The influence/impact of Semantic Web technologies on Social Media

La influencia/impacto de las tecnologías de la Web Semántica en las Redes Sociales

Alfonso Infante-Moro¹, Anca Zavate², Juan-Carlos Infante-Moro¹

¹ GITICE research group. University of Huelva, Spain
² University Alexandru Ioan Cuza, Iasi, Romania

alfonso.infante@uhu.es, anca.zavate@gmail.com, juancarlos.infante@uhu.es

ABSTRACT. The paper is based on a literature review of Semantic Web technologies and evaluation of how these technologies can be applied in the Social Media domain, whose purpose is to underline and to understand the way how Semantic Web technologies can influence and improve Social Media, and be relevant to effective searches (for example, for companies and their human resources departments).

The Semantic Web technologies allow data collecting about a certain subject from the entire Internet, and Social Media allows people to share information about them, to express their opinion and to influence others. Therefore, participation and collaboration for the selection of useful information from the Web is sought. But we found limitations for success in the high cost of apps that can serve as connecting points for these social networks (Facebook, LinkedIn, Google+, Twitter, etc.) and in the privacy of the shared content on these social networks.

KEYWORDS: Semantic Web, Social Media, Semantic Social Media, Facebook, Twitter, Google+.
1. Introduction

At first, the Web was seen as a package of Web sites which were offering a collection of Web documents and the purpose was to share that information. The Web was in one way only, users were able to only read the information and use it. Currently, the Web is everyday more and more interactive. At first, there was the content generated by the user on the Internet. Moreover, now users are able not only to read information, but also they use the Web to create content and interact through social networks (Yu, 2011). Also, the Web sites are offering more functionalities, from the payment of bills online, buying plane tickets, clothes to complex financial transactions.

A lot of Web sites have started to publish structured content so different business entities can share their content in order to gather more online transactions. For example, Amazon and eBay are publishing structured data through Web services so other applications can be build on top of this layer.

2. Functionalities of the current Web

The Web search is the most common use of the Web. The purpose is to find the information you are looking for. Finding different recipients, searching for classes, train programs, are all examples for Web searches. But, searching on the Web can become difficult. For example, the search by an acronym using a search engine can display results from many domains and not only from a certain domain. Only after selecting the information from many results you can find the wanted information. This happens because the search engines have implemented the concept to search after key words. Therefore, as long as a document contains the key word it will be included in the sets of results that are shown to the user. After the user reads and interprets the results, he can extract the needed information.

Information Integration is another functionality of the current Web. A company can provide structured data for other applications. It does not matter the programming language used as long as the standards of integration are checked. For example, the search of a location for vacation can be a challenge. First you have to look for the location, then for the hotels in the area, the weather. The challenge is that you cannot find all this information in one place. This can be accomplished with the use of web services.

Sharing information through social networks. Now the users can create content and share it.

Data mining means the extraction of useful information from a big data set or database. Therefore, the Web is seen as a giant distributed database, the concept of data mining is referring to the activity of extracting useful information from the Web.

Social Media is represented by the socializing Web sites. Social Media defines a new era of communication, a new way to connect people, with the purpose of sharing information, learning. The impact of online communication can be seen in a various of domains, like business, education, tourism and day to day activities (Infante-Moro, Martínez López & Infante-Moro, 2015; Infante, Martínez, García & Infante, 2014a). And all thanks to technological advances (Infante-Moro, Infante-Moro, Martínez-López & García-Ordaz, 2015; Infante, Martínez, García & Infante, 2014b).

The problem comes from the fact that the goal of the current Web are the human readers and it’s strictly oriented on the interface. To develop and invoke the useful Web services automatically, the first step is to define them in more efficient way. Currently, this kind of integration it’s hard to implement, because finding the components of a process isn’t efficient. From the point of view of the Web, all this components are equally created and there isn’t a way to teach the computer the meaning of every component.

