

BRINGGING SEMANTICS TO SOCIAL MEDIA CONTENT

VÂRLAN SIMONA ELENA^{1*}

¹ “Alexandru Ioan Cuza” University of Iași, Romania

Abstract: It is well known that Social Web is „growing” bigger and bigger and more diverse. An important issue relative to the huge amount of information user generated content is that knowledge coming from different communities of users in social media cannot be linked or shared because user generated content, in general, doesn’t have machine interpretable metadata that can be “understood” by computers. The connection can be done assigning meaning to web content with the help of the Semantic Web standards. The paper presents a study about role and uses of Semantic Web technologies within the framework of social media. A case study about the effective use of these technologies in a web application that lists the posts introduced by users, which can be found in different social sites is provided. The application was meant for usage within the Romanian research team.

Keywords: Semantic Web, Social Media, Social Semantic WEB, SIOC, Ontology, SPARQL, FOAF, OWL, RDF.

1. INTRODUCTION

In the last decade there has been a trend towards the Semantic Web or Web 3.0, which is focused on making the information contained in web pages machine interpretable and not only human readable. In parallel, the potential of Social Web, or Web 2.0, including social networks, wikis, blogs, folksonomies and all the sites where user is content-provider not only content-consumer, still capture attention and interest of millions of users as well as researchers and developers. The most important advantage of using Semantic Web technology in social media is the capability of represent the huge quantity of data in a machine interpretable form so the material distributed on the web is structured and tagged upon an agreed vocabulary enabling semantic queries for topic of interest. Computers only present the hypermedia information, they do not understand or reason about it unless it is represented in such way that computers can parse the text and process the data. The web content became machine readable only if web resources contain semantic mark-up from a defined vocabulary. A concept defined and described by a specific semantic mark-up is a member of a class, has properties and is associated with other defined concepts. This is where the Semantic Web bring its contribution.

Semantics can help social websites by using semantic standards to define concepts, to describe people, content objects and the connections that bind them all together, social media sites can interoperate by appealing to common semantics [1]. Because Semantic Web means well-defined data and social media in general means user generated content, we want to bring these ideas together and create a web application which manages user generated well-defined content [2]. With this application, which is a case study on the use of semantic web standards for social media interaction, users can share their posts/news in the simplest manner, using natural language which is then generated in an annotated form, creating a SIOC Ontology, where users can make queries upon it.

* simona.varlan@info.uaic.ro

2. PROBLEM IDENTIFICATION

One of the problems identified by [3], [4], [5] of the current microblogging platform is the lack of machine-readable meta-data about their posts, associations, creation date and the list can continue. If there is no semantics, then the web content is hard to share between other microformats or to semantically query it. This requires representation mechanisms to interconnect people and objects on the Web in an interoperable and extensible way [2].

Realizing that the main problem for the lack of knowledge and information exchange and interoperation among online-community sites is semantics, or better said the lack of semantics in web content, researches and web application developers founded the solution in applying Semantic Web technologies. We can use standards like RDF (Resource Description Framework), FOAF (Friend-of-a-Friend), OWL (Ontology Web Language) or SIOC (Semantically Interlinked Online Communities) to create ontologies for describing concepts and relations between concepts and make semantic queries through the help of SPARQL querying language.

3. HOW TO CONNECT SOCIAL MEDIA WITH SEMANTIC WEB

In this section we will present semantic web technologies that are being used to bring semantics in the web content of the social web sites. Web developers had started to use these standards realizing the potential that interlinked data can bring to facilitate location of related and relevant information.

The FOAF project provides a way to represent people and relationships that they share in a machine-interpretable way. User profile information (with relations between users), available in FOAF ontology, is available for social media sites like Facebook, Twitter, Flickr, hi5, identi.ca. Using FOAF, each person is represented as a *foaf:Person* and may be assigned a URI, has properties like *foaf:name*, *foaf:birthday* and the relationship with other persons is made through *foaf:knows* property. The FOAF ontology like any other semantic web ontology can be queried using SPARQL, a query language which is a W3C recommendation, to retrieved and share data.

Another ontology language to represent knowledge is OWL, which is a more complex language offering the ability to define properties like *owl:InverseFunctionalProperty* or *owl:sameAs* links.

The SIOC standard is used to describe various containers (i.e. wiki, blog, and forum) and web content items (i.e. WikiArticle, Post). It can be used in combination with FOAF vocabulary for describing people and their friends, with OWL to associate, for example, different web resources for the same concept, in combination with SKOS (Simple Knowledge Organization System) model to organize controlled vocabularies, folksonomies or thesaurus-like data, or extends the DC (Dublin Core) ontology.

In the literature review we find several projects that integrates semantic web technologies to represent social media content.

