Study of Sociocultural Ontology

Papa Fary Diallo, Seydina Moussa Ndiaye and Moussa Lo
Laboratoire d'Analyse Numérique et d'Informatique (LANI),
UFR SAT – Université Gaston Berger,
Saint-Louis –BP 234, Sénégal.
fary85.diallo@gmail.com, seydina.ndiaye@ugb.edu.sn, moussa.lo@ugb.edu.sn

***Abstract*— In this paper, we propose a process of sociocultural ontology development to popularize and perpetuate the culture of a country through a sharing of customs and history of different localities. It can be compared with the construction of a platform that would be straddled between "corporate memory" and a "social network", but applied in the context of a country. This process is based on the theory of Russian psychologist Lev Vygotsky called "Vygotskian Framework". This process allowed us to model our ontology in three axes - Community, Artefact and Infrastructure -, which allowed us to have two levels of social network analysis. An intra-community level allows us to have knowledge within a community and inter-community level, through our index of "similarity of interest," allows us to form clusters of our network.**

Keywords-Social Network; Social Network Analysis; Semantic Web; Ontology; Activity Theory.

I. INTRODUCTION

For lack of knowledge of the Africans territories, it is very common to meet African youth knowing more about the geography of the West than their own countries. Thus, to refresh the memory of our fellow citizens and revive the many stories that accompany the creation and daily life of the different African territories, we initiated the establishment of an online sociocultural encyclopedia.

Our goal is to develop a distributed infrastructure that will allow the Senegalese communities to share their sociocultural knowledge, tourist, economic, educational, agricultural, etc. The infrastructure developed can be compared with a platform which would be straddled between a "corporate memory" (or "memory organization") and a "social network", but applied to the context of a country.

We propose to have a new point of view of the concept of community in the context of the social Web where the community typically represents a set of individuals sharing the same aspirations. Our approach is less focused on individuals (which are classically central points) than on the beliefs and knowledge they share. This shift of view allows us to approach a community as an atomic entity and focus this time on the sharing of knowledge between communities.

Semantic representation is based on a sociocultural ontology of which development is the objective of this paper.

The term "ontology" is borrowed from philosophy, where an ontology is the study of existence. According to [1], in the context of knowledge engineering, ontology "may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definitions and indications of how concepts are interrelated which collectively impose a structure on the domain and constrain the possible interpretations of terms". Such characterization accounts of various objects such as glossaries, terminologies, thesauri and ontologies (in the strict sense), implemented by various professionals (knowledge engineers, librarians, translators) and distinguished according to whether the focus is on the terms and their meanings.

The construction of our ontology is structured around the process "Vygotskian Framework" [2] proposed by the Russian psychologist Lev Vygotsky. This process examines relationship between knowledge and the development of a society in three areas: a) human (subject), b) objects (buildings, park, etc.) and c) artefacts (abstract things).

This paper continues by a state of the art in which we mention the work of the social web and the semantic web that have guided us in our approach. The third section will present our sociocultural ontology with the presentation of Vygotsky theory, concepts and relationships that make up our ontology and we propose an approach to the analysis of our network. We end with a conclusion and perspectives of this work.

II. SOCIAL WEB, SEMANTIC WEB

We place ourselves in the context of setting up a semantic web platform sociocultural ontology-based to enable communities to share their knowledge as in classical social network.

Semantic representation of resources manipulated in our "social network", we allow having rich information contained in the network. On the other hand, the Semantic Web opens up a semantic approach to social network analysis.

We are interested here in the development process of sociocultural ontology. There is not, to the best of our knowledge, this ontology. In this section, we present the work in the field of the social web and the semantic web that have guided our methodological choices.

A. Social Web

There are several definitions of the social web. However, in our study we consider the definition of [3], who defines the social web as an ecosystem of participation, where value is created by the aggregation of many individual user contributions. In our case, contributions that can introduce users will certainly be new structures that are newly set up in

 ISBN: 978-1-61208-163-2

a locality (creating a new instance) and the relationships they have with those existing. Also, provide information on the events is unrolling or will unroll.

Once a social network is constituted, it can be analyzed by the study of social entities well as their interactions and their relationships. This is called social network analysis (SNA). Such analysis is related to the theory of social networks, which designs social relations in terms of nodes and links. The nodes are usually the social actors in the network but they may also represent institutions, and links are relationships between these nodes. This representation, called sociogram, has been proposed in [4]. Among the indicators of a social network, we can cite density and centrality.

