New Generation of Social Networks Based on Semantic Web Technologies: the Importance of Social Data Portability

Liana Razmerita[1], Martynas Jusevičius[2], Rokas Firantas[2]

Copenhagen Business School, Denmark[1]
IT University of Copenhagen, Copenhagen, Denmark[2]
liana.razmerita@cbs.dk, martynas@itu.dk, rokas@itu.dk

Abstract. This article investigates several well-known social network applications such as Last.fm, Flickr and identifies social data portability as one of the main technical issues that need to be addressed in the future. We argue that this issue can be addressed by building social networks as Semantic Web applications with FOAF, SIOC, and Linked Data technologies, and prove it by implementing a prototype application using Java and core Semantic Web standards. Furthermore, the developed prototype shows how features from semantic websites such as Freebase and DBpedia can be reused in social applications and lead to more relevant content and stronger social connections.

1 Introduction

Social networking sites are developing at a very fast pace, attracting more and more users. Their application domain can be music sharing, photos sharing, videos sharing, bookmarks sharing, professional networks, and others. Despite their tremendous success social networking websites have a number of limitations that are identified and discussed within this article. Most of the social networking applications are "walled websites" and the "online communities are like islands in a sea"[1]. Lack of interoperability between data in different social networks applications limits access to relevant content available on different social networking sites, and limits the integration and reuse of available data and information. This may result in a growing dissatisfaction of the user community and a reduced usability of the websites. More research combining social networks and Semantic web is required to address the above mentioned limitations. Research combining social networks and Semantic Web is an interdisciplinary field, atracting researchers from both social and information sciences. Current research is mostly related to extraction of semantic data from existing social applications, its representation and its analysis. For example, there is work done in extracting ontologies from user contributed folksonomies (collaborative tagging systems) and integrating ontologies together with folksonomies [2, 3]. Others propose approaches to development and evolution of lightweight ontologies in a similar collaborative way [4, 5]. Researchers seem to agree that folksonomies and (lightweight) ontologies share more common properties than differences and will be

further integrated, and thus community-based bottom-up development approach will prevail over top-down controlled engineering efforts.
Much of the current research for representing simple user profiles is based on the Friend of a Friend (FOAF) project[1] – a project aimed at "creating a Web of machine-readable pages describing people, the links between them and the things they create and do". FOAF is currently an important source of RDF data available on the Web which has already been used for social network analysis [6-8].
A related initiative is Semantically-Interlinked Online Communities (SIOC) project2, which provides an ontology for describing items and relationships from Internet discussion methods (such as blogs, forums, and mailing lists) to facilitate interconnection of these methods via publishing of metadata [9, 10]. Many recent papers show growing interest in portability issues among social network applications – they are being called "fundamental problems", and semantic technologies (mainly FOAF) are being proposed to solve them [11].

There is theoretical work done combining Semantic Web (SW) and social networks, especially in analysis of social networks and extraction of knowledge [12]. However, creation of new end-user semantic social applications as well as their design and implementation are not well explored. Existing social network applications do not employ SW technologies, although most of the standards infrastructure is already in place. Most of them are "walled" websites, which provide limited means for users and developers to control, publish, and access social data. This limits possibilities for reuse and integration, which are the driving forces behind Web 2.0 as well as Semantic Web, and results in growing dissatisfaction in the user community.
This article proves through the implementation of a prototype that Semantic Web technologies can be used to build a next generation of social networks that overcome limitations of social networks applications and enable new features currently not exploited by them.

2 Study of Social Networks

We analyzed applications which we personally use and which we think reflect the current state of the art in social networking: Last.fm, Flickr, Facebook, LinkedIn. These social networking applications have a number of technological limitations as summarized bellow:
• It is not possible to export/import profile data from one application into another
• It is not possible to export/import social relationships from one application into another
• There is usually less data available in machine-readable formats than the application contains
• Application Programming Interfaces (APIs) are based on variety of custom formats and protocols, some of them non-standard (such as FQL and FBML in Facebook)

1 http://www.foaf-project.org/
2 http://www.sioc-project.org/

Our observations fit well with statements by initiatives such as *Open Social Web*[3], *Social Network Portability*[4], *DataPortability*[5], *OpenID*[6], *OpenSocial*[7], born as a result of a growing dissatisfaction in user communities.

