Moveek: A semantic social network

Pablo Camarillo R., Abraham Sánchez L. and David Nuñez R.

Benemérita Universidad Autónoma de Puebla
Computer Science Department
MOVIS Group
p.camarillor@gmail.com, asanchez@cs.buap.mx, drpulpo@hotmail.com

Abstract. We know that it is currently difficult to create semantic-web content because pages must be semantically annotated through processes that are mostly manual and require a high degree of engineering skill. We must therefore devise means for transforming existing, non-semantic social networks into semantic social networks. We propose using information extraction ontologies to handle this challenge. In this work we show how a successful ontology-based data-extraction technique can automatically generate semantic annotations for social networks. We have implemented a prototype of our approach to demonstrate that our proposal works.

1 Introduction

The web is now a major medium of communication in our society and, as a consequence, an element of our socialization. As the web is becoming more and more social, we are now collecting huge amount of knowledge on-line [1]. Semantic web researchers provide models to capture such activities that have to be fully exploited in order to be turned into collective intelligence.

The "semantic web' represents a major advance in web utility, but it is currently difficult to create semantic-web content because pages must be semantically annotated through processes that are mostly manual and require a high degree of engineering skill. Semantic-web proponents propose making web content machine understandable through the use of ontologies, which are commonly shared, explicitly defined, generic conceptualizations [2]. But then one of the immediate problems we face is how to deal with current web pages. There are billions of pages on the current web, and it is impractical to ask web developers to rewrite their pages according to some new, semantic-web standard, especially if this would require tedious manual labelling of documents.

Web semantic annotation research attempts to resolve this problem. The goal of web semantic annotation is to add comments to web content so that it becomes machine understandable [4]. Unlike an annotation in the normal sense, which is an unrestricted note, a semantic annotation must be explicit, formal, and unambiguous: explicit makes a semantic annotation publicly accessible, formal makes a semantic annotation publicly agreeable, and unambiguous makes a semantic annotation publicly identifiable. These three properties enable machine understanding, and annotating with respect to an ontology makes this possible.

In this work, we present an approach to semantically relate contents posted on a social network. A semantics based on the representation of knowledge through an ontology of domain will be used. The domain covering by this ontology is a domain of scientist concepts mentioned in the publications made in the social network. Our work aims to create a platform based on the composition of Web services able to provide the functionality of a social network and at the same time provide various features of the semantic web, such as the semantic annotation and semantic queries to the information published on this network.

Our semantic social network Moveek are briefly presented in Section II and the process to semantically relate the contents. In Section III we discuss some experimental results and finally in section IV we give a conclusion and the future work.

2 A semantic social network

The term semantic social networks was coined independently by Stephen Downes and Marco Neumann in 2004 to describe the application of semantic Web technologies and online social networks [5].

The main motivation of this work was the curiosity produced by the strange behavior of the many social networks like Facebook, Twitter, Youtube, etc., i.e., at the time to suggest new friends, products or services. The most important function of our social network is located in the invocation of some Web services. Such Web services will perform the tasks of semantic annotation, information extraction, semantic query and social interaction with the social network, Moveek. Figure 1 shows our proposal to produce the semantic core of our social network.

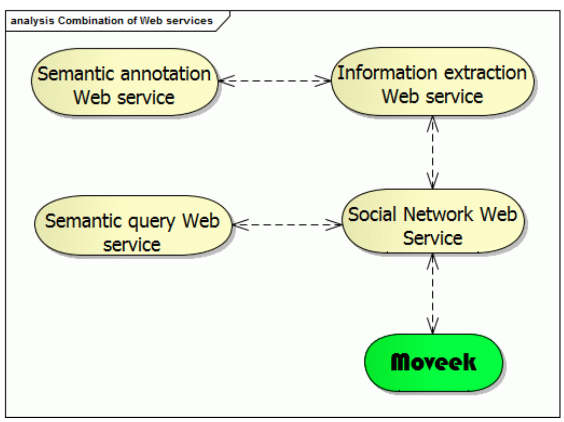

Fig. 1. The overall functionality of Moveek.

In semantic Web applications, ontologies describe formal semantics for applications, and thus make information sharable and machine-understandable. The work of semantic annotation is, however, more than just knowledge representation. Semantic annotation applications must also establish mappings between ontology concepts and data instances within documents so that these data instances become sharable and machine-understandable.