3. Defining the concept of Semantic Social Media

Social Media is a new form of big data, which creates a new way of aggregating and gathering of data so
business can understand what consumers and potential consumers are saying about them, who is buying the product, who isn’t and why? (Guess, 2014). A limitation of the current social websites is that they are isolated from one another, like separated entities. Different websites can contain complementary information. For example, information about a laptop, you can find comments on Twitter, pictures on Flickr, and videos on Youtube, but cannot find all this information in one place. The Social Web creates a set of single data silos that cannot interoperate with each other, where synergies are expensive to exploit, and where reuse and interlinking of data is difficult (Vrandei, Passant & Breslin, 2012).

The Semantic Web provides all the technologies and standards to add content understandable by machines, so in this way computers could understand Web documents and do tasks that would normally be done manually.

Semantic Web allows machines to understand and to satisfy the requirements of the users, through processing of content and semantics of the information. To the current Web it is added the responsibility of the machine. Computers become entities which extract and interpret information, and are not only posting devices (Fensel, Facca, Simperl & Toma, 2011).

The concept of Semantic Social Web is referring to the fact that social interactions on the Web lead to the creation of a rich content. Semantic Social Web can be seen as a collection of collective knowledge, which can provide useful information based on users contribution. The purpose of Semantic Social Media is to complete the vision of Semantic Web by adding a paradigm which is based on the description of technologies for the semantic browsing. A social-semantic system it’s an ongoing process of new information for a certain domain through the use of ontologies and taxonomies. Therefore, the importance of semantic created by human users is growing. Instead to be based entirely on automated content with a formal ontology, humans build in a collaborative way the semantic given by the social-semantic information systems. While the Semantic Web allows the integration of business processes with a precise automated logic, Semantic Social Web is open to social interactions of semantic business process, allowing the interoperability between the business objects, actions and users (Wikipedia, 2013).

Stephen Downes and Marco Neumann have acknowledged the term of Semantic Social Networks in order to describe the application of Semantic Web technologies and online social networks. Social Networks are compound of people or groups that are connected by a set of social relationships, such as friendship, co-working or information exchange (Finin, Ding, Zhou & Joshi, 2005). Social Networks and the technologies of the Semantic Web can add a new concept for social Web analysis by combining human intelligence with machine processing. Social Web brings data at a large scale, along with dynamics and precision impossible to reach from observing traditional communication. To get a full advantage Semantic Web provides the key in gathering information across heterogeneous sources, it incorporate user-generated metadata and other information left by the user (Mika, 2007).

While most of the social networks are based on relationships, semantic social networks are based around a certain topic. Semantic technologies and languages of processing can extract automatically the content of social Web conversations, for example Twitterrizer for Twitter. Another example of this kind of network is Scitable, from Nature. The purpose of Scitable is to gather undergrads, postgrads and university faculty around specific topics, with the initial emphasis on genetics. Scitable is an online textbook, a semantic social network that brings together researchers, post docs, students, teachers in one global community connected by a passion for teaching and learning science. As the director of Nature Education, Vikram Savkar, is saying in a podcast, the Semantic Social Media is the second generation of Social Media. If the first generation was focused in online socializing, the second generation sees the social network rather as a tool for learning, innovating, with a real purpose, where there can be built in tools and languages (Tobin, 2009).

The goal for the Semantic Social Media will be the interlinked documents, data, social interactions, all described using machine-readable formats. Users would be able to find information about a certain topic from

1 https://www.flickr.com/
4. Semantic Web technologies that can be used for Social Media

The Semantic Web technologies and languages of processing are able to extract automatically the meaning of social conversations (Twitter, Facebook, LinkedIn, etc.). Social media analysis adds much more than aggregated and retrospective report on brand evaluation. Among big sets of conversation, exists valuable signals, both opportunities and also threats, like competition, unsatisfied customers (Daedalus, 2013).

The technologies of Semantic Web are structured in layers. The first layer is URI\(^2\), then the XML\(^3\) and RDF\(^4\), next is the query language SPARQL\(^5\), RDFS\(^6\), ontology: OWL\(^7\), rule RIF\(^8\), then proof and trust and ultimately The User Interface and Applications. Most of these technologies can be applied to the area of Social Media in order to get full advantage of all the valuable information that is created and shared by users. In order to describe profiles, content and connections and bind them together, so social media sites can interoperate by appealing to common semantics is possible through the use of technologies specific to Semantic Web and ontologies.