Density allows defining the cohesion of a social network. It is defined as the number of existing links divided by the number of possible links. Its value is equal to one if all nodes are interconnected.

Centrality highlights the most important stakeholders of the network. Freeman [5] offers three definitions of centrality: (i) **degree centrality** treats the nodes which have the highest degrees of the graph, i.e., those which have more links in the network; (ii) **closeness centrality** indicates the degree whereby a node is close to all other nodes in a social network (directly or not). It is obtained by calculating the average distance of a node to all other nodes in the network; (iii) **Betweenness centrality** focuses on the ability of a node to serve as an intermediary in a graph. It is the shortest paths between any two nodes that pass through the given node.

However, [5] considers undirected relation. Yet in a social network, a relation-oriented alone contains much semantics. Relations-oriented leads to the concept of **prestige** that is more refined that centrality. We distinguish incoming (in-degree) and outgoing (out-degree) links. An actor is prestigious if it has the highest in-degree. The out-degree of an actor is often a measure of how influential the actor may have.

The emergence of the Semantic Web leads to apply the methods of analysis of networks on new traces generated by the use of the web.

B. Semantic Web

Berners-Lee [6] describes the web of tomorrow as a vast space for exchange of resources between humans and machines allowing exploitation large volumes of information and various services. The current Web is fundamentally syntactic, in the sense that document structure is well defined, but its content remains virtually inaccessible to machines treatment. Only humans can interpret their contents. Thus, the Semantic Web aims at overcoming this difficulty. Web resources are more easily accessible to both the man and machine, using the semantic representation of their contents. The Semantic Web is at first an infrastructure to allow the use of formalized knowledge in addition to the current informal web content, even if there is no consensus on the limits of this formalization. This infrastructure should allow first locating, identifying and transforming resources so robust and healthy while enhancing the openness of the Web with its diversity of users. It must be based on a certain level of consensus, for example, on representation languages or ontologies used. It should help ensure, as automatically as possible, interoperability and transformations between different formalisms and different ontologies. Thus, the Semantic Web provides the opportunity for machines to understand and exploit the resources of the web in an interoperable manner. For this, the W3C offers formalisms provided with XML syntax to model the concepts on the web, to instantiate it and query it [7].

C. The Semantic Web can be social

This question should be asked because papers such as [8][9][10] have defended the importance of social dimension in the construction of a Semantic Web life cycle and have proposed a new approach - the socio-semantic web - Authors radically oppose the traditional approach of the semantic web. In their approach, they subdivide the semantic web into two entities: the computational semantic web and the cognitive semantic web.

According to them computational semantic web "aims fundamentally at automating the search of information using software agents (...) and we will represent the ontologies and semantic networks using formal languages supporting inferences and powerful treatments, such as logical languages or object-oriented" [8] while cognitive semantic web "aims at supporting research activities of human users in complex and evolving corpus" [10]. Thus "while extending this perspective, socio-semantic web is positioned towards the Social Web (...) and it aims at supporting cooperation activities in which more structured interactions also rely on information or documents shared by a collective continuing, at least for a time, common goals" [9]. However, as shown in [11], there is a big difference between the Semantic Web and formal logic. According to [11], the semantic web is a family of languages of increasing expressiveness whose building blocks are not a logical but turn around the RDF model (model of triples to represent graphic descriptions of the resources) and semantic web does not object to the web dimensions semiotic, social or pragmatic. However, since [11], camp of socio-semantic web has changed its approach according to Manuel Zacklad "considering that there was indeed a form of complementary" between both even if he claims that "socio-semantic web initiative is a current particularly within semantic web". In this context, we see two forms of sociality.

The first form is the social network of communities. In this network, W3C formalisms allow us to model our social network, which is consistent with the position taken by Fabien Gandon who argues that "Semantic Web is not anti-social" [11] since the Semantic Web is not a revolution but a web evolution. Moreover, as we intend to use some indicators of social web we will need a powerful query language yet "cognitive Semantic Web does not usually make logically valid inferences automatically" [8].