The term Ontology has been used in several disciplines, from philosophy, to knowledge engineering, where ontology is comprised of concepts, concept properties, relationships between concepts and constraints. Ontologies are defined independently from the actual data and reflect a common understanding of the semantics of the domain of discourse. Ontology is an explicit specification of a representational vocabulary for a domain; definitions of classes, relations, functions, constraints and other objects.

The ontology-based knowledge representation needs a robust ontology. We named our ontology OntoScience, since we want to cover a scientific domain in this work. We develop our ontology according to methodology proposed by Rubén Darío Alvarado [6]. The steps of this methodology are shown in Figure 2.

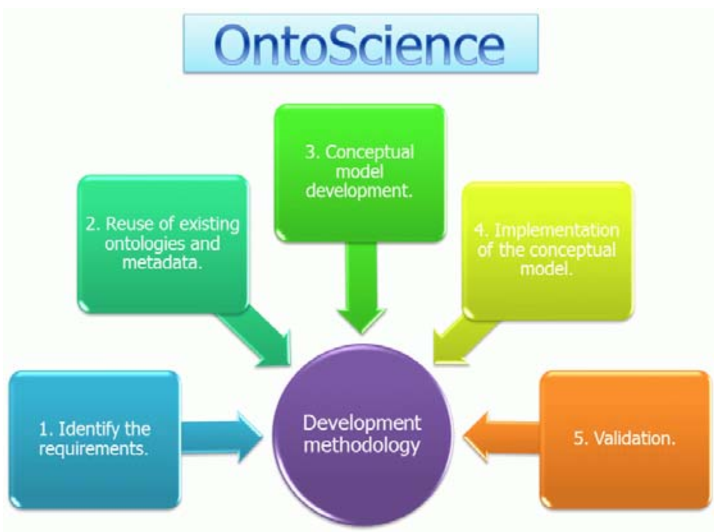

Fig. 2. The development process for the OntoScience.

First, we delimited the domain of our ontology and we set the questions that the ontology is able to answer. Second, we checked many existing ontologies and metadata like WikiOnt[1], Ontology of SCIENCE[2] and the Dublin Core Metadata[3]. Third, we started to collect the concepts and metadata that we needed

[1] http://sw.deri.org/2005/04/wikipedia/wikiont.html
[2] http://protege.stanford.edu/ontologies/ontologyOfScience/ontology_of_science.htm
[3] http://dublincore.org/

and we built the conceptual model with the classes and relationships necessaries for our proposal. After of developing the conceptual model, we implemented the ontology with the Protégé system [7]. Finally we validated the ontology structure using java routines that they programmatically verified the taxonomy and consistence of OntoScience. Information extraction For building the bridge between the modelled concepts in OntoScience and the posts published in our social network, we need to know what terms have a relationship with the concepts in OntoScience. To do this, we developed an information extraction tool that is published as a Web service. This tool receives the post as a string and use an application produced by the GATE environment [8]. The GATE application returns the information of the post as a XML document. This document contains the terms that are mentioned in the post and that are modelled in OntoScience. The document also contains the ID of the class that coincides with the term mentioned in the post. Figure 3 shows how the information extraction task is performed

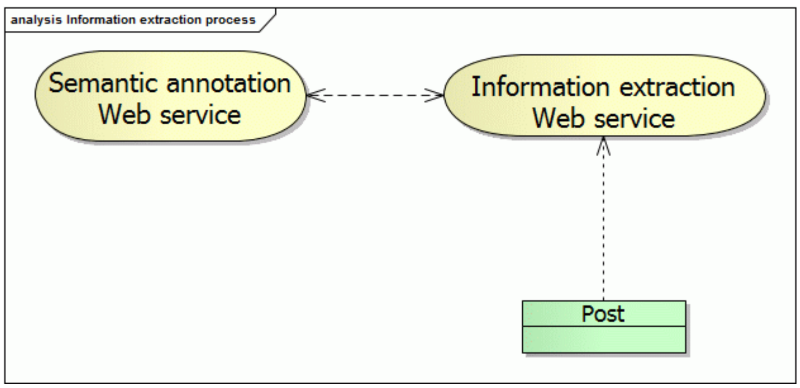

Fig. 3. Information extraction flow.

Pragmatically, queries and assertions are exchanged among software entities using the vocabulary defined by a common ontology. Ontologies are not limited to conservative definitions, which in the traditional logic sense only introduce terminology and do not add any knowledge about the world. To specify a conceptualization we need to state axioms that put constraints on the possible interpretations for the defined terms.