2.1 Semantic Social Networks

Freebase claims to be an open database of the world's information. It acquires structured data spanning different domains such as music, people, and locations from various sources such as Wikipedia and MusicBrainz[8]. The data is aggregated and identical or related concepts are linked together. In addition, users in the community can add, edit, and even upload data. Topics in Freebase are organized by types which are grouped into domains. An important feature is that users not only can fill already predefined types with instance data or edit it, but also create their own types and define their properties, i.e. they can create new schemas and extend Freebase's domain model using the same interface. However, it provides an open but proprietary API for its data and encourages its use in applications and mashups.

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. It provides an RDF dataset extracted from Wikipedia, which contains mostly free text but also structured information such as categories, lists, info boxes, links to external pages etc. DBPedia makes it possible to ask complex queries (such as "*German musicians who were born in Berlin*") over a SPARQL query interface. DBPedia is a prime example of Linked Data publishing and can be browsed using semantic browsers. It is interlinked with other semantic datasets such as Geonames[9], MusicBrainz etc.

3 Social network applications with semantic technologies

3.1 Creation of metadata

Folksonomies are the primary sources of metadata on Web 2.0, however they have issues with consistency, ambiguity, synonymity. A next step beyond Web 2.0 is Semantic Web. It has been observed how folksonomy tags evolve into *property:value* triple-tags, which serve the same purpose as *subject property object* triple statements in RDF. This phenomena has even been called "poor man's RDF" [13]. And thus

3 http://opensocialweb.org/
4 http://microformats.org/wiki/social-network-portability
5 http://www.dataportability.org/
6 http://openid.net/
7 http://code.google.com/apis/opensocial/
8 http://musicbrainz.org/
9 http://www.geonames.org/ontology/

folksonomies move towards becoming lightweight ontologies. Social networks will have to provide more sophisticated means to directly create RDF metadata, and collaborative tagging will evolve into lightweight ontology development and integrate into collaborative modeling of the social network domain.

3.2 User interface

Much work is left on the issue on how to present semantic data to the user in applications, not to mention editing it [14]. There exists a number of semantic browsers, such as Tabulator[10], Disco[11], OpenLink RDF Browser[12], Objectviewer[13], Zitgist[14]. They are able to render generic RDF data for human users and navigate through different data sources through RDF links, just as conventional Web browsers navigate through HTML links. However, this kind of presentation most likely too advanced for mainstream Web users (see Figure 1).

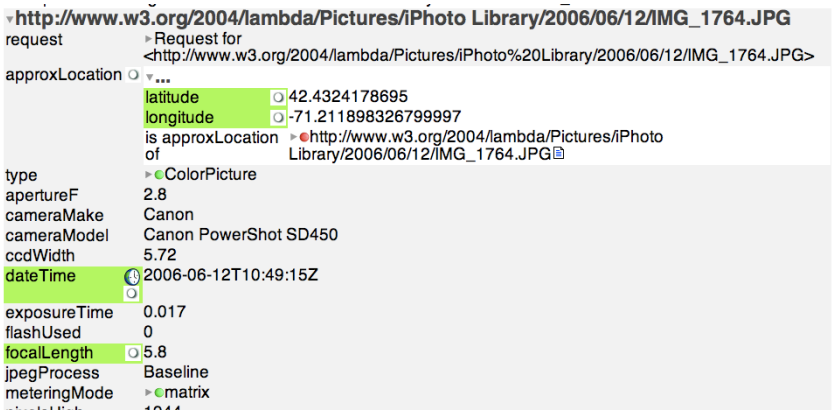

Figure 1 Tabulator view

We can assume that a Semantic Web application interface visualizes its domain ontology so that each class and instance would have its own page, linked to others through class-instance and instance-instance relationships. This approach, which we call *generic*, is used in many semantic websites, and probably best illustrated by Freebase. Another approach, which we call *specific*, is used by conventional Web applications, as well as social networks. Every type of information (such as a car, a user, or an event) has its own specific user interface. Each new type has to get a new interface, the same interface cannot be used for different types, and interfaces have to be fixed when the schema changes. This approach is obviously not feasible on the

10 http://www.w3.org/2005/ajar/tab
11 http://www4.wiwiss.fu-berlin.de/bizer/ng4j/disco/
12 http://demo.openlinksw.com/rdfbrowser/index.html
13 http://objectviewer.semwebcentral.org/
14 http://dataviewer.zitgist.com/

Semantic Web, where ontologies are meant to be extended, reused, and integrated from different sources. If social networks should become extensible semantic applications, it is likely that they should adopt this generic approach.

