SIOC project focuses on ways to integrate and merge these islands, providing bridges between the knowledge that exists there. As an extension of the FOAF ontology, work presented in [4] outlines an ontology capable of capturing knowledge associated with online communities, and offers formalisms for establishing links between communities.

---

[4] http://openid.net/

[5] http://www.livejournal.com

The SIOC ontology uses existing specifications such as FOAF and RSS to define classes and their properties. The SIOC ontology contains six main classes of knowledge associated with a community: Site, forum, post, role, usergroup and user. These abstract class definitions offer a flexible range for the capture of knowledge existing in online communities. The concept *sioc:User* can be thought of as the central point for the ontology, s*ioc:User* is a sub class of the *foaf:OnlineAccount* class. The SIOC ontology also defines properties that relate each of the classes to the user in a similar manner to how a user interacts within an online community. Such as *has_moderator* to describe the relationship a forum has with the person monitoring it, *has_administrator* to describe a site's relationship with the user who runs it, and so on. For data portability purposes, SIOC provides a vital specification when exporting data from sites containing groups or forums, or any services where messaging plays a crucial role. SIOC allows the capture of interactions between individuals, and is capable of expressing the role an individual plays in an online social space.

Implicit Social Information

Inclusion of social semantics within web pages has achieved wide spread adoption, particularly with the move by search engines to incorporate advanced knowledge retrieval in their results. New initiatives such as Search Monkey[6] and Google's Social Graph[7] readily use small snippets of semantics within web page code to either return search results containing knowledge of people (Search Monkey), or to construct a linked social graph (Social Graph). This subsection presents two technologies used to code the semantics of social information within a given web page. Each technology uses small-level semantics in order to achieve knowledge descriptions, complying with Tim Berners-Lee's ideology:

"The trick.... is to make sure that each limited mechanical part of the Web, each application, is within itself composed of simple parts that will never get too powerful." [2]

Microformats

The lowercase Semantic Web as paraphrased by [5] involves small-level semantics embedded in XHTML code. Microformats[8] contain useful representations of everyday knowledge commonly found in web pages. The web designer responsible for a given web page is able to add additional machine-readable information within the page content, invisible to the human reader as it is rendered within the browser. According to [5], Microformats make Semantic annotations possible for people and organisations, calendars and events, opinions, ratings and reviews, and tags and keywords.

---

[6] http://developer.yahoo.com/searchmonkey/
[7] http://code.google.com/apis/socialgraph/
[8] http://microformats.org/

Microformats can also be used to describe relationships between people through link structures[9] using the XFN Microformat. Small level semantics are included within XHMTL code by web designers to support semantic agents to access valid relationship metadata. Examples of this application appear in Blogrolls and link pages. XFN allows the web designers to include details about a relationship such as the type of relationship, whether it is a friend, a colleague or a member of the person's family, the origin of the relationship and the geographical semantics of the relationship.

RDFa

One of the most widely adopted specifications for the semantic description of knowledge is the Resource Description Framework (RDF) [6]. RDF offers a useful formalisation for knowledge capture and has been adapted by [1] to allow inclusion within XHTML through lightweight semantics. Unlike Microformats and particularly XFN, the web designer incorporating RDFa is able to include references to classes that appear in any ontology on the web, therefore supporting ontology reuse. In essence RDFa allows a much more expressive formalisation of semantic information, capable of encapsulating any concept from any ontology providing its namespace exists. Microformats and XFN offer a much lighter approach, one that a non-Semantic Web-savvy web designer would be more comfortable using due to the exclusion of ontology references. All three lightweight formalisms presented in this section offer useful means for the inclusion of personal semantic information within existing web pages, facilitating the extraction and portability of social information.

3 Extracting Social Information

This section details the current state of the art associated with extracting social information. Our work focuses on the portability of an individual's identity, of which we regard social network information to be a crucial component. As described in [9] there are 3 tiers of identity that exist in the modern realm of the Social Web:

- My Identity
- Shared Identity
- Abstracted Identity

The top tier, My Identity, contains information that is persistent about a given person. This information is rarely altered and is unique to the person, aiding their disambiguation from others. The second tier depicts identity information that is more temporal, and contains features and relations that help to bind a given person with others; i.e. relationships. One of the main challenges that the social network mining community has faced is how to ascertain knowledge from this tier. We now present relevant work in social network mining, and explain existing technologies designed to aid with Shared Identity knowledge capture. We conclude by describing existing approaches for porting data from Web 2.0 services into RDF.