URI is the acronym for Uniform Resource Identifier, that is a compact string of characters utilized for identifying or naming a resource (Wikipedia, 2014).

XML (Extensible Markup Language) is a markup language that defines a set of rules for documents readable across platforms. XML is used to store and share data. The difference between XML and RDF – Resource Description Language – is that XML is a syntax and RDF is a data model. RDF has several syntaxes and XML is one: RDF/XML (Sequeda, 2012).

RDF (Resource Description Framework) is the modeling language of data for Semantic Web. Therefore, RDF is a platform for representing the information on the Web. Semantic information is stored and shared with the use of RDF, making easier the merger of databases even if the schemas are different. Using this model, structured and unstructured can be used together in different applications (RDF Working Group, 2014).

RDF uses the next key concepts (Klyne& Carroll, 2004):

- The graph data model. The structure of RDF expressions is a triple expression, named a RDF graph, which contains a subject, a predicate and an object. This is illustrated, in the next picture, as a diagram with two nodes and an arrow between:

  ![Figure 1. The RDF Graph.](image)

- Vocabulary based on URI. An URI reference used like a predicate identifies a relationship between the connected nodes.

---

\(^{2}\) Uniform Resource Identifier

\(^{3}\) Extensible Markup Language

\(^{4}\) Resource Description Framework

\(^{5}\) SPARQL Protocol and RDF Query Language

\(^{6}\) RDF Schema

\(^{7}\) Web Ontology Language

\(^{8}\) Rule Interchange Format
- **Data types** – are used by RDF to represent values as integers, float numbers and date. A data type consists of a lexical space, a space value ({{T, F}}) and a lexical value mapped ({{"true", T}, {"1", T}, {"0", F}, {"false", F}}). For instance, the lexical value mapped for the schema: xsd:boolean.

- **Literals** – are used to identify values as numbers and dates by their lexical content. Anything represented as a literal can also be represented as an URI. A literal can be an object of a RDF declaration, but not the subject or the predicate.

RDF represents for data what HTML represents to Web pages. HTML is a standard for publishing content on the Web. RDF functions as a common data model (Sequeda, 2011). According to W3School RDF is a standard for describing Web resources, can be used to describe title, author, content, etc. RDF is written in XML and is part of the W3C Semantics Web Activity. With the use of XML, information in format RDF can be exchanged very easy between different platforms and applications. The document http://www.w3schools.com/rdf could be described in RDF format as follows (w3schools.com, 2014):

![Diagram of RDF format](image)

Figure 2. Example of RDF format.

In order to get information from the Web you need a query language. **SPARQL** - Protocol and RDF Query Language is an RDF query language, that capable to retrieve and manipulate data stored in Resource Description Framework format (Wikipedia, 2014).

**OWL** - Web Ontology Language is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things (W3C, 2011).

5. Analysis of applications already existing in the market

If we talk about Semantic Web in Online Social Networks, we talk about collaborative-shared tagged systems (folksonomies) in each of them. These tags are used to classify/categorize Internet content and sort, search and find all the information related to these resources contents through these tags. Each tag can be bound to new contents that others have tagged the same way and interact with them (social tagging) or respond to certain social affiliations of their members. Therefore, we're talking about tools for content management of these Online Social Networks, providing a data mining (González, Labra & Álvarez, 2012; Cernea, Del Moral & Labra, 2007).

Nowadays, all this Social Semantic Web is collected on new apps of the great giants of Online Social Networks (we focus on the study of the top 3: Google, Twitter and Facebook), apps with meta tags that indi-
cate what information is displayed in their available fields (Abarca, 2013):

**Google Authorship**

This app of Google consists an authoring system, which allows an author to link all his own content published on the Internet with his Google+ profile, regardless of where it has been published.

It displays the Google+ profile and image of the author in the results of Google, if his web page or blog appear in them. In addition, this also improves the positioning and increases brand image of the author and the number of visits to his content. At short, this application provides authoring, visibility and clicks (Google, 2014; Gil, 2013).

