Some Considerations Regarding the Usefulness of Applying Semiotics in Web Ontologies

necula, sabina-cristiana and popescul, daniela

Alexandru Ioan Cuza University of Iasi

September 2011

Some Considerations Regarding the Usefulness of Applying Semiotics in Web Ontologies

Sabina Cristiana Necula

PhD Researcher
Research Department, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iaşi, Romania
Address: B330a, Bd. Carol I nr. 22, 700505, Iasi, Romania
Phone: ++ <40 232 201070>; Fax: ++ <40 232 217000>
E-mail: sabina.mihalache@gmail.com

Daniela Popescul

PhD Lecturer
Business Information Department, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iaşi, Romania
Address: B324a, Bd. Carol I nr. 22, 700505, Iasi, Romania
Phone: ++ <40 232 201070>; Fax: ++ <40 232 217000>
E-mail: rdaniela@uaic.ro

Abstract

The current paper aims to study Web ontologies from a semiotic perspective. On the one hand, the study of signs is constantly used lately in the analysis of the social life, of organisations and informational systems. On the other hand, Web ontologies may be seen as systems of signs, wishing to catch the reality of a certain field by identifying fundamental categories of identities and the relations between them. Their intention is to describe their significance as accurately as possible, so as to ensure the best communication between computing systems. Since any act of communication includes and involves signs and codes (Fiske, J., quoted in Borţun, 2006), in the current paper we bring to discussion the opportunity of using semiotics, as the science that studies signs, in the study of the Web ontologies.

Keywords: semiotics, info-semiotics, Web ontologies, RDF, knowledge representation

Ontologies and semiotics – short descriptions

Briefly, the term *ontology* used in the computing context (mainly in the artificial intelligence and knowledge representation) refers to the attempt to draw a conceptual, exhaustive and accurate scheme inside a given field, having as purpose the facilitation of the communication and sharing of information among various systems. In the meaning associated with computers, ontologies wish to catch the reality of a certain field by identifying the fundamental categories of entities and the relations among them. More precisely, they contain categories, fundamental concepts, concept properties and relations between concepts, and also axioms. W3C (2009) affirms that ontologies are shared models of a domain that encode a view common to a set of different parties. As Andone (2005) clarifies it, ontologies aim to explain, as accurately as possible, the significance of concepts so that the communication between separate computing systems is at its best. Ontologies have spawned fair degree of research and implementation interests, as evidenced in the literature. For instance, while Shum et al (2000) use ontology principles to develop ScholOnto, an ontology-based digital library server to support scholar interpretation and discourse, Lai and Yang (2001) experimented the use of ontology principles for deriving ontology-based metadata for the Chinese information services in Chinese digital libraries. Yeh

(2002) designed an ontology-based portal for digital archive services. Uszkoreit et al (2003) successfully employed ontologies for creating a knowledge portal for the field of Language Technology. Kalfoglou et al (2004) developed an ontology-driven web-based system for personalized news services. As it can be inferred from the examples shown, the main use of ontologies is in the field of Semantic Web, which came with the idea of semantically describing web sources. Considering that the Web is a big source of data that is not semantically described, it would be ideal that someone could describe these sources in such a way so as to be understood by a machine. The emergence of terms such as "significance", "meaning" and "good communication" in the ontologists' definitions brought us to the idea of confronting it with similar concepts from semiotics.