[9] http://gmpg.org/xfn/

Social Network Mining

Work by [14] and [11] presents a three-step approach by mining the web for social network information to isolate links between two people, these links are monitored for interactions both in the real world and on the web. [15] presents a two-part process by mining the web and retrieving semantic documents containing FOAF. Relationship strengths between two people are derived using name co-occurrence from web queries. A methodology in [10] crawls the FOAF-web, extracting information from each FOAF file and aggregating with information from other FOAF files. Assertions are made about discovered individuals using the supplied semantic information. Work by both [12] and [16] identifies labels for relations between two people, not only providing a social network, but also providing extra information for each edge in the social graph.

Exporting Social Networks from the Social Web

Flickr Exporter

Work by [7] has created a web application designed to export a user's personal and community information from Flickr[10]. The 'Flickr Exporter'[11] works using a simple user interface requesting the username of the user to export data for. The exporter generates RDF using both the FOAF and SIOC specifications [8]. The user's personal information is described using FOAF, detailing their name, and location, along with their friends on Flickr. The *foaf:knows* relation is used to depict those relationships. User data is interlinked by describing the geographical location of the user using Geonames[12], giving a unique resolvable URI to the location. The SIOC specification is used to describe additional properties of the user detailing the location of the user's photo gallery and their profile image. The exporter also details any communities that the user has joined within Flickr, as this forms an integral part of the site allowing users to share and discuss specific photo topics. The SIOC specification is well suited to this task, as it facilitates the expression of forums and groups, and whether a given user is a member, and what their role is.

Twitter Exporter

Similarly, an exporter of RDF[13] has also been written for the micro-blogging service Twitter. This exporter uses both the FOAF and SIOC ontologies to describe the friends a user is 'following' on Twitter, i.e. their social network. Each friend is described as an instance of *foaf:Agent* containing FOAF semantics such as *foaf:name*, *foaf:nick* and *foaf:homepage*. A blog post made by a user is expressed as an instance

[10] http://www.flickr.com
[11] http://apassant.net/home/2007/12/flickrdf/
[12] http://www.geonames.org/
[13] http://tools.opiumfield.com/twitter/mattroweshow/rdf

of *rdf:Description* containing *sioc:content* to express the content of the post, with *foaf:maker* linking back to the namespace of the poster.

4 Exporting My Identity

Porting data from an existing web service is the main focus of this paper. We present in this section, a successful approach to perform this task involving a social networking site containing information hidden within a 'walled garden'. Our approach makes use of Semantic Web technologies to support reuse of the produced RDF as part of an initiative to make social network information portable from all of the major social networking and social web sites.

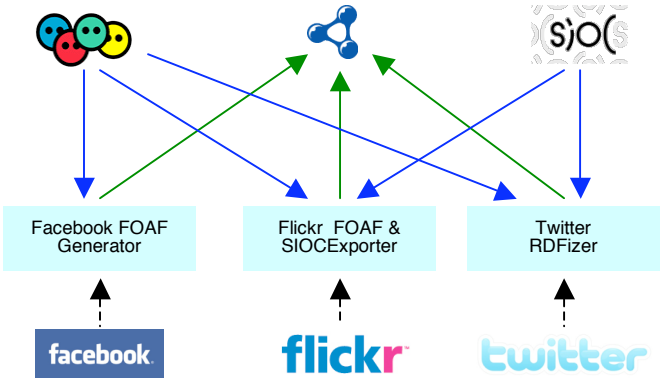

Figure 1 – Current status of data portability from Web 2.0

As figure 1 demonstrates the current model of data portability from social web sites employs various existing Semantic Web ontologies and specifications to generate reusable RDF. Each exporter or generator shown within the middle tier operates by producing RDF according to either the FOAF or SIOC ontologies. In essence the middle tier provides a mapping between the implicit knowledge offered by the service and Semantic Web ontologies, the mapping is commonly between an XML schema or API specification and several ontology concepts. Should a service in the lower tier alter, for example by adding new features, the exporter can be easily adapted to capture this new knowledge. By separating the model as such, this allows for greater flexibility, important when considering the rate of growth of the social web, and its susceptibility to change.