3.3 Domain model

Social network applications (Last.fm, Flickr, LinkedIn etc.) are usually developed for different application domains (music, photos, business). However, they share a common property: the domains are fixed and non-extensible. Users are encouraged to contribute and improve application data, but this is only limited to instance data for predefined types. Semantic applications such as Freebase take a different approach and allow users to edit the domain model itself: not only fill in instance data, but extend and edit types, add new ones, and define properties in the underlying ontology. Following this approach, social network applications would empower users to express their identities by creating or reusing concepts and relationships relevant to them, and share them with others. The domain model could be left to the community to control and to develop it in a direction which is currently of most interest to it, keeping it relevant over time. People would connect through things they have in common, achieving object-centered sociality [10]. This may be achieved in the future by integrating lightweight ontology development into the means of user collaboration and content contribution. To implement this approach, applications need to be modelled with semantic-enhanced languages such as: RDF/OWL, which offers more expressivity than object-oriented and relational models, is based on formal semantics and therefore interpreted unambiguously by different agents. Furthermore, they need to reuse FOAF and SIOC ontologies, which currently are state of the art representations of social networks on the Semantic Web, as well as other relevant ontologies.
Most of the current SW applications are also static and fixed in the sense that ontologies are known and mapped manually at design time [15]. Although semantic technologies are designed with extensibility and openness in mind, current programming languages and tools are not able to fully exploit it. It is expected that future semantic applications will be using multiple ontologies, discover them and integrate on request.

3.4 Publishing and reusing data and metadata

Large amounts of meaningfully interlinked RDF data available on the Web are crucial for achieving the Semantic Web vision. However, many social networks do not offer interfaces and APIs to access application data. Others make the contents of the website (such as lists of users, songs, or pictures) available via simple read-only REST interface in a software-processable data format, usually a custom schema of XML, Atom, or RSS. Some provide full APIs with add/update methods, invoked via various interfaces such as REST, XML-RPC, SOAP, Atom, or OpenSocial. A variety of publishing formats (especially non-standard) make reuse difficult. We argue that semantic social networks should publish their data in RDF, designed specifically for

distributed knowledge representation. Furthermore, all resources in social networks (including non-information, "real-world" resources) should be given URIs, distinguished from URIs of representations that describe them, and published as Linked Data[15]. APIs should be replaced by SPARQL endpoints, which would allow running remote structured queries against application data. Semantic data representation and advanced interfaces would help to overcome portability issues of proprietary APIs and interconnect social networks with different data sources, enable use of semantic browsers, and facilitate semantic mashups.

4 Prototype application design and implementation

The application prototype is a social network that allows users to browse events (such as concerts and conferences) and places of interest (such as venues and hotels) and find those that are most relevant to them. The prototype features a generic user interface. Users are able to browse OWL ontology classes and their instances and see properties with values, as represented in Figure 2.

