

Towards an Integrated Social Semantic Web

Antonio Maccioni

Università Roma Tre, Rome, Italy
maccioni@dia.uniroma3.it

Abstract

The Social Semantic Web is the data space on the Web where human produced information are enriched and modeled using Semantic Web standards. This vision is enabled by the convergence between Web 2.0 and Web 3.0 but still, it's far from be put into practice. Moreover, this miss causes drawbacks such as data redundancies and overhead in the development of social applications. In this paper we discuss how, in pursuing the Social Semantic Web vision, a different approach for opening social data to the Semantic Web should be adopted. Furthermore, without reinventing the wheel, we will use existing technologies to define a framework for integrating Social Web and Semantic Web. This would help to reengineer or extend the Social Network infrastructures in order to fully benefit from a social Semantic Web.

1 Introduction

Social networking provides services that we could not even imagine few years ago. A Social Network (SN) is commonly intended as a communication platform such as Facebook,¹ Twitter² or LinkedIn³ but, more generally, they are all the applications that foster collaborative and participative behaviour of profiled human users. Users in a SN can easily interact to each other in order to achieve a goal. For example, let us consider a Social CRM⁴ where customer care activities are managed by customers themselves.

The interactions within a SN produce huge amount of meaningful data. The aim to formalize the meaning of such data in order to allow machines to understand and process the data has led to the concept of Social Semantic Web (SSW) [3]. SSW inherits principles and methodologies from Semantic Web (SW), the global knowledge base where data is freely available on the Web and semantically organized through the so called ontologies of reference, described with RDF and “linked” to other data. The Semantic Web is fed by organizations and practitioners that transform or, in rare cases, produce natively RDF data to publish on the Web following the Linked Data principles [2]. This approach works well if conducted by Semantic Web experts, but it is impractical for non expert users (e.g., also SN users). Efforts have been made by people to make social Data available on the Web, but they make the data loose of authenticity and

¹ www.facebook.com

² www.twitter.com

³ www.linkedin.com

⁴ Customer Relationship Management

become easily obsolete. In fact, the common approach works well on static data (e.g., historical data) as it is noticeable from the Linked Open Data initiative⁵, but it is not the best option for dynamic and evolving social data.

For social data we can count, at most, on series of APIs that SNs expose to engage developers. In this way we can obtain social data that is usually structured by using JSON or XML formats. These data are rarely compliant to Semantic Web standards and, if published, generate ambiguities and replications. For example, let us consider when a user has to add the same personal information in all the social networks he registers to. Another paradoxical case is when SNs refer to each other as when we insert on the LinkedIn profile our Twitter ID and vice versa. By using the identification principle of Semantic Web, we could easily refer to the same resource identifier (i.e. an URI representing oneself) when inserting our personal information and an application can get from the Web our public information including related profiles. This is possible when the application knows a reference ontology (e.g., FOAF) to model and query such kind of information.

In this paper we review the current state of the Social Semantic Web, showing its main drawbacks and we depict a framework where Social Web and Semantic Web are integrated with each other using existing technologies. Very little efforts can be made by SN players to implement the forementioned framework. These efforts may consist on: (i) refining data and application standards for SNs, (ii) extending and adopting existing ontologies and (iii) deploying a Semantic Web infrastructure for each SN. Then, this framework can be used to reengineer SN infrastructures in order to natively take advantage of the Semantic Web. Note that, at the current state of the art, there are interesting attempts, such as the Social Graph⁶ (promoted by Facebook and implemented through the Open Graph), to enrich the SSW but they are not very effective as our proposed framework.

The main benefits to share a SSW framework of reference are:

- easier integration among SN applications;
- higher social data quality;
- empowering of Social Network Analysis (SNA) that is the area which tries to infer information from SN data (e.g., a classic example of SNA is the problem to find, over a SN, a team of experts on a certain topic). If the SN data are available in a public space and connected to other related available data, SNA can exploit the surrounding information to infer more knowledge;
- enabling the creation of services in the context of smart cities, where applications exploit data in a smart way to provide useful services to citizens. Let us imagine services like automatic recruitment (that can be done by matching people's info with job requests), personalized event or activity recommendation, etc.

We will show, by analysing real case scenarios, the validity of the proposed solution.