Semiotics is the study of signs and it deals with their creation/production, representation and interpretation (Cordeiro and Filipe, 2004). Borţun (2006, p. 17) emphasises the fact that semiotics deals mainly with "the text" and gives the receiver or "the reader" a much more active role than in most process communication models. For semiotics, the transmitted message is a construction of signs that, as a result of the interaction with the receiver, produces the meaning. The meaning is born or is "discovered" during the "reading" process, which is a negotiation process between the "reader" and the "text". The first one brings its cultural experience referring to the signs and codes of the latter. We can see that significance is not a static, absolute concept, clearly delimited in the message, but it is rather obtained from an active process, which semioticians describe using verbs such as "to create", "to generate" and "to negotiate". The semiotic moments passed the abstractedness sphere and have been constantly used in the recent years in the analysis of the social and economic life. Leaving apart the easy-to-understand approach which regards semiotics as a way to help young people decode the metaphors created by the media or to protect themselves from other such ways of using power through semiotic means (Gazendam and Liu, 2004), real subjects have also been created, such as organisational and informational semiotics, computer semiotics or infosemiotics. Gazendam (2004) states that organizational semiotics tries to understand organizations starting from the use of signs, texts, documents, artefacts and sign-based communication. In order to do this, it uses the results of subjects such as psychology, economics, and informational systems. One of the purposes of organizational semiotics is to show what we do when we try to understand, design or change organizations using, for example, models and metaphors. Understanding can help us free from the unconscious use of certain models and metaphors, leaving room for the free design. Informational semiotics or info-semiotics analyses information systems from a particular perspective, that is, as interpretation targets (Andersen, 2000). Semiotics views computer systems as sign-vehicles whose main function is to be perceived and interpreted by a group of users. It has nothing to say about data in itself, only in its capacity of being interpreted and used as a source of knowledge or guide for action. Design and programming are considered activities specific to semiosis, processes that help the formation and interpretation of signs. The individual is considered a creator, interpreter and referent of signs, as a user and reproducer of a common meaning potential and code, using the results of a semiotic labour done by others. The Internet itself is seen by Sowa (2000) as "a giant semiotic system".

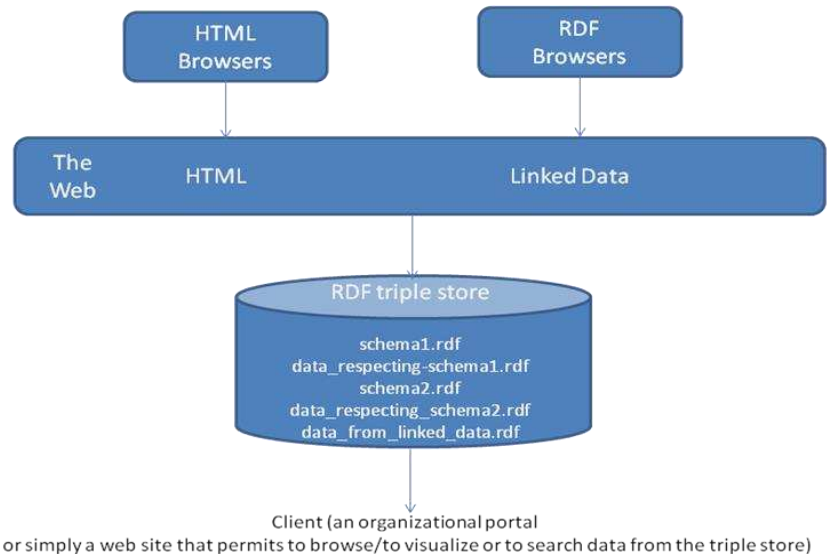
Interferences between Web ontologies and semiotics – some reflections

For the beginning, we will choose the knowledge-based approach out of the three approaches of organizational semiotics (system-oriented, behaviour-oriented and knowledge-oriented)