Each exporter and generator within the middle tier can be regarded as a wrapper, purporting knowledge reuse through the generated RDF. In the remainder of this section we present our approach to aiding data portability for social networks through the generation of RDF content according to the FOAF specification from the social networking site Facebook using our application; the Facebook FOAF Generator. In the context of Figure 1, the Facebook FOAF Generator only uses the FOAF

specification for RDF generation; this is due to the lack of community data accessible from the site. Should this data become available, the incorporation of knowledge according to SIOC ontology would also be included. At present only the porting of social network data is supported.

Facebook FOAF Generator

This subsection presents the methodology and processes involved in exporting information from Facebook. The actual exporting of personal information involves mapping the existing Facebook XML data schema with concepts from the FOAF specification to enable sufficient knowledge representation. Figure 2 presents an overview of the process involved when generating FOAF from Facebook.

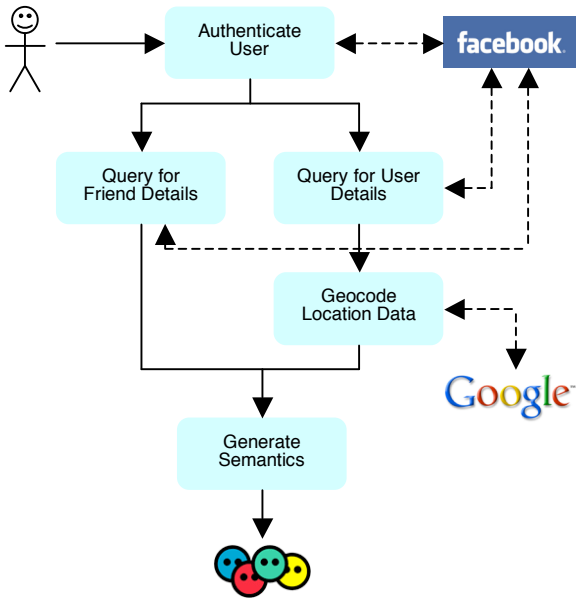

Figure 2 - Overview of the process of FOAF generation from Facebook

The process of FOAF generation begins by authenticating the user through Facebook's API. Upon authentication of the user the process can then begin by querying the API to retrieve the user's personal information and a list of the user's friends together with their details. First the user details are converted into FOAF, the *foaf:name* class contains the name of the user, *foaf:gender* describes the sex of the user, and *foaf:img* is used to contain a URL depicting an image the user.

In order to uniquely identify this FOAF file from others containing similar properties, the *foaf:mbox* or *foaf:mbox_sha1sum* classes are normally used. Facebook does not allow the exporting of email addresses, they are not offered by the API. Instead, our approach uses the user identification number assigned by Facebook to the

user. This is a simple incremental integer that is often seen in browser query strings when using the site. Therefore we used the *foaf:holdsAccount* property, containing the details of the user's account with Facebook. The account is described using the class; *foaf:OnlineAccount* within which the account service, in this case Facebook, is identified using *foaf:accountServiceHomepage*, and the *foaf:accountName* property contains the user identification number. The user's interests are also expressed using the *foaf:interest* property.

The next stage in the process is to create a geocoded representation of the given user's location. In order to do this we took the location information returned by the API and queried the Google Map[14] web service. The service returned a geocoded object containing the longitude and latitude of the given location. Using the *foaf:based_near* property to express a reference to the location, we included *geo:Point* class from the geocode ontology[15] and populated the *geo:lat* and *geo:long* attributes with the latitude and longitude respectively.