The second form of sociality is located within the community. For its consideration, it will certainly be necessary to use socio-semantic web for the different views of members of a community that will bring "mutual understanding that encompasses all issues related to cultural

 ISBN: 978-1-61208-163-2

and linguistic to establish an agreement between participants" according to Manuel Zacklad.

D. Semantic representation and social network analysis

FOAF (Friend Of A Friend) project is one of the largest projects on the Semantic Web. FOAF has become a widely accepted standard vocabulary for representing social networks [12][13][14]. However, it is an RDF vocabulary for describing people and the relationships they maintain between them while in our approach we want to model sociocultural knowledge of the different localities. The use FOAF ontology is therefore not appropriate in our context. It is why we propose the use of OWL ontology [15] in our modeling.

OWL (Web Ontology Language), a W3C Recommendation, is a language for defining and instantiating Web ontologies. An OWL ontology may include descriptions of classes, properties and their instances. Classifications expressed in OWL are based on a strict separation class/instance, inheritance of properties, the expression of cardinality constraints and logical constraints on the relationships between properties, etc. OWL provides three increasingly expressive sublanguages designed for use by specific communities of implements and users [15], among which we can mention OWL-DL language that we use. It supports those users who want the maximum expressiveness without losing computational completeness (all entailments are guaranteed to be computed) and decidability (all computations will finish in finite time) of reasoning systems.

Furthermore, social network analysis focuses on the nodes rather than types of node. Thus, we can use it in our case. However, as part of our work to apply the formalism of the Social Web in our ontology, it would be interesting to do so for a specific entity. Because in traditional social network where there is only one object type (people), in this network, it would be interesting to see the nodes that meet the different metrics across the network. In our "social network" where there are several types of concepts, find degree betweenness of the node for example has no real meaning as in conventional networks. However, it would be very interesting for the different components in each locality to see which one is more active by calculating the degree centrality. In the same sense considering the degree centrality, we can know the localities where there are many more sociocultural activities.

However, current approaches to analysis algorithms of social networks are based on definitions and characteristics of graphs representing social networks. The semantics of measured indicators are not taken into account. Social data described in RDF form a typed-graph that provides more powerful and richer representations compared to conventional models for graphical analysis of social networks. The majority of the research aims to calculate the metrics of social networks using the relations "knows" and "interest" of FOAF ontology [16] with the query language SPARQL (SPARQL Protocol And RDF Query Language), a W3C Recommendation, in particular, allows querying of RDF descriptions. However, SPARQL shows some limits on the semantic analysis of social networks. As shown in [17], RDF and SPARQL present all the characteristics for sharing, interoperability, query processing and social data on the web. However, they also show that the standard version of SPARQL is not expressive enough to query "global" on a social network, necessary to calculate the metric of the most SNA. Likewise, SPARQL lacks some key features for building powerful Semantic Web applications. Thus a new version, SPARQL 1.1 [18] in development since March 2009, seeks to rectify these omissions. It adds, among other things, an update language and supports aggregation, subqueries, creating values by complex expressions, extensible value testing, and constraining queries by source RDF graph. This new version is promising for the SNA mainly with the aggregation functionality.

We cannot conclude this section without mentioning works that have been done around Towntology project [19], even if they have not been developed within the Semantic Web. We mention them because their finality - design an urban ontology - seems to be a part of our work, since modeling sociocultural aspects of a community necessarily involves a consideration of the urban aspect. However, during the development of this ontology, designers have felt the need to develop their own language based on XML, so it is impossible to reuse it in our context. Nevertheless, in our modeling we will use some concepts (classes) of the project to build our ontology.

III. SOCIOCULTURAL ONTOLOGY

We propose a methodology to identify features that represent a community in its social aspects (in the broadest sense), modeling of these characteristics will represent our sociocultural ontology. The approach we use is based on a process called "Vygotskian Framework" proposed by Vygotsky. This process examines the relationship between knowledge and development of a society in three areas: a) human (subject), b) objects (buildings, park, etc.) c) artefacts (abstract things).

We mean by methodology, work procedures and steps that describe why and how of conceptualization then of artefact built. By lack of common guideline, there is no “correct” way or methodology for developing ontologies [20]. Thus, we will rely on the Vygotsky process for our methodology.