The difficulty in sharing and processing Web content, or resources, derives at least in part from the fact by using the resources are unstructured, and consist of text, iby usingideo etc. The semantic annotation process is performed by using the extracted XML document; this document is retrieved from the information extraction Web service. Figure 4 illustrates the steps for inserting the semantic annotation in the post.

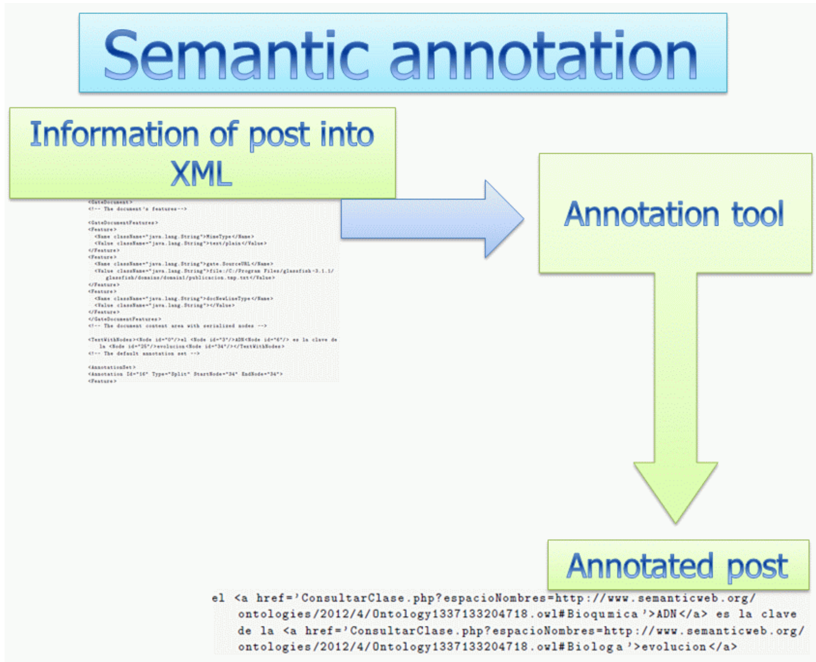

Fig. 4. The proposed semantic annotation solution.

First, we receive the XML document with the information of the post and programmatically we rebuild this post with an HTML link that, after that, it can be used to formulate semantic queries. This HTML link contains a string that uses the class ID extracted as a GET variable for the PHP engine that will performs the semantic query. In the Figure 5, one can see an example of an annotated post.

A domain ontology provides the standardized terminology and conceptualization of a particular domain. We suppose that a certain domain ontology is agreed and used to reconcile the semantics of model contents. The ontological concepts are referenced by model contents through simple URIs or semantic relationships.

```
el <a href='ConsultarClase.php?espacioNombres=http://www.semanticweb.org/
   ontologies/2012/4/Ontology1337133204718.owl#Bioqumica'>ADN</a> es la clave
   de la <a href='ConsultarClase.php?espacioNombres=http://www.semanticweb.org/
   ontologies/2012/4/Ontology1337133204718.owl#Biologa'>evolucion</a>
```

Fig. 5. An example of an annotated post.

3 Experimental results

Moveek was developed using PHP server-side scripts and AJAX client-side scripts. Our Web services are developed as follows: social interaction Web services were developed with C# and .NET framework 3.5 technologies for a quickly development; semantic Web services, information extraction information Web services and semantic queries Web service were developed with Java technologies since we used the API of GATE and Protégé. We should note that one of the advantages of Web services is the interoperability between many development environments such as our architecture.

The reason for using ontology is because we need answers to many questions about the social network behavior. These answers are solved by using the relationships modelled in the ontology. Following the previous strategy to solve the semantic issues, we used a Web service to execute the semantic queries. At this point, the Moveek users have many annotated terms and related with their corresponding ontology class. The figure 6 shows the path that follows a post to send the request of a query and how the response of the semantic Web service is used to show the list of post(s) related with the term originally selected.