identified by Gazendam (2004). The approach refers to a triad-system formed from 1) the cognitive architecture of the human actor, 2) the signs "hidden" in its mind, that cannot be "exported" to the semiotic environment (which, in short, we can call tacit knowledge) and 3) the signs from the actor's semiotic Umwelt, made and perceived by the actor (similar to the explicit knowledge or information, made to circulate among actors, to express their intentions and allow the communication). Applying this approach to the ontologies example, we compare them to the explicit knowledge from the above-mentioned triad. *The categories and relations that make the ontology form a semiotic Umwelt which, considering the current level of development of the Web ontologies, we can assess as being imperfect*, *even for describing a limited field of activity*. However, its imperfection is diminished by 1) the intention of ontologies to represent only the existent things, which have a proven value of truth and 2) by the preoccupations in the ontologic engineering field, whose purpose is to offer standard/pattern ontologies for a certain field of study, catching the knowledge and meanings from that field as accurately as it can. By materializing these intentions, the semiotic Umwelt given by the Web ontologies will create a clear, consistent and coherent context for a certain field, simplifying communication through the existence of certain standards/patterns recognized as such, and at the same time giving the other actors involved in the semiosis the freedom to decide the necessary degree of specialization and instantiation of the fundamental concepts.

Another element that information semiotics emphasizes is the following: *semioses should be directed towards their user, the computer science thus getting a humanist aspect*. The essence of design and programming, from a semiotic perspective, is not to make models strictly to the benefit of the computer professional, but rather to use the machine (the computer) in order to say something to people. To illustrate this assertion, Andersen (2000, p. 18) introduces a metaphor, where a system is viewed as a kind of theatre, its executions are performances interpreted by an audience, and the designer is a stage director whose success does not ultimately depend on the look of the props and set-pieces from backstage, but on their communicative effect on the audience. In this spirit, ontologies were taken from the artificial intelligence area and processed to make the Semantic Web, whose desiderate is to give the "traditional", unstructured Web a meaning – a purpose which is undoubtedly a humanistic one, meant to make the human-to-computer and human-to-human communication, mediated by the computer networks, more meaningful and consistent.

However, here are some technical problems worth taking into consideration. First of all, an ontology should be made public at a web address. This means that, in order to have data from a web source that is semantically described, its provider must host this ontology on the web server and an another file that contains data from his/her web site described respecting the ontology in



order for the data to be used by other applications. At this moment, not all the website providers are aware of the importance or benefit of the ontologies. For this reason, at this point there are either implementations made by scientific researchers who try to improve the research domain, or governmental initiatives such as data.gov or data.gov.uk. Second of all, in order to manipulate data described by following a certain ontology, there is a need for using technologies that follow the Semantic Web stack layers as described by Tim Berners Lee. We can see in Figure 1 what the main parts of the Semantic Web application architecture are.

Figure 1 An Semantic Web application architecture

For example, we may imagine a schema for describing Companies and an example of data described by observing the defined schema/ontology. Since RDF (Resource Description Format) is the standard accepted by W3C – World Wide Web Consortium, we used this standard to semantically accomplish requirements that we pose for developing a kind of portal intended to be used by the government for some analyses on companies' data.

Figure 2 An example of a simple ontology

RDF is a general method to decompose any type of knowledge into small pieces, with some rules about the semantics or meaning of those pieces. The point is to have a method so simple that it can express any fact, and yet so structured that computer applications can do useful things with it. While there are not many implementations in the field, we benefit from the existing grounding

technologies such as ontology, RDF/OWL descriptions, or SPARQL language. Related research includes research on ontology generation, ontology mapping, and ontology evolution. An ontology can be generated manually using an authoring tool or (semi)automatically, from various knowledge sources (e.g., database schemas). Techniques used for ontology mapping, including ontology alignment and ontology merging, overlap to a large extent with those techniques for schema matching. Finally, ontology evolution, also called ontology versioning, involves changes on representation, structure, and semantics of ontologies. Each step of such an evolution must ensure the consistency between the old version and the improved version of the ontology, just as if a database schema's evolution must guarantee the consistency of the new schema with the data.