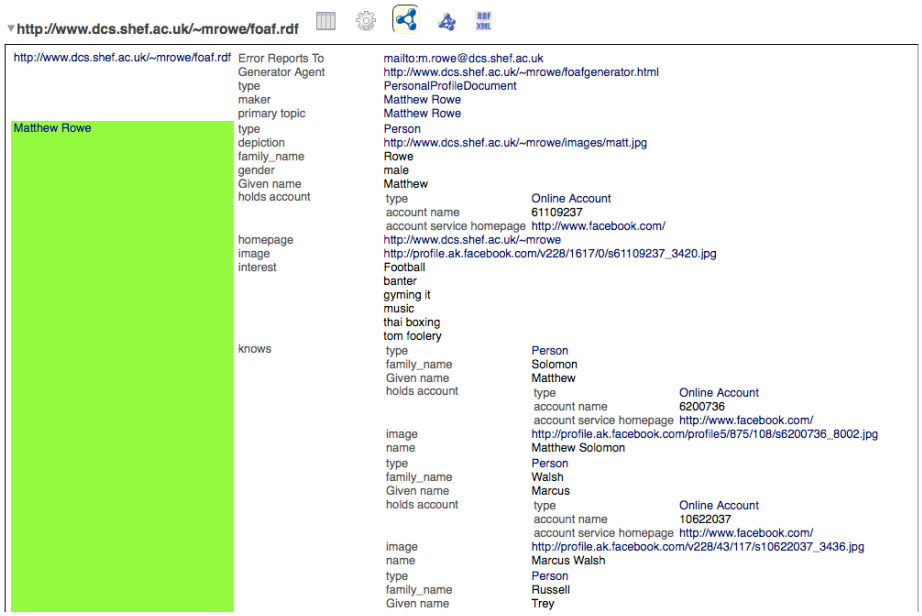

Figure 3 - Example FOAF produced by the Facebook FOAF Generator. Visualisation is achieved using the Tabulator[16] Addon for Mozilla Firefox.

The following stage involves capturing details of each of the user's friends in their social network, essential as we are using a 'Friend of a Friend' specification! The FOAF ontology contains the *foaf:knows* property to specify a relationship between

[14] http://code.google.com/apis/maps/
[15] http://www.w3.org/2003/01/geo/
[16] http://www.w3.org/2005/ajar/tab

two people. We use this property for each of the user's friends, and include the *foaf:Person* class to express the existence of a person. Within this class we include the *foaf:name* property to identity the friend, and the *foaf:img* property to contain an image of the friend. In order to uniquely identify the individual we use the same technique as mentioned previously through the user identification number provided by Facebook and store this value in the *foaf:accountName* property, within the *foaf:OnlineAccount* class.

The completion of this process generates a FOAF file that can then be downloaded by the user and placed on the web for usage by Semantic web agents and crawlers, to open to up the user's social graph enabling social network analysis and knowledge reuse.

5 Response

In this section we present the response given by users of the FOAF generator. Evaluating for usage is a fairly trivial process given that the user has a minimal amount of input, clicking a button, and the process of generating the RDF is fairly straightforward. Therefore we present results from RDF validators on 10 randomly selected FOAF files generated by the application, and the usage statistics since the applications launch in April 2008.

We tested 10 randomly selected FOAF files from the generator. We ran each file through the RDF:About Validator[17], and the W3C RDF Validation Service[18]. All the FOAF files that we tested using both validators were validated as being correct, and no errors were found. Figure 4 presents the usage statistics so far, with time displayed along the x-axis and user numbers along the y-axis. To date the FOAF generator has had 639 users, with an average of 6 active daily users, and 11 people have expressed their satisfaction and admiration for the application by becoming a 'fan' (feature in Facebook). Of those users we are unsure how many are Semantic Web 'enthusiasts'.

Figure 4 - FOAF Generator Usage

[17] http://www.rdfabout.com/demo/validator/
[18] http://www.w3.org/RDF/Validator/

6 Conclusions

In this paper we presented our approach to making social information portable from the social networking site Facebook. Our service has been widely adopted and appreciated by the web community based on the response we have had. The approach provides valid RDF that can be reused across the web, following the Bill of Rights for Users of the Social Web. Using FOAF to provide the semantics of the identity to be exported allows for the expression of relationships and the properties of those individuals. At the time of developing the application there was no need to incorporate the SIOC ontology as the Facebook API did not provide sufficient access to groups and communities the user may have joined and participated in, like Flickr allows. Future work would include the SIOC ontology for describing such community information should this information become available through the API.

One issue that needs to be addressed in the future is the usage of *foaf:accountName* property within the *foaf:OnlineAccount* class, which presents the unique identifier for the exported user and each of the user's friends within the RDF. In a similar manner to *foaf:mbox* we propose that this value should also have a hashed alternative able to protect the web presence of the user if they requested. However, such a process could be leveraged by the adoption of *foaf:openid* as an alternative identification property, by resolving a given profile URL from Facebook as the OpenID URI. Other future work will include additional geographical semantics similar to the technique adopted in [7] to enable interlinking of data. Place names will be resolved to specific geographical concepts using the Geonames service.