In [3] Semantic MicroBlogging is a distributed microblogging system that uses SIOC - to define semantics for representing posts, FOAF - for defining people, Online Presence Ontology (OPO) - for describing user's presence, Meaning Of A Tag (MOAT) - to model semantic tagging capabilities.

miKrow [6] is a semantic microblogging tool that allows its users to share information/notes about them. Other research projects that uses semantic web technologies are described in [7], [8] which combines the social contribution of humans through ratings and semantics issued from linked data.

An interesting add-on tool that turns the web browser into a collaborative semantic data editor is presented in [9]. Collaborative tagging and ontology representation of social content was used in [10], [11], [12] and [13] as a way of bringing semantics to social media platforms.

There are also various tools and exporters that have been created to share SIOC data from social media sites. We have a series of APIs for the main web programming languages, PHP, Java, Perl and Ruby and data exporters for WordPress, Drupal, phpBB and BlogEngine.NET that creates SIOC representation blog posts, comments, content and author of the post and other information about content of the site.

Despite the utility of semantic web technologies, there is a major problem in using them, that is represented by the end user. The simple, common user, that must be convinced to use these technologies, who, maybe, is not familiar and doesn't have technical skills to be able to manage these technologies and who is not wheeling to spend time and effort in installing APIs or export metadata. The solution is to create applications that generate automated semantic formats that can be automated queried for information retrieval.

Another important aspect regarding SIOC is that data can be mapped to RSS (Really Simple Syndication) news reader. RSS allows us to navigate a subset of SIOC information inside a regular RSS news reader, but with some lost information because SIOC has a richer data model than RSS [2].

4. APPLING SEMANTICS TO SOCIAL MEDIA CONTENT

Our research study was intended to demonstrate how a user can create SIOC ontology about his posts from different web sources. The SIOC ontology must be automatically generated, to represent some simple fact (it is a demonstrating study) and queried in order to find the exact information that the user is looking for. The premises of the case study were that the simple user doesn't know anything about metadata, ontologies and especially how to create them. The semantic annotated information is not visible for the regular user. It represents the backend of the application where is processed and then output to the user in a natural, human-readable form. The users are not used to, or technically prepared to these semantic web standards, they do not know how to produce semantic mark-up content, so we need to create a simple application, for a simple user, but that implies user contribution by providing necessary information for SIOC ontology creation. The architecture of the application is very simple and involves annotated web content. The result is the SIOC ontology, which is generated automatically and queried using SPARQL semantic query language. The architecture is presented in Figure 1. To make queries we also need a SPARQL endpoint where the query is processed and the corresponding result (a Json array) is returned to the user.

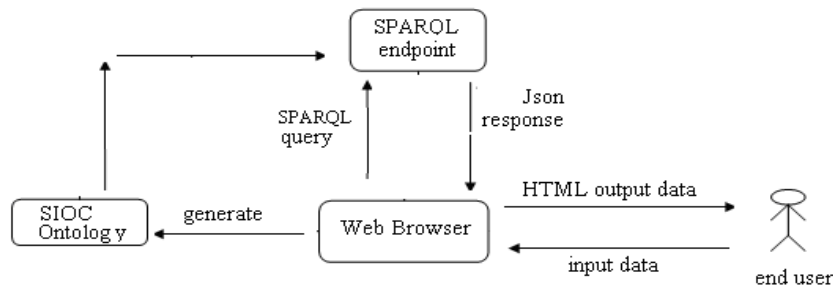


Fig.1. The case study architecture

The user introduces in a simple and very intuitive HTML form the necessary and basic information about the post/new/comments that he/she posted on his/her blogs or other social media sites. This is the only implication that the user had to do to create the SIOC ontology. Also, in this manner, we bring information from multiple sources to one, with the aim to create in the end the ontology upon a semantic standard format like SIOS that can be easily query, process and share among different social communities. The generated SIOC ontology is presented in the following code lines. The information offered by the user in natural language is then semantically marked-up. The SIOC statements that we used are *sioc:Post*, to define each post concept, *sioc:site*, used here for the corresponding URI of the post, *sioc:has_creator* for the name of the creator of the post, *sioc:content* for the description of the post, and so on.

```

<sioc:Post rdf:about="http://plumee.ub.ro/2019/">
<sioc:Site>http://plumee.ub.ro/2019/ </sioc:Site>
<dcterms:title>PLUMEE 2019 </dcterms:title>
<dcterms:created>10.04.2019 </dcterms:created>
<sioc:has_creator>Simona Varlan </sioc:has_creator>
<sioc:content>I am participating at PLUMEE 2019. </sioc:content>
</sioc:Post>
  
```

Web content being semantically annotated the next step is to make queries upon the SIOC ontology. We used the SPARQL protocol to interrogate the ontology. The query asks for all the input posts.

```

select ?title ?creator ?data ?description ?url
where {
  ?post dcterms:title ?title.
  ?post sioc:has_creator ?creator.
  ?post dcterms:created ?data.
  ?post sioc:content ?description.
  ?post sioc:Site ?url }
  
```

It is obvious that we can make a lot of different other queries, searching, for example, only the post that a specific user posted, and, the beauty of the ontology and semantic web format is that it can be developed by integrating other semantic vocabularies like FOAF, SKOS or OWL.