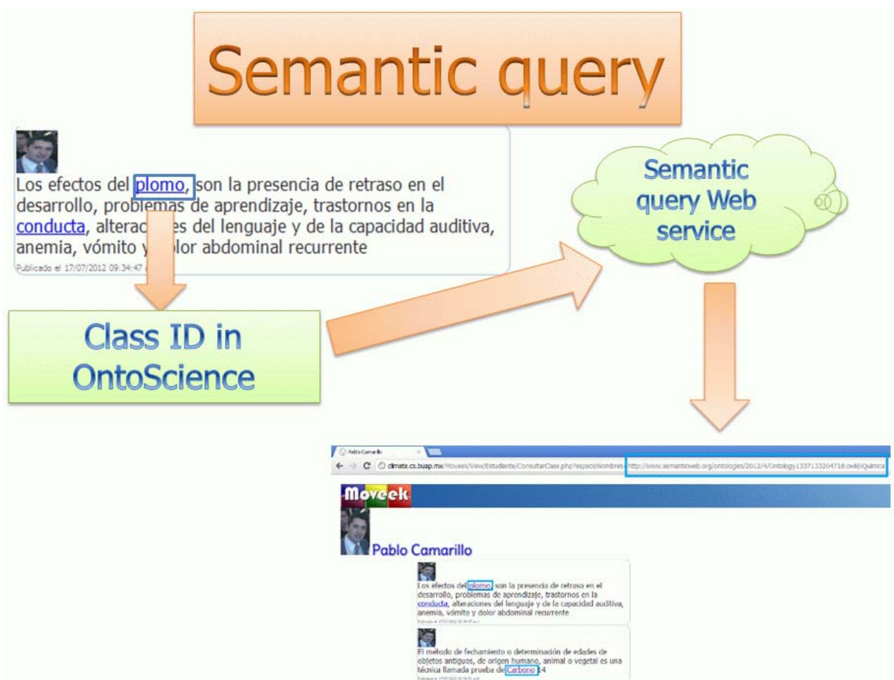

Fig. 6. An example of our semantic query strategy.

For discovering the knowledge, our proposal is completed when we use all Web services mentioned previously. Every Web service has a key role in our architecture. In the Figure 7, one can see how the architecture combines the Web services and its relationships with the Moveek social network.

Web services are used to distribute services over the Internet. They make operations of applications available or enable information systems to be invoked over the network. There are two ways to combine such services: either through orchestration or choreography. In orchestration, the involved web services are under control of a single endpoint central process (another web service). Choreography, in contrast, does not depend on a central orchestrator. Each Web service that participates in the choreography has to know exactly when to become active and with whom to interoperate. In our proposed architecture, we used orchestration for composing Web services.

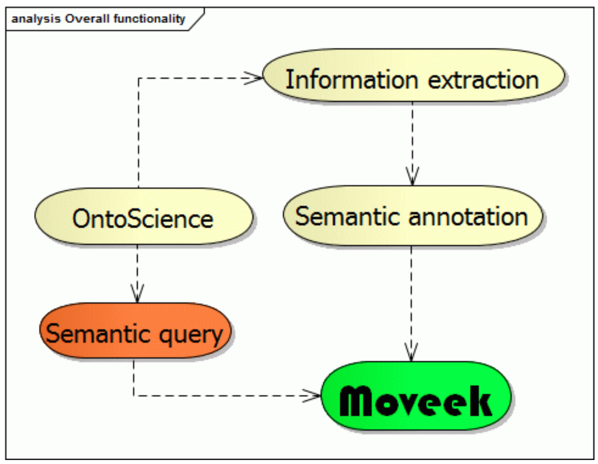

Fig. 7. Combination of the Web services.

We saw that since there is a considerable range of semantic annotation tools that each have its own characteristics but whose goal is becoming more oriented to support human annotators when they are creating annotations. The level of automation depends on information extraction engine embedded into the tool. The annotations can be stored in an annotation server, in loaded or unloaded mode from the source document, or a knowledge base. They can, also be used to improve information retrieval systems, to populate (enrich) an existing ontology, or even help in building ontologies. Despite a rapid evolution of languages and rules, these tools do not always suit the specific uses of the real world. And for companies who have desires to benefit from the new model specified by the founder of the semantic Web, they can integrate into their applications, some still existing constraints should be rejected.

1. Annotations tools there are intrinsically linked to the information extraction engine used to extract the information from text and annotated them.
2. We thought that it is necessary dissociate both extraction information tool and annotations tool; in order to the annotation tool can use any extraction information tool according to the applications needs.
3. Almost all information extraction engines are based in supervised learning process. These systems are powerful for the structured content or even semi-structured. But in the Web and in the companies, the new strategic information is stored in non-structured content, in particular, semantic relations among entities, whose semantic is the most difficult to extract.
4. We consider that in this case, it is necessary give priority to the extraction systems based in different fines linguistics analysis that they allow define this semantic into the document of the treated documents.
5. Semantic annotations must be follow encouraging approaches domain-ontologies-based, not generic ontologies. This is about in the case of the applications aimed to companies. These last are interested in inherent facts in their own applications domain and not in a general domain.