A. Vygotskian Framework

Vygotsky theory, sometimes called Activity Theory, is a metaphorical space representing the location of cognitive development, a site occupied by subjects, experts, and any other device capable of contributing in development. Activity theory sees human action as being mediated by objects such as tools that carry with them the cultural history of mankind. It describes two processes that are inseparably intertwined: internalization and externalization. Internalization is the process by which culture determines human action and ensures continuity. Externalization is the process by which human actions construct new instruments and forms of activity at collective and individual levels and thereby initiate social change.

 ISBN: 978-1-61208-163-2

Activity Theory is not a methodology; it is a philosophical framework for studying human practices as development processes. It offers, at least in principle, the possibility to conceptualize a scientific way of metacognitive processes, which allows to bind this cognitive development dimension in general and understand the origin of this capacity to control its own internal processes by the schema of Figure 1 and describes the transition from external and interpersonal control to individual intrapsychic control.

Thus, we could say that the Vygotsky theory is a "socio-historical-cultural development theory" [21]. With the three axes of Vygotsky theory, we can model the different concepts of our ontology:

- Subject: as in our "social network" we will replace human by communities. That axis represents communities.
- Object: can be different infrastructures of a community.
- Artefact: can be physical (tools), symbolic (text, taxes) or mental (architectural styles). For our modeling, it will represent cultural activities, historic events and localities of our country.

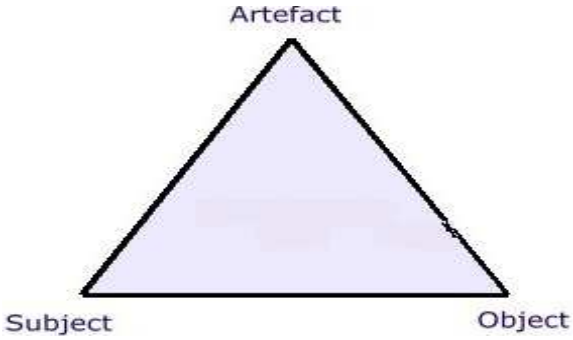

Figure 1. Vygotsky mediating triangle

B. *Concepts and Relationships*

The three axes of the Vygotsky process will be the fundamental classes in our ontology. As shown in Figure 2, an excerpt of the ontology includes main classes and their subclasses. Note *Historic_Site* class is the union of *Built_Area* and *Unbuilt_Area* classes. With these classes and their under classes we can capture all the sociocultural knowledge of a city.

