Open Educational Resources Ontology

Ricardo J. Rocha Amorim[1,2], Thomas A. Rabelo[1], Dinani Gomes Amorim[1,3]

[1]Faculdade de Ciências Aplicadas e Sociais de Petrolina (FACAPE)
Campus Universitário S/N - Petrolina- PE – Brazil

[2]Departamento de Educação - Universidade do Estado da Bahia (UNEB)
Senhor do Bonfim, BA - Brazil

[3]Departamento de Tecnologias e Ciências Sociais - Universidade do Estado da Bahia (UNEB) – Juazeiro, BA- Brazil

{amorim.ricardo,trabelo,dinaniamorim}@gmail.com

Abstract. *This paper proposes an ontology based on the specifications of DocBook and LOM metadata, used as consensual knowledge for the semantic description of educational content in order to facilitate editing, remix and share of Open Educational Resources.*

1. Introduction

In education context, the abundance of educational resources has attracted the interest of experts in the application of methodologies, techniques and teaching practices based on collaboration and social interaction in which learning gains another meaning, so that new skills and abilities are demanded. In this sense, Litto (2006) discusses about a new ecology of knowledge based on open content which aims to increase the democratic access of knowledge and Weller (2011) presents the pedagogy of abundance based on free, varied and abundant content, produced collaboratively from social interaction. In the same way, Brown and Adler (2008) defend an open education based on social learning and free educational resources where both the content and the process where it is created are equally visible. This educational practice is based on the use of Open Educational Resources (OER) [Caswell et al. 2008; Santana et al. 2012], where the emphasis is on the selection, the aggregation and the interpretation of existing materials, much more than the development of specific material.

However, selecting appropriate material to a particular learning situation is not a simple task, considering the abundance of existing resources on the Internet. Locating resources takes time and, in many cases, is applicable only on limited and specific situations. Therefore, a great problem is to find the most appropriate content to each situation of teaching and learning. Furthermore, the reuse of a part of a learning material, in many cases, can be a complex task, considering that certain formats are closed and require specific tools for editing.

Thus, we propose a semantic description model of educational materials as possible solution to address these issues in the OER scope. It is based on *DocBook* [Walsh and Muellner 1999] and LOM [IEEE-LOM 2005], metadata specifications that facilitates the development of applications focused on the selection and aggregation of existing materials associated with a particular educational context. With this model, during the classes planning, teachers could use a specific application that produce

indexed educational resources easier to find and use, from independent and reusable atomic pieces that are aggregated in the OER composition.

In this paper, section 2 discusses about the concept of OER, its main open issues and possibilities; section 3, about ontologies; section 4 describes *DocBook*, shown as consensual knowledge to the OER representation; section 5 presents ONTOER, a proposed *DocBook* based ontology that describes and enables easy editing, remix and share OER; and, finally, section 6, is about the ONTOER application and future work.

2. Open Educational Resources

The movement of production and distribution of free and open resources for education, from which originated the OER - Open Educational Resources, arose in the context of a open knowledge culture originated in the 90s, that is based on freedom use, reuse and (re)distribution of knowledge without legal restrictions, social and technological - the same principles applied to the Free Open Source Software (FOSS) [Crowston and Howison 2005]. The OER have many common aspects to Learning Objects [Hodgins 2002], designed with the idea of producing materials for reuse easily in a wide range of teaching and learning situations.

The concept of OER was defined from the work developed by the Hewlett Foundation in order to democratize the access of quality educational content. It involves educational resources of public domain or that have been distributed under an intellectual property license allowing their free use. This concept includes full courses, didactic materials, textbooks, video streaming, tests, software and any other tools, materials or techniques used to support access to knowledge [Santana et al. 2012; Litto 2006]. Thus, OER are educational materials which allows use, modify, reuse (remix) and freeform share.

Currently, many studies have been dedicated to the development of tools and techniques oriented to OER production. However, many semantic issues related to searching and navigation schemes of content with an educational context associated, are still open. In this sense, some technological solutions of Semantic Web such as ontologies [Barros et al. 1998; Gómez-Pérez et al. 2004] and metadata [Greenberg and Robertson 2002] can be applied in the context of education to facilitate search and reuse of applicable content in a given situation of teaching and learning. Nowadays, in the abundance world [Baumgartner et al. 2007; Sicilia 2006; Zhong et al. 2007], tools oriented to facilitate search and reuse of content are crucial. In the preparation of online learning materials they are justified as means of minimize effort and time consuming.

3. Ontology

From the philosophy perspective, ontology is a term that defines the existence of a systematic way. In computer science, an ontology can be described through the definition of a representative terms set (eg., classes, relations, functions and other objects) with texts in natural language that describes the meanings of names, formal axioms that confirm the interpretation and effective use of these terms [Gruber 1993; Lu et al. 2002]. In the context of online learning, there are widely accepted metadata standards that can be used to facilitate the structuring, search, use and reuse of educational resources [IEEE-LOM, 2005; IMS-LD, 2003]. However, these standards contains many natural language definitions that make the implementation of application

software for such a complex task and prone to misinterpretation which can easily be implemented by programmers. It is a kind of error that produces inconsistencies between different software that implement those standards. To solve this problem, the ontologies [Gómez-Pérez et al. 2004] allow describe formally and explicitly the structure and meaning of metadata elements [Amorim et al. 2006; 2007b].