⁵ lod-cloud.net/

⁶ en.wikipedia.org/wiki/Social_graph

The rest of the paper is organized as follows. Section 2 discusses the current state of the Social Semantic Web pointing out its drawbacks. In Section 3 we illustrate the proposed framework to integrate Web 2.0 with Web 3.0 and in Section 4 we show how this SSW framework is beneficial in real case scenarios. Finally, Section 5 sketches conclusions and future works.

2 The current Social Semantic Web

In this section we review the current state of the Social Semantic Web in the perspective of whom, in order to design new services, has to exploit SSW data.

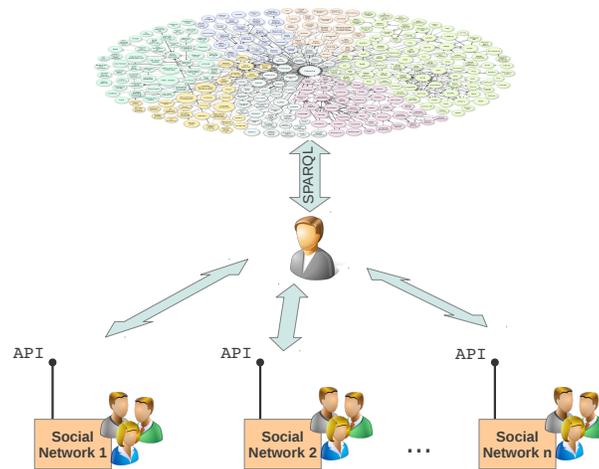


Fig. 1. The current Social Semantic Web.

Fig. 1 helps to understand the scenario. We have many social networks (i.e. SocialNetwork_1 , SocialNetwork_2 , \dots SocialNetwork_n) providing a set of APIs. An user (i.e. in the middle of Fig. 1) can use these APIs to retrieve social network's data. The user can also exploit the Semantic Web (i.e. the graph cloud on the top of Fig. 1) information and integrate them with API calls' results. Actually, the user is implementing an ad-hoc integration of the Social Semantic Web consisting in the following steps:

1. query all the SNs through API calls and retrieve data from the SW (e.g., through SPARQL);
2. transform data into a common interoperable format;
3. integrate data by using matching and similarity techniques.

For instance, let us consider the concrete case in which a user wants to mashup social and encyclopedic information from three data sources: Flickr⁷,

⁷ www.flickr.com

Facebook and DBPedia⁸ [1] (i.e., the linked data extracted from Wikipedia⁹). To do so, we have to query the Flickr APIs¹⁰ to return data in one of the following formats: REST, XML-RPC, SOAP, JSON or serialized PHP. Facebook data can be extracted in several ways: with the Open Graph APIs¹¹ or through the FQL¹² (Facebook Query Language). The user can obtain DBPedia data from its SPARQL endpoint¹³ in different formats (e.g., XML, CSV, JSON). Finally, the integration of all the data is computed to employ the service.

It is easy to notice that the computational load of this process is heavy and the software engineer has to learn different API usages. He could use a single SPARQL query over the Semantic Web to obtain in one shot all the useful data for the service. This integrated vision will be explained in the next section.

3 The Integrated Social Semantic Web

In the previous section we have seen that, to exploit the current SSW, it is required to go through many subsequent steps. To avoid this long chain of steps, volunteer users transform data that they get from SN APIs into RDF data, that then they publish on the Semantic Web.

This approach, led by volunteer users, fails under many aspects:

- social data evolves faster than these transformed data;
- infrastructures provided by volunteers cannot manage high computational load;
- most of the times a single user does not provide a “fully furnished” SW platform (e.g., exploiting content negotiation or providing SPARQL endpoints);
- data coming from different social networks are, sometimes, not interoperable.

This section proposes an integrated framework for the Social Semantic Web as depicted in Fig. 2. In this framework, each SN acts independently from the others but the SSW data is natively produced and automatically integrated.

The technologies and tools that we use for this framework are:

- OpenSocial¹⁴ : it is a formal model to represent social application primitives. It was initially proposed by Google¹⁵ and MySpace¹⁶ in order to standardize back-ends of social applications.
- Apache Shindig¹⁷ : it is the reference implementation of OpenSocial. It is a container for hosting social applications (both client and server side) which