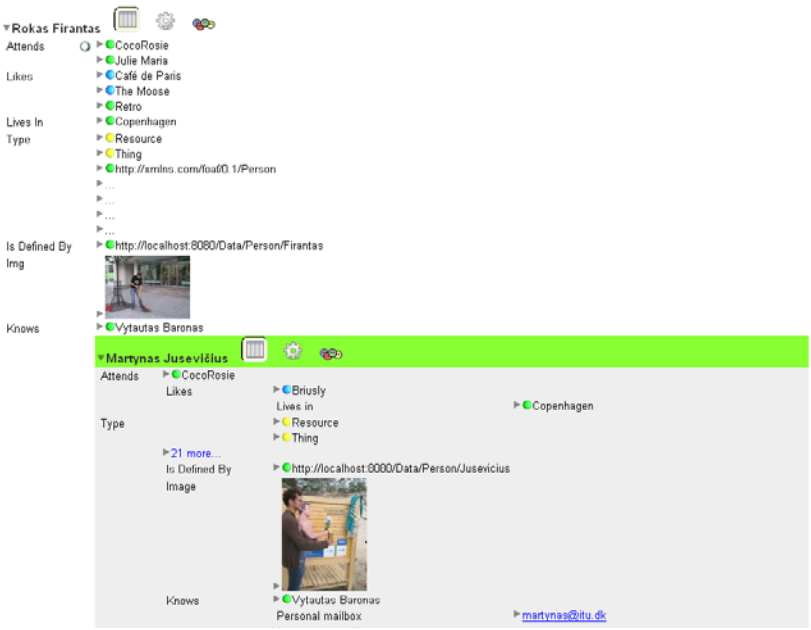

Figure 2 Browsing the ontology in a Tabulator view

The application is able to import a list of friends from an external FOAF source. It provides Linked Data access by serving interlinked RDF/XML which can be

15 http://linkeddata.org/

visualized and browsed using semantic browsers such as Tabulator or reused in other applications. It provides a SPARQL endpoint, which allows running structured queries. It also implements a semantic mashup: when a page of an instance is requested, the application queries remote DBPedia SPARQL endpoints, retrieves its description and homepage address in real-time and presents it to the user in the same fashion as local properties and values. Prototype is based on a RESTful Web framework which treats HTTP resources as first-class objects and follows a Model View Controller (MVC) pattern and W3C standards. Within the prototype the Model is the ontology layer, the Java code is the Controller and Views are generated by integrating SPARQL queries results and transforming them into XHTML using XSLT.The application domain is modelled as a RDF/OWL ontology, stored in a RDF triple store, accessed using Jena[16], and queried using SPARQL.

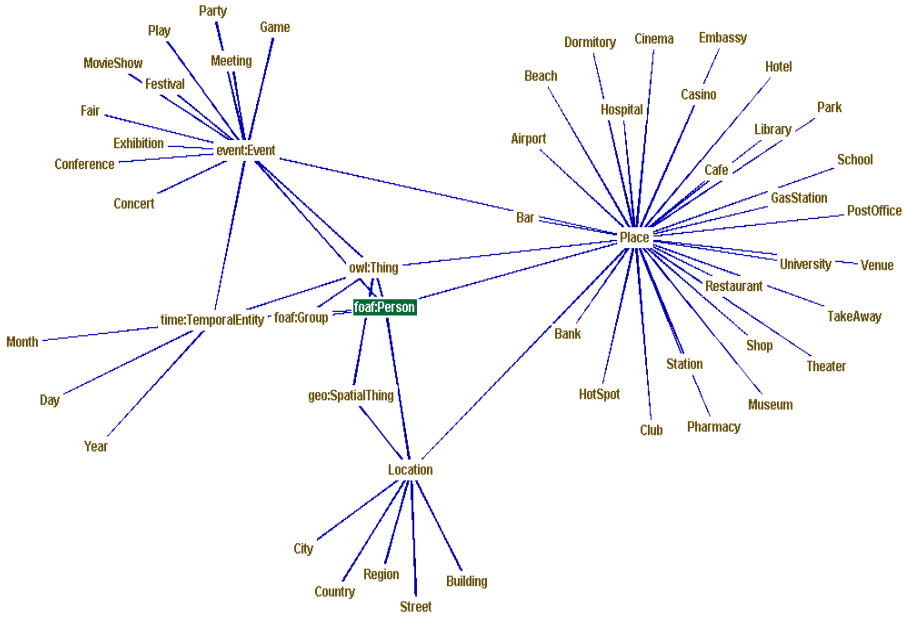

Figure 3 Domain ontology graphical view

The domain ontology is based on FOAF and SIOC. It is implemented in OWL and extends classes such as foaf:Person and adds a number of new classes such as: Place and Event as represented in Figure 3. FOAF and SIOC classes and properties were reused. Views become representations of REST resources (XHTML, RDF). They join several SPARQL XML results and transform them directly to output XHTML using XSLT, or serve raw RDF/XML for the Linked Data interface, depending on the HTTP *Accept* header. Controller dispatches requests to resources which have explicit

16 http://jena.sourceforge.net/

URIs, implement HTTP methods, can be related to domain instances using *foaf:topic*, and return view representations.