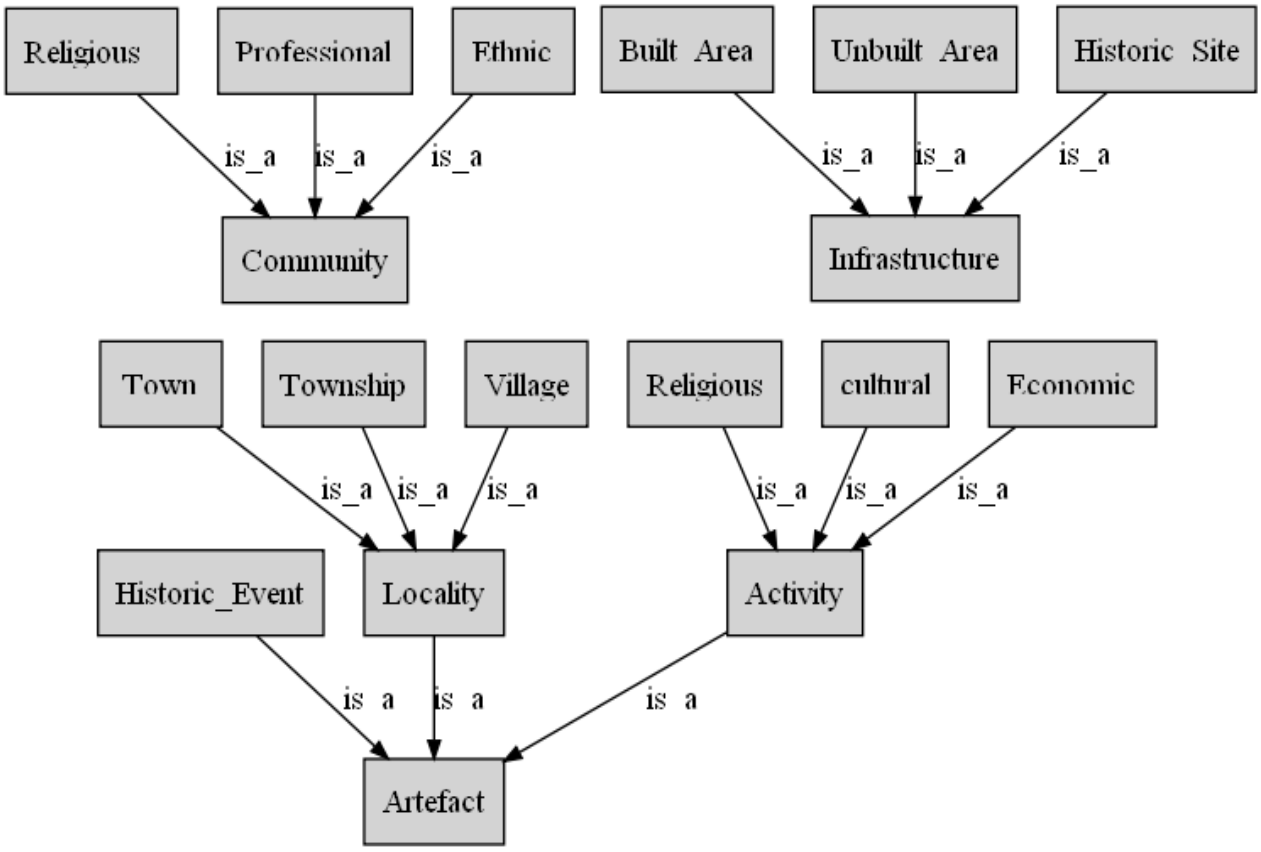

Figure 2. Extract of our sociocultural ontology (Basic Concepts).

 ISBN: 978-1-61208-163-2

Classes alone will not provide enough information. They should be associating of attributes that play an essential role in ontology development. They describe properties of classes and instances. However, due to lack of space, we cannot detail the attributes of our various classes. Nevertheless, we will present two ontologies reused, which are W3C recommendations, OWL-Time ontology [22] proposed for modeling complex temporal phenomena for Semantic Web and GeoRSS-Simple ontology [23] a reference vocabulary for geospatial description of properties of Web resources.

We exploit OWL-Time ontology by defining two relations between, respectively, *Activity* and *Historic_Event* concepts of sociocultural ontology and *DateTimeDescription* and *Interval* concepts of OWL-Time ontology, due to *owl:equivalentClass* constructor of OWL-DL language. With first relationship, the *Activity* concept has properties (attributes) such as *hasBeginning* and *hasEnd* that mark respectively the beginning and the end of an activity. Since the *DateTimeDescription* concept is used to describe intervals implied, such as "May 8, 2007 at 12am 03mn 08s", which represent an interval of 24 hours, with the second relationship we enjoy this type of description for our *Historic_Event* concept.

With respect to GeoRSS-Simple ontology, we define a relationship between the *Infrastructure* concept of sociocultural ontology and the *gml:_Feature* concept of GeoRSS-Simple ontology. Due to *owl:equivalentClass* constructor of OWL-DL language, we get all properties of the *gml:_Feature* concept. Thus, many attributes- box, point, line and polygon- can be used to attach *Infrastructure* instances concrete geometries specified using strings following a certain format. Also we benefit of *WHERE* relationship that can bind *Infrastructure* instances to different geometries of *gml:_Geometry* concept.

Relations are, as classes, most important concepts in ontology development. A design choice that must be made during ontology development is to define when knowledge should be modeled in a property (attribute) or used as relation pointing to another concept. A criterion may be saying that is a property when values are of a type called primitive (integer, string), and it is a relationship when values are of a type said complex, i.e., another concept of ontology. However, this border can be questioned. Thus, Figure 3 illustrates the different relationships that may exist between our classes.

With these relations, we find Vygotsky mediating triangle at different levels. They allow representing different sociocultural knowledge:

- *Organize*, *Localize* and *Occur* relationships allow knowing the different interests of the *Community* based on their *Activity* it organizes. Likewise, we have an idea of the events that occur in a *Locality*.
- *Concern* and *Refer* relationships provide different historical narratives of a *Locality* or *Community*.
- *Is-in* relationship provides different communities of a *Locality*.