Using ontologies for semantic integration of heterogeneous data promises to be a way to solve the problem of managing knowledge at the level of a corporation, government, social network, or a big actor that has a lot of data described according to an ontology. Why do we have better ads offered by Facebook instead of ads offered by Google? It seems that Facebook is currently using rdfa (Lunn, 2010) and for the moment Facebook has its own "database with relations/ preferences of its users" in order to recommend better ads. This is the problem that Semantic Web addresses: big data sets described according to schema(s). Its main applications are managing the currently linked data, learning how to extract information from the currently linked data, sense-making of events (there is, a lot of life data streams like tweets) in order to provide a solution able to gather, collect and analyze in real time a large number of live data streams (e.g. twits), to extract the information contained and to map any reference to both a) a geographical location/point of interest, etc. and b) a domain specific facts (e.g. music events or violence in demonstration). The goal is to identify events happening in a specific area (e.g. a specific city) in a short time (e.g. some hours). Sense making of the events is provided via mapping events over location and time. There is a lot of use for social purposes, such as emergency operators or governmental bodies. We must say that Semantic Web is intended to be used by people in the way which the web intended to be at its origins.

We can assert that the actors involved in semiosis in the ontologies and semantic Web field are more aware than the „usual" computer science workers (for example, the database creators) of their role as stage-directors working for the public. To support this statement, we use John Sowa's example, who noticed, in Sowa (2000), that the proposals for ontologies and metadata used in Internet must take into consideration all the features of signs, saw by Peirce as (1) entities that represent (2) other entities to (3) agents. He draws the attention on the fact that designers neglect the represented entities and agents whom the signs are intended to. Starting from the assertion that ontologies are „tightly interconnected collections of signs" and „the primary connections are in the minds of the people who interpret them", Sowa greets the ontologies creators' intention to make these connections explicit by tagging the data with more signs, but he considers this process a risky one unless it is performed in a rigorous and involved manner. He asks ontology developers to see beyond the syntax (the way the signs relate to one another), to the semantics (the way the signs are related to things in the world) and pragmatics (the study that relates signs to the agents who use them to refer to things in the world and to communicate their intentions about those things to other agents who may have similar or different intentions concerning the same or different things). The fact that ontology developers (also known as knowledge engineers) usually come from their specific field of activity (medicine, law etc.) and

have a background in logic and philosophy, makes them more aware of the semiotic aspects of their work and more conscious of the ontologies' adaptation to the realities of their own domain.

In our example shown in Figure 2, a semantic web application that would use our ontology would bring data from two sources: one is ours, with data about companies, and the second one is *geonames*, containing data concerning countries, regions, longitude and latitude. If the developer of the semantic web application wants to visualize the company on a map, it would use a procedure in PHP, for example, that would show the user the location on the map.

The main advantage is that we do not have to store data about countries, locations and we do not have to update it. Another provider of this data, namely *geonames*, is concerned about this. We simply create a link in our ontology in which we state that the property *hasCountryCode* is related to the property *countryCode* from *geonames*. The meaning of part of the data is constantly taken from the geonames; the elements we added through our own ontology can be taken by other users – the signs will thus also be semantically, not only syntactically coherent. When creating an ontology, a good knowledge about the represented domain guarantees the consistency of the represented concepts and eliminates redundancy and confusions.

Conclusions

The web is effective at bringing any resource to the web user, but if the information the user needs is not represented in a single place, the job of integration belongs to the user.

How much from the intended meaning of the message could be provided by using semantic web technologies? Pretty much as much as it is intended to be represented. The power of represented linked data should be in discovering relations in existing represented data. The whole idea of linked data consists in consuming and publishing data. There is no intention to represent a standard/ de facto ontology that is the best in the modeled domain, the whole intention is to link our data with others'.

The main problem remains that of scrapping data from the Web. In order not to scrap for data, every web source provider should have his/her data represented in a standard semantic web format. In this way, semantic web applications could gather data described semantically and share it to the user, as semantic web application developers intended to do.

Our paper represents an approach to semantics and semiotics fields of study from the point of view of final users who are able to receive and understand a message. It depends on every user what message is information or a valuable piece of knowledge, but the role of the ontologies designers is to make them consistent and clear enough.