The metadata used are author and URL of Google+.

**Open Graph**

This is a Facebook app for the personalization of the platform, so that the interests of users about leisure, music, sports,... can be available and perfectly organized in it. To do this, Facebook requires that users agree to share their privacy. Additionally this app allows web pages can have the same functionality as the contents published on Facebook.

It consists of a combination of plugins that allow you to see who consume that service, being a consumer and registering it in the user profile, consuming all its activity and receiving recommendations to other services. The button 'like' is the cause of all this. (Iskold, 2010).

This app is based on RDF, and metadata used are (Open Graph protocol, 2014):

- Basic Metadata: title, type, image (image URL), url.
- Optional Metadata: audio, description, determiner, locale, locale:alternate, site_name, video.
- Object Types:


This metadata can describe what object is on that page (film, food, business, book, artist, event,...), and this allows categorize these objects and provide a basis for other applications (recommendations or other applications). For the analysis and publication of this metadata, this app is supported by the following tools (tools that allow and make the whole process) (Open Graph protocol, 2014):

Facebook Object Debugger, Google Rich Snippets Testing Tool, OpenGraph.in, PHP Validator and Markup Generator, PHP Consumer, OpenGraphNode in PHP, PyOpenGraph, OpenGraph Ruby, OpenGraph for Java, RDF-RDFa-Parser and Wordpress plugin.
**Twitter Cards**

In the case of Twitter, we find Twitter Cards, this app allows more than 140 characters in a tweet. An additional content (description of the link of the tweet, title, videos, images, audio,...) through tags that make Twitter can extract it and display it. This allows a preview of the content of the tweet. (Sánchez, 2013; Carrodeguas, 2012).

There are 7 card types (Twitter Developers, 2014; Rocafull, 2014):

- **Summary Card**: Tweets with a title, description, image and Twitter profile.
- **Summary Card with Large Image**: It's like a Summary Card, but can to offer the larger image.
- **Photo Card**: Tweets with big photo.
- **Gallery Card**: Tweets with a collection of photos.
- **App Card**: Tweets with a profile of an application.
- **Player Card**: Tweets with video/audio/media player.
- **Product Card**: Tweets to better represent product content.

These cards cause more and better communication with the fans, and the choice of each one of them depends on the feature to highlight in each tweet. We can say that its purpose is the communicative.

The tags that can appear in Twitter Cards are (Twitter Developers, 2013):

"Card (card type), site (@username of website), site:id (Same as ‘site’, but the user’s Twitter ID), creator (@username of content creator), creator:id (Twitter user ID of content creator), description (Description of content), title (Title of content), image:src (URL of image to use in the card), image:width (Width of image in pixels), image:height (Height of image in pixels), image0, image1, image2, image3 (1st, 2nd, 3rd, and 4th image in the gallery, respectively), player:HTTP URL of player iframe), player:width (Width of iframe in pixels), player:height (Height of iframe in pixels), player:stream (URL to raw video or audio stream), data1 [Top customizable field, can be a relatively short string (ie "$3.99")], label1 [Customizable label or units for the information in twitter:datal (best practice: use all caps)], data2 [Bottom customizable data field, can be a relatively short string (ie "Seattle, WA")], label2 [Customizable label or units for the information in twitter:datal (best practice: use all caps)], app:name:iphone (Name of your iPhone app), app:id:iphone (Your app ID in the iTunes App Store), app:url:iphone (Your app’s custom URL scheme), app:name:ipad (Name of your iPad optimized app), app:id:ipad (Your app ID in the iTunes App Store), app:url:ipad (Your app’s custom URL scheme), app:name:googleplay (Name of your Android app), app:id:googleplay (Your app ID in the Google Play Store), app:url:googleplay (Your app’s custom URL scheme)."

This metadata can describe what object is on that page, making a more detailed description of what users will find the other side of the link. This allows a preview of these objects.