⁸ dbpedia.org/

⁹ www.wikipedia.org/

¹⁰ www.flickr.com/services/api/

¹¹ developers.facebook.com/docs/reference/opengraph/action-type

¹² developers.facebook.com/docs/technical-guides/fql/

¹³ dbpedia.org/sparql

¹⁴ opensocial.org/

¹⁵ www.google.com/

¹⁶ www.myspace.com/

¹⁷ shindig.apache.org/

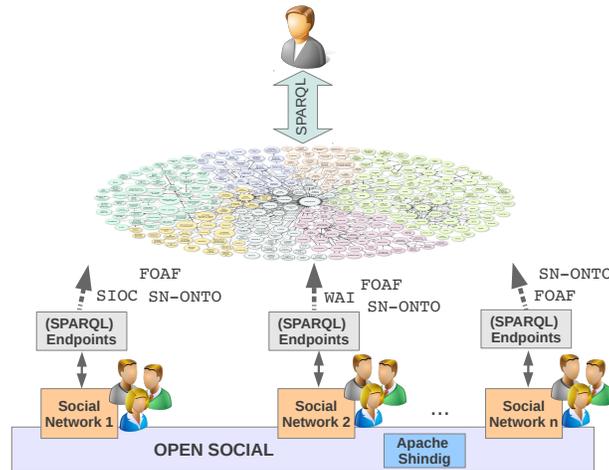


Fig. 2. The Integrated Social Semantic Web.

implements primitives to manage: user profiles, user relationships, user activities, user authentications, user authorizations, etc.

- FOAF ontology: is a very famous ontology to represent personal data and relationships among people.
- WAI (Who Am I) Ontology¹⁸ : is an ontology that extends FOAF considering that a person can have multiple roles and multiple profiles.
- SIOC [4] ontology: is an ontology to represent information of online communities (blogs, wikis, forums, mailing lists, etc.) and their relationships.
- VCARD ONTOLOGY¹⁹ : is an ontology derived from the IETF standard for business cards and, therefore, it allows to represent personal information such as company affiliations.
- SPARQL endpoint: is a Web service that is able (i) to receive SPARQL queries over HTTP, (ii) to process them and (iii) to convey the answers back to the user.

Social Network players should agree on this framework and on these technologies in order to extend their own existing infrastructures. This is analogous to the schema.org²⁰ project where search engine players formed a consortium to define common metadata to embed within Web pages (by using microformats or RDFa) in order to make understand the Web contents to the crawlers.

By adopting this framework, SN players can natively feed the Semantic Web with their social data. Still, a little effort has to be made in the development of a common model to represent Social Web data.

¹⁸ vocab.ctic.es/wai/wai.html

¹⁹ www.w3.org/TR/vcard-rdf/

²⁰ schema.org/

Let us call this model *Social Network ontology* (`sn-onto`). We propose, in order to agree on the specifications of `sn-onto`, the way tracked by SIOC-A (SIOC in Action) [5]. SIOC-A extends SIOC considering active communities. Basically, it embeds the concept of *action* (i.e. `sioca:Action`) within SIOC. Since OpenSocial primitives already consider the dinamicity and the actions performed within a social application, the Social Network ontology should join (re-using the existing concepts, when possible) concepts from: SIOC, SIOC-A, FOAF and the OpenSocial primitives. Note that an activity-centric ontology reflects the dinamicity of SNs and SN data, in opposition to the staticity of other domains where the current approach to open data on the Semantic Web works well. In particular, the elements connected to `sioca:Action` should not be general artifacts as in SIOC-A, but social network resources, e.g., instances of a possible class `sn-onto:SNResource`. Then, each SN can implement subconcepts of `sn-onto:SNResource` according to its domain. For example, these resources can be posts, tweets, etc. Similarly, also the actions (i.e. `sioca:Action`) can be specialized according to the domain of reference, e.g., like, comment, retweet, etc.

Furthermore, SNs could exploit the SSW very smartly, exploiting and computing linking of data already on the Web. For example, let us consider a SN user that likes a Wikipedia page. From this action, we can automatically extract the DBPedia [1] resource corresponding to the Wikipedia page and then link the SN user URI to the DBPedia resource by using a property of `sn-onto`. Then, for example, another SN can exploit this information to recomend readings to the user.

A similar case can happen georeferencing objects: they can be automatically linked to geographic datasets such as Geonames²¹. Note that these linkings would be impossible to obtain if left to be generated by single users.

4 Real Case Scenarios

In this section we consider real case scenarios in order to validate our framework with a proof of concept. In particular, two scenarios will be considered as follows.