Regarding this, we assess that it is extremely important to familiarize computer science workers involved in the creation of Web ontologies and semantic Web (which aims to be, in the future, a Web of trust) with the semiotic approach by which signs are related to the entities they represent and to their addressees. In the ontologies domain, we appreciate that the part referring to the syntax is covered by the complete creation of the RDF and OWL standards, and the semantics and pragmatics requirements can be met by the conscious projection of ontologies by their creators. Thus, the individuals who "surf" the Internet are protected from the danger to get lost and confused, by the accurate and well-intended marking of the pages they visit. Since

ontologies also connect computers, we can consider the latter … semiotic bench-marks in the semantic Web.

Acknowledgement

This work was supported by CNCSIS-UEFISCSU, project number PN II-RU code 188/2010.

References:

1. Andersen, P. B. (2000), *Computer Semiotics*, at http://imv.au.dk/~pba/Homepagematerial/publicationfolder/Computersemiotics.pdf
2. Andone, I. (2005), *Ontologiile şi modelarea informaţională a întreprinderii*, in Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi, Ştiinţe Economice, 2005-2006, Tomul LII/LIII; pp. 275-281
3. Borţun, D. (2006), *Semiotica. Teorii ale limbajului*, Bucureşti, 2006
4. Cordeiro, J., Filipe, J. (2004), *The Semiotic Pentagram Framework. A perspective on the use of semiotics within organisational semiotics*, at http://www.orgsem.org/papers/19.pdf
5. Gazendam, H., Liu, K. (2004), *The Evolution of Oganisational Semiotics – A Brief Review of the Contribution of Ronald Stamper*, at http://www.orgsem.org/papers/00.pdf
6. Gazendam, H. (2004), *Organizational Semiotics: a state of the art report*, at http://www.semioticon.com/semiotix/semiotix1/sem-1-05.html
7. Heath, T., Bizer, C. (2011), *Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology*, 1:1, 1-136
8. Kalfoglou, Y., Domingue, J., Motta, E., Vargas-Vera, M., Buckingham Shum, S. (2004), *myPlanet: An Ontology-driven Web-based Personalized News Service. Knowledge Media Institute (KMi)*, The Open University, Milton Keynes, U.K.
9. Lai, M., Yang, X. (2001), *Ontology-based Metadata Schema for Chinese Digital Libraries*, Department of Information Management, Peking University, Beijing, http://www.cs.vu.nl/franksh/postscript/K-CAP01.pdf
10. Lunn, B. (2010), *RDFa Momentum: Facebook Now Publishes RDFa*, at http://semanticweb.com/rdfa-momentum-facebook-now-publishes-rdfa_b649
11. Shum, S.B., Motta, E., Domingue J. (2000), *ScholOnto: An Ontology-based Digital Library Server for Research Documents and Discourse*, International Journal of Digital Libraries. 3(3), pp. 237-248
12. Sowa, J. (2000), *Ontology, Metadata, and Semiotics*, in Ganter, B., Mineau, G.W., eds., *Conceptual Structures: Logical, Linguistic, and Computational Issues*, Lecture Notes in AI #1867, Springer-Verlag, Berlin, 2000, pp. 55-81
13. Uszkoreit, H., Jorg, B., Erbach, G. (2003), *An Ontology-based Knowledge Portal for Language Technology*, German Research Center for Artificial Intelligence and Saarland University, Germany, http://www.mcgreg.net/pub/COLLATE-EnablerElsnet03.pdf.
14. W3C (2009), *W3C Semantic Web Activity News - SPARQL is a Recommendation*, W3.org. 2008-01-15, at http://www.w3.org/blog/SW/2008/01/15/sparql_is_a_recommendation. Retrieved 2009-10-01
15. Yeh, C. (2002), *Development of an Ontology-Based Portal for Digital Archive Services*, Department of Computer Science and Engineering, Tatung University, http://www.iis.sinica.edu.tw/APE02/Program/chingyeh.pdf