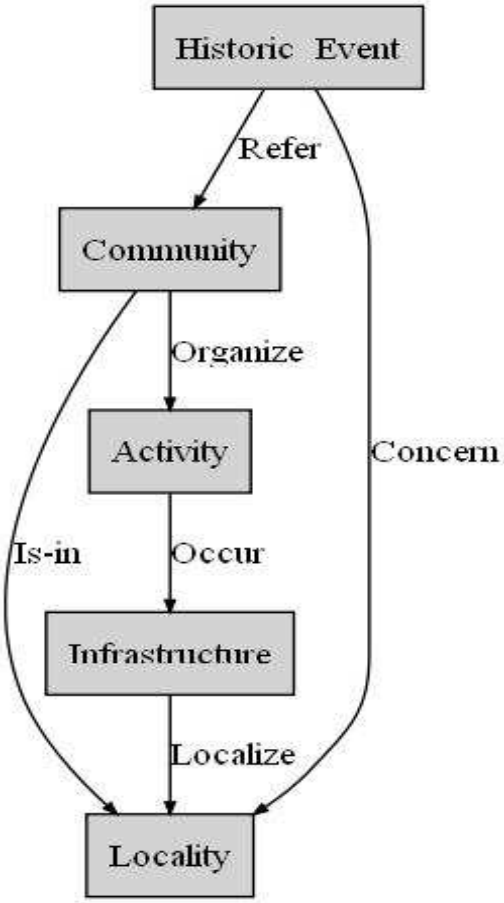

Figure 3. Relationships between classes.

C. "Social network" analysis

Analysis of our "social network" will be made by considering two levels in the definition of metrics.

A first level is to consider a social network within the Community. In this case, the basic elements considered are the different components of a community. Associations that are in place are the components. Considering associations we can calculate the degree centrality to see those that are more active in considering the relationship *Organize*, i.e., those which organize more activities. Similarly, for a locality, we know the various infrastructures and organizations that are there and their numbers thanks to the degree centrality. Therefore, with the degree centrality it is possible to find various information about a locality.

A second level is to consider an inter-community network. In this case, the basic elements considered are communities that make up our "social network". As in our first level we have important activities of the various components of a community, the idea here is to create a new indicator that can show the "similarity of interests" between communities. The similarity measure is built on the various activities within the Community. This indicator allows us to divide our communities clustered according to their center of interest. For example by calculating the

 ISBN: 978-1-61208-163-2

degree centrality of different activities that are organized in a community, if we realize that religious associations are more active we can say that community has a religious interest. This "similarity of interest" can also be calculated using a descriptive vector. Thus from the vector, we can describe the different characteristics for which the similarity of interest will be calculated between communities. For example, we can take a vector whose first field contains the communities in which their cultural activity is between 20% and 25% of their activity, the second field in which economic activity is between 30% and 40% of their activity, etc. We can define several fields and so see the communities that share the same characteristics as this vector.

With different levels of our two metrics, we can have the interests of each community and with our index of "similarity of interests", it will be possible to divide our network according to their centers of interests.

IV. CONCLUSION

In this paper, we presented a method for developing sociocultural ontology in order to popularize and perpetuate the culture of a country through the sharing of customs and history of different localities of the country. This method is based on the process "Vygotskian Framework" which allowed us to model the main concepts of our ontology. Under the Social Web, we defined a new point of view of the concept of community. Thus, in our approach we have substituted a person to a community of people. Thus, focus on the information within a community but also between communities.

As a result, we have divided our system into two levels. The first level gives us rich information at a community level and the second allows us to divide our communities according to their focus through our index of "similarity of interests."

We have just completed a survey in the region of Louga in Senegal and we envision, at first, use the monograph obtained to populate and validate the sociocultural ontology. Likewise, we are going to try to see how to integrate socio-semantic aspects of our study.

Afterwards, we must imagine a semantic web platform around this ontology, a framework for sharing knowledge of Senegalese communities.

REFERENCES

[1] M. Uschold, “Knowledge level modeling: Concepts and terminology”. The Knowledge Engineering Review, Vol. 13: 1, 1998, pp. 5-29. Printed in the United Kingdom. Copyright 1998, Cambridge University Press, 1998.