The walled garden social networking model inhibits the portability of social network information, forcing researchers to find alternatives through the creation of mashups and web applications to perform the exporting process. By offering a service to export RDF according to the FOAF specification, the social graph linked to a given person can then be reused by other service aiding such functions as relationship derivation, and social network analysis. Once we had developed the application issues arose when attempting to enlist the application with Facebook's application directory. Facebook were unhappy with the exportation of what they believed to be their information onto a separate server for reuse. The application was not blacklisted however; it was simply not listed to inhibit its distribution.

The future of access to social information really depends on the adoption of open standards capable of expressing identity implicitly. As mentioned previously, formalisms such as MicroFormts, XFN and RDFa go some way to allowing the expression of identity features within XHTML. With the creation of standards such as Google's OpenSocial[19] and the steady uptake of shared standards the move towards a more portable Social Web is becoming a reality, with a focus of user-centric identity at the forefront.

[19] http://code.google.com/apis/opensocial/

References

1. Abida. B., Birbeck. M.: RDFa Primer: Bridging the Human and Data Webs. http://www.w3.org/TR/xhtml-rdfa-primer/ (2008)
2. Berners-Lee. T.: Weaving the Web. Harper, San Francisco. (1999).
3. Brickley. D., Miller. L.: FOAF Vocabulary Specification. (2004).
4. Breslin. J., Harth. A., Bojars. U., Decker. S.: Towards Semantically Interlinked Online Communities. The Semantic Web: Research and Applications. (2005).
5. Khare. R.; Microformats: the next (small) thing on the Semantic Web? Internet Computing, IEEE. Vol. 10. Issue 1. Pp 68-75. (2006).
6. Lasilla. O., Swick. R.: Resource Description Framework (RDF) Model and Syntax Specification. W3C Working Draft WD-rdf-syntax-19981008. (1998).
7. Passant. A.: RDF Export of Flickr Profiles with FOAF and SIOC. http://apassant.net/blog/2007/12/18/rdf-export-of-flickr-profiles-with-foaf-and-sioc/ (2007).
8. Passant. A.: me owl:sameAs flickr:33669349@N00. Linked Data on the Web Workshop. WWW08, Beijing, China. (2008).
9. Windley. P. J.: Digital Identity. O'Reilly Media (2005).
10. Finin. T., Ding. L., Zhou. L., Joshi. A.: Social Networking on the Semantic Web. The Learning Organisation, vol. 1 , no. 5, pp. 418-435 (2005).
11. Hamasaki. M., Matsuo. Y., Ishida. K., Nakamura. Y., Nishimura. Y., Takeda. H.: Community Focused Social Network Extraction. Proceedings of 2006 Asian Semantic Web Conference (2006).
12. Jin. Y., Matsuo. Y., Ishizuka. M.: Extracting Social Networks among Various Entities on Web. The Semantic Web. International Semantic Web Conference 2006. pp. 487-500 (2006).
13. K Jordan, J Hauser, and S Foster. The Augmented Social Network: Building Identity and Trust into the Next-Generation Internet. (2003).
14. Matsuo. Y., Hamasaki. M., Nakamura. Y.: Spinning Multiple Social Networks for the Semantic Web. Proceedings of the 2006 Asian Artificial Intelligence Conference (2006).
15. Mika. P.: Bootstrapping the FOAF-Web: An Experiment in Social Network Mining. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web, Galway, Ireland (2004).
16. Mori. J., Tsujishita. T., Matsuo. Y., Ishizuka. M.: Extracting Relations in Social Networks from Web using Similarity between Collective Contexts. International Semantic Web Conference (2006).
17. Smarr. J.: Bill of Rights for Users of the Social Web. http://opensocialweb.org/2007/09/05/bill-of-rights/ . (2007)
18. Sweeney. M.: Facebook sees first dip in UK users. Guardian Newspaper. 21st Feb. http://www.guardian.co.uk/media/2008/feb/21/facebook.digitalmedia. (2008).