Ontology-based CMS news authoring environment

Edgard Costa Oliveira
University of Brasília, Software Engineering Faculty UnB Gama
Brasília – Brazil
5561992196093
ecosta@unb.br

Edison Ishikawa
Universidade de Brasília
Dept. of Computer Science
Brasília. Brasil
5561 993928784
ishikawa@unb.br

Lucas Hiroshi Hironouchi
University of Brasília, Software Engineering Faculty UnB Gama
Brasília – Brazil
5561 99903-1143
lucashh@hotmail.com

Thabata Hellen Granja
University of Brasília, Software Engineering Faculty UnB Gama
Brasília – Brazil
5561 98197-7034
thabata.helen@gmail.com

Marcos V. de A. Nunes
Universidade de Brasília
Dept. of Computer Science
Brasília. Brasil
5561 98651-5221
marcosnunesmbs@gmail.com

Daniel Rodriguez
University of Alcalá
Dept of Computer Science
Madrid - Spain
34 918856933
daniel.rodriguezg@uah.es

Rafael Batista Menegassi
University of Brasília
Dept. Computer Science
Brasília – Brazil
5561 992196093
batista7r@gmail.com

Luciano Gois
Brasilia Heath Department
SES-DF, Suplans
Brasília – Brazil
5561 99981-9081
luciano.gois@saude.df.gov.br

George Ghinea
Brunel University London
Dept. of Computer Science UK
& Faculty of Technology, Westerdals Oslo School of Arts, Communication and Technology,Norway
44(0)1895266033
george.ghinea@brunel.ac.uk

ABSTRACT
This paper describes the specification, modelling and design of an ontology-based news authoring environment for the Semantic Web, that takes into account the construction and use of an ontology of the Zika disease. CMSs are being adapted in order to receive semantic features, such as automatic generations of keywords, semantic annotation and tagging, content reviewing etc. We present here the infrastructure designed to foster research on semantic CMSs as well as semantic web technologies that can be integrated into an ontology-based news authoring environment.

Categories and Subject Descriptors
D.3.3 [**Software and its engineering- Software notations and tools**]: - Formal language definitions – *Semantics.*

General Terms
Semantic-based environment; Requirement specifications, Authoring environment architecture;

Keywords
Semantic Web, Ontology, Authoring tool, Content Management System – CMS, Semantic authoring.

INTRODUCTION
Nowadays, text authoring can be seen as a similar practice to those taken 100 years ago, with a slight difference: we have shifted from the hand-pen-paper model in cellulose (that still exists), to the digital finger-keyboard-cursor-white page. In the support level a lot has changed - such as making links to other documents; making and sending as many copies as desired - as we can see from the development of editing resources, which were in the past restricted to editing houses and their complex software. In the syntactic level, we can benefit from searching and ordering key words. However, in the semantic level, text production is the same as before: it depends on the writer´s ability to associate his contents to existing formal concepts structures (links with other documents, links to web pages, associating text to dictionaries, terminologies, taxonomies, indexes, etc).

In the Semantic Web, we are facing a new opportunity to use concept referencing in text – and not only its objects and components such as summaries, images, links, descriptive terms and their meanings. The main problem we are facing today is that the available content on the Web is generated by one person, indexed by another and retrieved by computers that do not make a difference between variant terms.

Particularly, when it comes to preparing a journalistic text, users of CMSs – Content Management Systems – in newsrooms, they count with a blank screen to insert texts with basic formatting options that current editors offer. However, the problem is that these tools limit the use of correct terms, by not giving the author the awareness of using the best term to identify a certain subject as well as its variations. To identify the best keywords to label the subject, to produce tags that are semantically linked, other than hanging loose and ambiguous. This happens to be the case of the subject Zika, disease or virus. The impact of this problem is

978-1-5090-4284-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ICSC.2017.91

related to the news production: they may contain useful information but they were not well represented via keywords or hash tags. Our question is: can we propose an ontology based CMS for the production of news articles that is able to link a term with its semantically related classes or instances in the ontology?

1. OBJECTIVES AND GENERAL METHODOLOGIES OF THIS PROPOSAL

This research began with a general specification of the environment (figure 1), by using as the general requirements of ontology based authoring environments. In a nutshell, the proposal of this tool is to annotate terms and concepts used by writers/journalists and to relate these terms with the ontology of the subject, to create links with other information resources about that subject: existing news pages or any other page selected by the writer.