4.1 Music Domain

Nowadays many musical social networks and musical applications exist: Last.fm²², Spotify²³, Jamendo²⁴, SoundCloud²⁵, Deezer²⁶ just to mention the most famous. They have data representing artists (name, origin, members, genre, etc.) and their production (albums, songs, etc.). All of these SNs provide

²¹ www.geonames.org/

²² www.lastfm.it/

²³ www.spotify.com

²⁴ www.jamendo.com/

²⁵ soundcloud.com/

²⁶ www.deezer.com/

APIs to developers in order to make them implement additional features. Even if some of the SNs built its own database gathering metadata of user's MP3 files, none of those SNs provide directly their datasets.

The DBTune project²⁷ has published on the Semantic Web data extracted from Last.fm and Jamendo. The project has also published an RDF version of MusicBrainz [10], the open music encyclopedia, that is now considered an hub within the SW, since several music datasets refer to it. DBTune datasets have been linked to other famous datasets such as Geonames and DBPedia.

Another important step in the music domain of the Semantic Web has been already accomplished: the creation of an ontology of reference, the Music Ontology [9]. DBTune already uses this ontology to model the data.

In this scenario, we can notice that each SN creates the database from scratch and that the content of these databases are not aligned. Moreover, they do not make use of the music ontology. Clearly, all of them would benefit from the use of a common dataset with a common data model. For example, since they exploit users listenings in order to provide suggestions on similar artists, they could, sharing the data, recommend an artist that a user has listened from another social network.

4.2 Research Publications Domain

There exists social networks that make use of scientific publications, e.g., LinkedIn (the SN for professionals) and ResearchGate²⁸ (the SN for researchers). The author has to add his publications manually in each SN.

This is quite absurd because there exist a large number of databases of scientific publications that could be exploited: Arnetminer²⁹, DBLP³⁰, the ACM repository³¹, the IEEE repository³², Google Scholar³³. All of them work as data silos and are not officially shared on the Semantic Web. Moreover, they are also highly inconsistent among each other as we can notice by the number of publications that an author has in each database. As a result, for each author, the number of citations and parameters such as the h-index change across the databases.

Some of the above mentioned databases have a non-official RDF version, but if there was a common official database on the Web, publications and authors could be addressed with the same URI and exploited by all the applications, including SNs.

²⁷ dbtune.org/

²⁸ www.researchgate.net/

²⁹ arnetminer.org/

³⁰ www.informatik.uni-trier.de/~ley/db/

³¹ www.acm.org/

³² www.ieee.org/

³³ scholar.google.it/

5 Conclusions and Future Work

In this paper, we showed that existing technologies can be employed to pursue the vision of the Social Semantic Web. In this vision Social Web and Semantic Web are integrated to exploit benefits of both of them at the same time. There exists other aspects of the SSW left outside the framework presented in this paper such as the management of folksonomies (e.g., folksonomies generated from SNs, such as delicious³⁴, to be integrated automatically on the SW) or the semantic extraction from text [8] (e.g., extraction of knowledge from SNs such as Quora³⁵). They will be considered in future work. Other future work are represented by the adaptation of new emerging applications based on social data such as RandomDB [6] or the use of social aspects applied to existing Semantic Web problems such as the keyword search [7].

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.G.: Dbpedia: A nucleus for a web of open data. In: ISWC/ASWC. pp. 722–735 (2007)
2. Berners-Lee, T.: Design issues: Linked data (2006)
3. Breslin, J., Passant, A., Decker, S.: The Social Semantic Web. Springer-Verlag, Heidelberg (2009)
4. Breslin, J.G., Harth, A., Bojars, U., Decker, S.: Towards semantically-interlinked online communities. In: ESWC. pp. 500–514 (2005)
5. Champin, P.A., Passant, A.: Sioc in action representing the dynamics of online communities. In: I-SEMANTICS (2010)
6. De Virgilio, R., Maccioni, A.: Generation of reliable randomness via social phenomena. In: MEDI (2013), to appear.
7. De Virgilio, R., Maccioni, A., Cappellari, P.: A linear and monotonic strategy to keyword search over rdf data. In: ICWE. pp. 338–353 (2013)
8. Gangemi, A.: A comparison of knowledge extraction tools for the semantic web. In: ESWC. pp. 351–366 (2013)
9. Raimond, Y., Abdallah, S.A., Sandler, M.B., Giasson, F.: The music ontology. In: ISMIR. pp. 417–422 (2007)
10. Swartz, A.: Musicbrainz: A semantic web service. IEEE Intelligent Systems 17(1), 76–77 (2002)

³⁴ delicious.com/

³⁵ www.quora.com/