It is necessary to provide maximum support to the user and, in particular, when the aim is also to populate the ontology. The processes and interfaces must not only submit suggestions, proposals but must also be able to guide it taking advantage of the difficulties and constraints modeled in the reference ontology. We are convinced that the question of the integrity of the annotations, and more specifically of the knowledge base that supports new instances resulting from these notes, it is essential to exploit the results provided from such application.

3.1 Discussion

In the study of the semantic Web, the aim is to describe the content of annotating resources with unambiguous information to facilitate the exploitation of these resources with software agents. However, current data from the Web are frequently written in natural language, since these data are intended for humans. Natural language is inherently ambiguous and it is important to consider an alternative formal and explicit semantic to implement mechanisms to reduce these ambiguities, it is equally important to consider the relationship between the content of the resources and their annotations. The task of the annotation for the semantic Web is therefore to take as input a documentary resource and provide as output the same content enriched with semantic annotations based on the more or less formal knowledge representation.

An important component for the semantic Web, is an ontology, i.e., a collection of information. Artificial-intelligence and Web researchers have another definition of the ontology term, they define an ontology as a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules. This taxonomy defines classes of objects and relations among them. Classes, subclasses and relations among entities are a very powerful tool for Web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties. Inference rules in ontologies supply further power.

The computer does not truly “understand” any information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user. With ontology pages on the Web, solutions to terminology (and other) problems begin to emerge. The meaning of terms or XML codes used on a Web page can be defined by pointers from the page to an ontology.

Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches-the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules [1].

We have just presented in this work, an interesting solution for the semantic annotation problem and we noted that the semantic Web problem, is intrinsically linked to the modelling of an ontology. That is, this ontology will represent the concepts, attributes and relationships of an area by using a language of knowledge representation oriented to the Web as OWL. The semantic annotations are structured with an ontology and its values point to instances of the reference ontology or, in some cases, directly towards the own concepts. Other terminology resources or ontological, and thesaurus, can also be used as the semantic annotation tools to provide a different perspective to the end user of the subject of a documentary resource.

4 Conclusion and future work

We have developed an interesting social network called Moveek and its objective is to create networks of users who share and disseminate publications with some scientist contents. Each publication, automatically, will highlight the concepts that the system has detected as scientists contents in the ontology. In this way, the user can consult more publications semantically related with the highlighted concepts [6].

The study and development of the Semantic Web applications includes various areas ranging from correct modelling of a Web application to the development of a ‘good’ representation of knowledge and the relationship that should be both ideas to make the Web more human. In this work, we involve all these aspects concerning the semantic Web with the objective to understand, but in fact to develop an application that would reflect the advantages (and disadvantages) of the semantic Web. In the research and in the development of our social network, we have realized that, while the semantic Web meets its mission of making the information queries more efficient on the content of the Web, it also entails certain disadvantages, especially concerning the necessary infrastructure to not affect the performance of the Web applications.

Another interesting aspect that we highlight of this work is the growing presence of multimedia resources on the Web, this makes more challenging the development of the semantic Web. With the latest idea, we can introduce the future work of this work. First, we must improve the social network modelling in order to scale the application and think of more features such as the management of courses to create networks of scientific collaboration on our social network. Another crucial point in our future work is our interest to automate the process of learning of our ontology and thus enrich the scientific domain which it currently covers.

References

1. Peter Mika, "Social networks and the semantic Web (Semantic Web and beyond)", *Springer-Verlag*, (2007)
2. T. R. Gruber, "translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5, No. 2, (1993) 199–220
3. S. Garlatti. Y. Prie, "Méta-données et annotations dans le Web sémantique", *Cépadùes*, (2004) 45–68
4. J. G. Breslin, A. Passant and S. Decker, "The social semantic Web", *Springer-Verlag*, (2009)
5. S. Downes, "Semantic networks and social networks", *The Learning Organization*, Vol. 12, No. 5, (2005) 411–417
6. P. Camarillo R., "Descubriendo conocimiento en redes sociales a partir de ontologías", *BS Thesis FCC-BUAP*, in spanish, (2012)
7. The protégé ontology editor and knowledge acquisition system, *http://protege.stanford.edu/*
8. General architecture for text engineering (GATE), *http://gate.ac.uk/*