As future directions of development we can try to combine our mark-up sets of data with other ontologies like DBpedia or Wikipedia and to realize complex queries from multiple data sets to connect our ontology with other social media ontologies.

4. CONCLUSIONS

Semantic Social Media refers to how to create the semantic layer or better say how to use annotations to describe user generated content about people, places, events, user knowledge in general. We have the SIOC ontology, validated by the Semantic Web Community, but we also need the support of browsers developers to implement functionalities that permits users to generate machine-interpretable metadata. This is not an easy task because not every user complies with the Semantic Web Vision.

We begin our research by identifying the problem (or challenge) regarding the access of data from different social networks. A solution to this problem was presented: the use of semantic web technologies like SIOC and FOAF to bring semantics to web content, as it is referenced in literature review. The solution was effectively applied in a case study to demonstrate how SIOC ontology can be generated through user implication bringing information from different sources in one. SIOC data are then extracted using SPARQL protocol to query data.

RÉFÉRENCES

- [1] Ravi, Sharma, Analyzing the Role of Semantic Web in Social Networking Sites, *International Journal of Engineering and Technology*, 1, (4), 505-517, 2012
- [2] John, Breslin, What is the Social Semantic Web, and why do we need it?, posted on Technology web site at the address < <http://technologyvoice.com/2011/09/09/what-is-the-social-semantic-web-and-why-do-we-need-it> >, September 2011
- [3] Passant, Alexandre, Breslin, John, Decker, Stefan, Rethinking Microblogging: Open, Distributed, Semantic, *Proceedings of the 10th International Conference Web Engineering*, Vienna, Austria, 2010
- [4] Smith, Daniel A., Van, Kleek, M., Seneviratne, Oshani., Bertails, Alexandre, Berners-Lee, Tim., Hall, Wendy, Shadbolt, Nigel, WebBox: Supporting Decentralised and Privacy-respecting Micro-sharing with Existing Web Standards, 2012
- [5] Cheong, M., Ray, S., A Literature Review of Recent Microblogging Developments, Technical Report TR-2011-263, Clayton School of Information Technology, Monash University, 2011
- [6] Penela, Victor, Álvaro, Guillermo, Ruiz, Carlos, Córdoba, Carmen, Carbone, Francesco, Castagnone, Michelangelo, Contreras, Jesus, miKrow: Semantic intra-enterprise micro-knowledge management system, *The Semantic Web: Research and Applications*, 2011
- [7] Philippe Lemoisson, Guillaume Surroca, Clement Jonquet, Stefano A. Cerri. ViewpointS: When Social Ranking Meets the Semantic Web. *FLAIRS: Florida Artificial Intelligence Research Society*, May 2017, Marco Island, FL, United States. 30th International Florida Artificial Intelligence Research Society Conference, FLAIRS'17, 2017. <lirmm-01598842>
- [8] Georgi, Kobilarov, Tom, Scott, Yves Raimond, Silver, Oliver, Chris, Sizemore, Michael, Smethurst, Christian, Bizer, Robert, Lee, Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to MakeConnections, *The Semantic Web: Research and Applications*, 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31-June 4, 2009
- [9] Torres, Diego, Diaz, Alicia, Skaf-Molli, Hala, Molli, Pascal, Semdrops: A Social Semantic Tagging Approach for Emerging Semantic Data, *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence*, WI 2011, Campus Scientifique de la Doua, Lyon, France, August 22-27, 2011
- [10] Ashfaq Ahmad, Roslina Othman, Mohamad Fauzan, A Social Semantic Web based Conceptual Architecture of Disaster Trail Management System, *International Journal of Advanced Computer Science and Applications* 8(4), April 2017
- [11] Rehab, Ghaly, Emad, Elabd, Emad Elabd, Mostafa, Abdelazim, Mostafa, Tweets classification, hashtags suggestion and tweets linking in social semantic web, *SAI Computing Conference (SAI)*, July 2016
- [12] Federico, Bertola, Viviana, Patti, Ontology-based affective models to organize artworks in the social semantic web, *Information Processing & Management* 52(1), November 2015
- [13] Khaled, Halimi, Hassina, Seridi-Bouchelaghem, Faron-Zucker, Catherine, An enhanced personal learning environment using social semantic web technologies, *Interactive Learning Environments* 22(2), March 2014