Twitter Cards parameters have an equivalent format to those of Open Graph, as they are based on the same protocol. Twitter can get to read up to 5 Open Graph metadata: description, title, image:src, image:width, image:height. (Sánchez, 2013).

**Hashtags**

Finally, it is important to name this tool that is based on tags that are created by the author himself and describe and spread resources, and that are used to classify a resource (for identification on a topic) and unite users with same interests. Both Facebook and Google+ have copied this tool of Twitter, besides other Online Social Networks (Rivero, 2011).

These hashtags are character strings formed by a pound sign (#) and one or more concatenated words.
(example: #SemanticWeb) that can be searched by users and whose results will be all the resources that have the same hashtag (in this example: #SemanticWeb).

With all this, we can say the apps that we find about Semantic Web in Online Social Networks allow categorize resources posted by their users and provide them with a authoring, but it is difficult to obtain and select information about a particular person when we delimit it to professional experience, preference for music or film, ... because of the way how the data is distributed in the different applications of the various Online Social Networks.


Web search is a key technology of the Web and it’s the primary way to access information from the variety of data on the Web. Current search technologies are based on key words. Web search will change radically once with the development of powerful tools like Semantic Web, which is a common platform that allows data to be distributed and reused in different applications, companies and communities (d’Amato, Fanizzi, Fazzinga, Gottlob & Łukasiwicz, 2010). Semantic Web Search has the purpose to make the information to be understood by machine, thought the use of ontologies in order to annotate the content from the Web (Kerrigan, Mocan, Simperl & Fensel, 2009).

These ontologies are languages needed (protocols) to get information that can be interpreted by computers without human intervention, it is necessary that the content of web pages is encoded by ontologies to be interpreted by them. Ontologies define the content of the different domains and their relationships, which allows reaching deductions to computers and using it, for example, to find professionals that are within a certain distance if you use the right web agent (Ponce Toste & Fernández Peña, 2012; Tello Lozano, 2001; Studer, Benjamins & Fensel, 1998; Gruber, 1993).

Currently, all Semantic Web is based on ontologies and labels, these ontologies are conceptual schemes that are designed to facilitate communication between different systems and these tags are used to structure the content of web pages and are based upon ontologies. One way in which, computers are able to connect web content and tags and, consequently, further adjust the search to what the user requires. This technology is also being used in social platforms (Online Social Networks), and areas such as education, industry, R&D and e-commerce is developing specialized models based on it (InterBlog, 2014).

With classic search you get the result list with everything that contains the typed keyword. Semantic Search considers also the context of the search and also the meaning of the documents that are searched (Gonzalez, 2012). Others consider that there are two kinds of Semantic Search engines: Ontology Search Engines— are using ontologies to search within Semantic Web documents and Semantic Search Engines – that use Semantic Web technologies to improve the search result (Esmaili & Abolhassani, 2006).

Semantic search means that a search engine is able to determine what you mean when you search for something and list right the results without entirely matching with the words used in the query search. Recently, Google added phrases to result pages – the Knowledge Graph, but their search still depends on keywords and not semantics. Also Facebook added the search Graph, where you can add option when you search for a specific group of people. For instance, people that live in a certain city, they studied and economics and they work at a bank. Also with the classic search if you look up the word java for example the search engine would not know if you refer to the coffee or the programing language and here semantic search adds the context.

Semantic search resolves the ambiguities from searching by a keyword by specifying the kind of result that are required. Three types of search can be supported: search by entities, search by relations and parameterized searches (Uren, Lei, Lopez, Liu, Motta & Giordanino, 2007). Semantic search is all about operating in an envi-
environment where the meaning of particular symbols need to be defined. For example, if an employer wants to find the perfect person to hire, he needs to know besides the technical knowledge also what hobbies the future employee has, what he does in the free time, so he can also be part of the team socially. The problem is that you cannot find all this information in a single Web site, if you want personal information you have to look on Facebook, if you want information about his knowledge you need to check on LinkedIn and so on. A semantic social search application would resolve that.

This new solution for searching on the Web, the semantic search, promises to change the way we use the Web. Instead of searching by keywords, will search in a specific environment and result set will contain only useful information.