[2] V. John-Steiner and H. Mahn, "Sociocultural Approaches to Learning and Development: A Vygotskian Framework". Educational Psychologist, 3I (3/4), pp. 191-206. Lawrence Erlbaum Associates, Inc, 1996

[3] T. Gruber, “Collective Knowledge Systems: Where the Social Web meets the Semantic Web”. Journal of Web Semantics, 6(1) pp. 4–13, 2008.

[4] J. L. Moreno, “Emotions mapped by new geography”, New York Times, pp. 17, 1933.

[5] L. C. Freeman, “Centrality in social networks: Conceptual Clarification”. Social Networks. 1, pp. 215-239, 1979.

[6] T. Berners-Lee, J. Hendler and O. Lassila, “The Semantic Web”. Scientific American, Vol. 284, n 5, pp. 34-43, 2001.

[7] W3C. Semantic Web Activity, http://www.w3.org/2001/sw/ and http://www.w3.org/2001/sw/Activity, 30.08.2011.

[8] J. Caussanel, J. P. Cahier, M. Zacklad and J. Charlet, “Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique?” Conférence Ingénierie des Connaissances IC’2002, pp. 3-14, 2002.

[9] M. Zacklad and X. Barbaud, “Vers une application du Web Socio Sémantique pour la réalisation d’un système d’information destiné aux réseaux de santé”, Second séminaire francophone du Web Sémantique Médical, Rouen – 9 mars 2004.

[10] M. Zacklad, “Introduction aux ontologies sémiotiques dans le Web socio sémantique“. Conférence Ingénierie des Connaissances, pp. 241-252, Nice : PUG, 2005.

[11] F. Gandon, “Le Web sémantique n’est pas antisocial”, in actes des 17ème journées francophones de l’ingénierie des connaissances, IC 2006, pp.131-140, 2006.

[12] T. Finin, L. Ding and L. Zou, “Social networking on the semantic web”. The Learning organization journal. Vol. 12, Number 5, pp. 418-435, 2005.

[13] J. Goldbeck and M. Rothstein, “Linking social Networks on the web with FOAF: a semantic web case study”. Proceedings of the twenty-third conference on artificial intelligence, Vol. 2, pp. 1138-1143. 2008.

[14] G. Erétéo, F. Gandon, M. Buffa, P. Grohan, M. Leitzelman and P. Sander “A state of the art on Social Network Analysis and its applications on a semantic web”. Proc. Social Data on the Web, Workshop held at the 7th International Semantic Web Conference, pp. 13-18, 2008.

[15] M. K. Smith, C. Welty and D. L. McGuinness, “OWL Web Ontology Language”. W3C Recommendation 10 February 2004. http://www.w3.org/TR/owl-guide/, 30.08.2011.

[16] G. Erétéo, F. Gandon, M. Buffa and P. Grohan, “Analyse des réseaux sociaux et web sémantique : un état de l’art”. Technical report, Project ISICIL (ANR), 2009.

[17] M. San-Martin and C. Gutierrez, “Representing, Querying and Transforming Social Networks with RDF / SPARQL” ESWC’09, pp. 293-307, 2009.

[18] S. Harris and A. Searborne, “SPARQL 1.1 Query Language”. W3C Working Draft 12 May 2011, http://www.w3.org/TR/sparql11-query/, 30.08.2011.

[19] A. K. Keita, C. Roussey and R. Laurini, “Un outil d’aide à la construction d’ontologies pré-consensuelle: le projet Towntology”. In actes du 24ème congrès d’Informatique des Organisations et Systèmes d’Information et de Décision (INFORSID), pp. 7-22, 2006.

[20] N. F. Noy and D. L. McGuinness, “Ontology Development 101: a guide to creating your first ontology”. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 2001.

[21] I. Ivic, “Lev. S. Vygotsky. Perspectives”, revue trimestrielle d’éducation compare (Paris, UNESCO : Bureau international d’éducation), vol. XXIV, n° 3/4, 1994 (91/92), pp. 793-820, 1994.

[22] J. R. Hobbs and F. Pan, “Time Ontology in OWL”. W3C Working Draft, http://www.w3.org/TR/owl-time/, 30.08.2011.

[23] GeoRSS-Simple Documentation, http://www.georss.org/simple. 30.08.2011.

 ISBN: 978-1-61208-163-2