Apache Jena Fuseki version 1.1.0 [3] was used to overcome the low performance verified by RDFlib. It is a large-scale Java platform, designed for optimized performance. Indexation takes place via semantic annotation. [4,5]. To perform the search in the database, we used SPARQL, which is a standard search language and a data access protocol [1,2]. We used the 101 thodology [10] to create the Zika ontology and we based our research on the structure of existing ontologies and resource documents. The ontology project was constructed using the Web Protegé and is available via the project´s ontology page http://webprotege.stanford.edu/#Edit:projectId=7515ad86-1bdc-431f-a7a3-b9b8167ec068. This ontology was then validated via Hermit reasoner and by the group responsible for its construction, including the medical staff and other specialists in the subject area.

In order to implement the project, we found two viable paths: one using Python [12] and another one based in the Java platform [11]. The Python platform is composed of machines based in an Intel X86 Architecture, Linux Ubuntu O.S., Apache Server, applications in Python and libraries in Python RDFlib for the data treatment in the RDF and Django formats for the CMS [13].

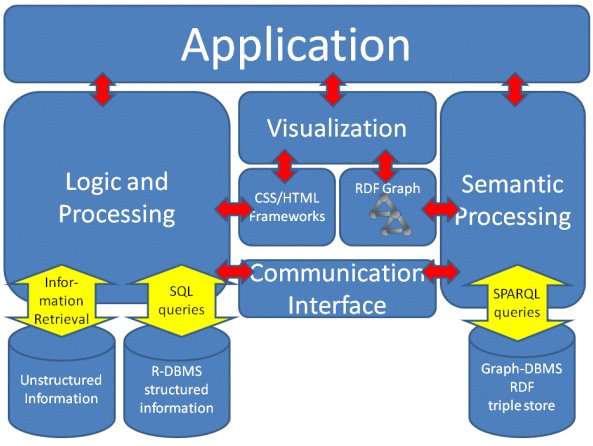

Figure 1. Models showing the use of the application via structural databases and with the use of ontology

2. CONCLUSIONS

We brought here the following contributions to the area: the specification, modelling of an authoring environment for a text editor supported by a semantic and lexical interpreter for the edition of news articles about Zika, supported by a specific ontology created by the students. The following steps in this research is to implement the authoring environment in full, allowing real-time concept recognition from text annotation with the ontology.

This experience resulted in an environment that allows the use of a text editor, integrated to a semantic CMS, in which terms can be typed and in parallel be automatically recognized and associated to classes and instances of the Zika ontology. From the relationships created between the ontology and the text, one is able to obtain for this annotation a list of keywords and conceptual #tags that identify specific subjects of the text, the scope of the article in relation to the general context of the Zika ontology. It also correlates the text with already existing texts and articles or pages so that they can be interconnected via non ambiguous semantic relationships.

3. REFERENCES

[1] F. Gandon,; G. Schreiber (2014). RDF 1.1 XML Syntax. Rio de Janeiro: W3C, 2014. Disponível em: http://www.w3.org/TR/rdf-syntax-grammar/. Acess in May 2016.

[2] M. Schiessl. Lexicalização de ontologias: o relacionamento entre conteúdo e significado no contexto da recuperação da informação. 2015. 261 f., il. Tese (Doutorado em Ciência da Informação) Universidade de Brasília, Brasília, 2015.

[3] Apache Jena. A free and open source Java framework for building Semantic Web and Linked Data applications. Configuring Fuseki http://jena.apache.org/index.html> Access in May 2016.

[4] W3C. RDF Schema 1.1 Recommendation 25 February 2014. http://www.w3.org/TR/rdf-schema/ .Access in April 2016.

[5] M. C. Daconta, L. J. Obrst, K. T. Smith. The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management. Willey, 2003.

[6] N. Noy, D. McGuinness, D. (2001) "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford University.

[7] S. Isotani e I. Bittencourt (2015) "Dados Abertos Conectados", Novatec Editora, São Paulo.

[8] J.Talas, T.Gregar, and T.Pitner (2011). Semantically Enriched Tools for the Knowledge Society: Case of Project Management and Presentation. In: Knowledge Management, Information Systems, E-Learning, and Sustainability Research Volume 111. Springer, 2011.

[9] IKS. Developing Semantic CMS Applications: The IKS Handbook (2013). Editors WernherBehrendt and VioletaDamjanovic. Salzburg Research Forschungsgesellschaftm.b.H. 2013.