Most of the current object-oriented languages are statically typed and do not allow classes be changed or extended at run-time. I.e. it is not easy with the existing tools to map Event class in OWL to an Event class in Java so that it could be changed or extended at run-time. Tools such as RDFReactor[17] and Elmo[18] generate object-oriented Java code from our ontologies, but this code is static and not extensible at run time and therefore it was not used.

5 Conclusions and future work

Social data portability issues are leading to dissatisfaction in both user and developer communities. They are caused by limited amounts of social data published openly and lack of tools to import it, as well as formats and APIs of limited interoperability. Social networks would benefit from Semantic Web technologies. FOAF, SIOC, and Linked Data can solve portability issues and enable data reuse.

New generation of social applications could also take advantage of the advanced means to model data that SW technologies provide. Semantic data representation and advanced interfaces would help to overcome portability issues of proprietary APIs and interconnect social networks with different data sources, enable use of semantic browsers, and facilitate semantic mashups. Domain model could be collaboratively developed by users of the application. This approach requires a new generic user interface based on classes, instances, and properties. It could lead to more up-to-date and relevant content, which would in turn facilitate social connections through points of common interest. Other interesting directions that we have not yet pursued include AJAX-enabled application interface, a form-based interface for SPARQL, and dynamic, run-time object-ontology mapping tools.

References

[1] U. Bojars, J. Breslin, G.,, V. Peristeras, G. Tummarello, and S. Decker, "Interlinking the Social Web with Semantics," *IEEE Intelligent Systems,* vol. 23, pp. 29-40, 2008.

[2] L. Specia and E. Motta, "Integrating Folksonomies with the Semantic Web," in *Lecture Notes In Computer Science*, 2007, pp. 624-639.

[3] Z. Xu, Y. Fu, J. Mao, and D. Su, "Towards the semantic web: Collaborative tag suggestions," in *Collaborative Web Tagging Workshop at WWW 2006*, Edinburgh, UK, 2006.

[4] S. Angeletou, M. Sabou, L. Specia, and E. Motta, "Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report," in *ESWC 2007, Workshop on Bridging the Gap between Semantic Web and Web 2.0*, 2007.

17 http://ontoware.org/projects/rdfreactor/

18 http://www.openrdf.org/about.jsp

[5] P. Mika, "Ontologies are us: A unified model of social networks and semantics," *Journal of Web Semantics,* vol. 5, pp. 5-15, 2007.

[6] T. Finin, L. Ding, L. Zhou, and A. Joshi, "Social networking on the semantic web," *The Learning Organization,* vol. 12, pp. 418-435, 2005.

[7] L. Ding, L. Zhou, T. Finin, and A. Joshi, "How the Semantic Web is Being Used: An Analysis of FOAF Documents," 2005, pp. 113c-113c.

[8] J. C. Paolillo, S. Mercure, and E. Wright, "The social semantics of livejournal foaf: Structure and change from 2004 to 2005," 2005, pp. 69–80.

[9] J. G. Breslin, A. Harth, U. Bojars, and S. Decker, "Towards Semantically-Interlinked Online Communities," 2005, pp. 500–514.

[10] J. Breslin and S. Decker, "The Future of Social Networks on the Internet: The Need for Semantics," *IEEE Internet Computing magazine,* vol. 11 pp. 86-90, 2007.

[11] P. Mika, "Flink: Semantic Web technology for the extraction and analysis of social networks," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 3, pp. 211-223, 2005.

[12] L. Ding, T. Finin, and A. Joshi, "Analyzing Social Networks on the Semantic Web?," *IEEE Intelligent Systems (Trends & Controversies),* vol. 8, 2004.

[13] B. Andersen, "Meta Tags: The Poor Man's RDF?," http://weblog.scifihifi.com/2005/08/05/meta-tags-the-poor-mans-rdf/, 2005.

[14] A. Ankolekar, M. Krötzsch, T. Tran, and D. Vrandecic, "The two cultures: Mashing up Web 2.0 and the Semantic Web," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 6, pp. 70-75, 2008.

[15] L. Razmerita, A. Angehrn, and A. Maedche, "Ontology-based User Modeling for Knowledge Management Systems," in *UM* Pittsburgh, USA, 2003, pp. 213-217.