7. Ways to develop a Semantic Web application. Use case scenario

A Semantic Web application can determine the meaning of the text and create relations for users. Another key function of Semantic Web applications is data portability and connectivity and using the Web as a platform. There are two approaches for development (MacManus, 2007):

- Bottom up – which need semantic annotations for data (meta-data)
- Top down – which consists of analyzing the information, like a natural language processor, which is able to understand text like people do.

In the picture bellow are illustrated different approaches for developing a Semantic Web application.

![Diagram](image)

Figure 3. Different approaches to Semantic Web (MacManus, 2007).

The three standards used by Semantic Web, RDF, SPARQL and OWL, are all open-ended. Open-ended refers to a specific issue at different levels and times, for example, the volume of transaction, use of data, types of users. Therefore, Semantic Web technologies would be suitable for (Gonzalez, 2012):

- Complete Data Model Unknown: there is no knowledge about the size of data in the future, for example for now you keep track only for customers and orders, but in the future you will also keep track about the marketing department.
- Complete Usage Model Unknown: the views and reports needed by the users are unknown.
- Complete User Base Unknown: the category of users that will use the application is unknown. New data and new usage of data can be added.
**Goal of the application**

First the goal of the application must be set up. Like, the goal of the application is to find suitable candidates for a job. The profile of the candidate should be done before he or she is called for the interview. With the use of Social Media and Semantic Web technologies we want to design a Semantic Web application that can extract information about a person from the social media sites.

**User’s scenarios for Social Media**

In the last few years Social Media has evolved, everyday more and more users are present on the social Web. The variety of social application makes difficult to extract the information you need, mainly because the information is dispersed in many Web sites. The picture bellow illustrates where categories of information can be found.

![Social Media world](image)

Figure 4. Social Media world.

**Use case scenario**

A human resources employee wants to find suitable candidates for the jobs of java software programming. The team does not necessarily want a person with a lot of knowledge and experience but a person that will feat with the personality of the rest of team and is capable to learn fast. Therefore, the information needed about the profile of the candidates would be:

- Knows java (LinkedIn)
- Education (LinkedIn)
- Minimum one year of experience (LinkedIn)
- Knows English (LinkedIn, Facebook, Twitter)
- Likes to travel (Facebook, Google +)
- Insides of personality (Facebook, Google +, Twitter)
- Language use (Facebook, Google +, Twitter)
- General knowledge (Twitter)
- Interests (Twitter)
- Hobbies (Facebook, Google +)

The message from the well-known quote “Tell me who your friends are, and I will tell you who you are” can be very well translated into a “social media quote”. What you share, post, like or comment on Facebook, the tweets, following and followers from Twitter, the online CV from LinkedIn can say more about the personality
of one person than a job interview.

Sequence of Events is a good representation of the user’s actions. Below is shown the sequence of events for the process of searching on the Internet.

![Sequence Diagram: Accessing Data.](image)

The information that one want may be private. Also, not all the information on the Web is true. But because the top concepts of Semantic Web are Proof and Trust we could hope that this would not be an issue.

8. Conclusions

Semantic Web technologies represent the key to better find information on the Web, to create and share content. Social Media is presently very popular and applying the technologies of Semantic Web makes social networking information more organized and easy to find. The advantage of Semantic Social Web networks is that they are based on a certain topic, like all the people that visited a certain place, all the java programmer or all the people that are customers at a certain company.

The fact that there are already some semantic applications, for example Open Graph, Google Authorship or Twitter Cards, demonstrates that the development of semantic web applications is quite close.

Even though the Semantic Web is a controversial domain, because of the cost implying the development of such applications or because of the lack of resources, the concept of will always remain. That is to be able to search on the Internet and find everything you need related to your subject of interest. This off course in the Social Media is another challenge, that being the privacy of the shared content. This means that only some people will be able to access some information.

To conclude, the presented use case scenario shows the need of Semantic Web technologies in Social Media domain. Also, there are already some methodologies that can guide developers to design and implement Semantic application.
Cómo citar este artículo / How to cite this paper


References