4. DocBook

DocBook is a markup language defined in SGML (Standard Generalized Markup Language) or XML Document Type Definition (DTD). The *DocBook* DTD defines a particularly very well defined vocabulary for books and documentation about computational hardware and software, although not limited to this kind of use. *DocBook* was originally created as an SGML application developed by O'Reilly and Hal Computer Systems in 1991 aiming to facilitate the exchange of software manuals [Walsh and Muellner 1999]. It has been widely used and tested by companies like SUN [Oracle 2012], Microsoft [Microsoft 2012], HP [HP 2012], Novell [Novell 2012], SCO and Caldera [Sco 2012], Red Hat [Redhat 2012], and Linux Documentation Project [Linuxdoc 2012]. This specification was conceived from requirements that fit with the OER principles [Santana et al. 2012]:

- Content reuse and adaptation. *DocBook* is based on the separation of content from presentation. This feature allows the reuse of information in multiple presentation formats. Moreover, the reuse of content in pieces from different formats facilitates adaptation.
- Content sharing. *DocBook* documents can be built in a neutral way that facilitates the exchange of content independently of authoring tool or proprietary format.
- Content modularization. *DocBook* allows the creation of content based on specific needs and structured in units that can be automatically packaged.
- Support automation. *DocBook* is an open format - open source specification with a wide variety of documentation available. Moreover, several authoring tools and other free resources have been developed aiming facilitate the automation of processes involving the creation, use and reuse of documents.

By the other side, *DocBook* is modular and extensible. Despite the *DocBook* DTD be large and complex considering various options and features contained in their technical documentation, this can be adjusted (reduced) to the specific characteristics of each particular project. Also, it is possible to extend the language to add new tags or attributes if necessary. A software tool that implements these features of DocBook, plus functionalities for the semantic annotation of contents could make searching, sharing and reuse of OER easier. It is better discussed in the following sections.

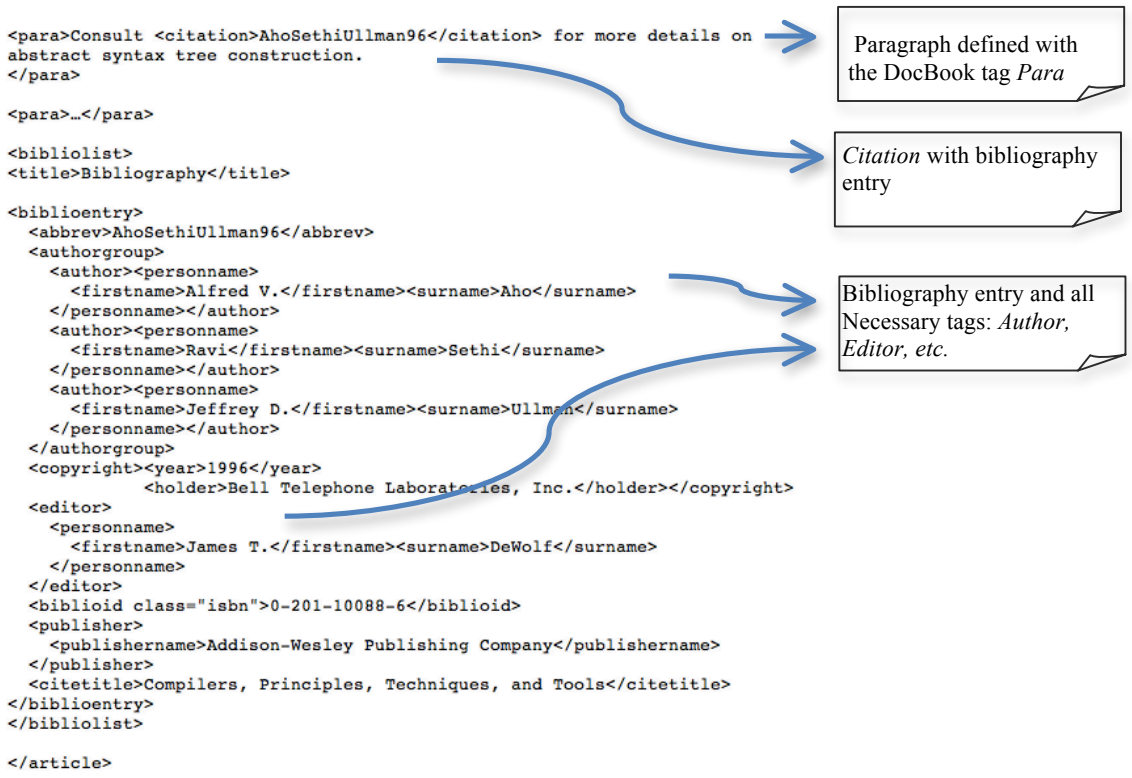
5. ONTOER: Open Educational Resources Ontology

For the ontology modelling, we have considered as crucial in creation, reuse, remix and sharing of OER between different platforms, the following aspects: the semantics of terms, relations, constraints, and the interoperability of the material produced. These factors were determinant in the choice of *DocBook* and the LOM standard. In addition, we have considered that ontology requires consensus in communities related to the domain knowledge. In this case, such specifications represent consensus, considering their wide acceptance in industry and academia: (i) *DocBook* is a stable specification,

has been adopted nearly two decades ago, consisting of a large amount of markup tags (around 380 tags and 3053 entities). It describes various structures and relations of books appropriated to compose more general concepts as learning content; and, (ii) the LOM standard was chosen taken in account that it was specified to facilitate search, evaluation, acquisition and use of learning objects for students, instructors or automated software processes.

The ontology modelling aimed to obtain a knowledge model for building applications oriented to search, use and reuse of educational contents based on existing semantics between their elements. Therefore, an application that implements the ontology could be used in developing of educational materials from other materials developed by recognized authors in a way to ensure the quality of the material obtained without incurring extra effort to the teacher (or contents Designer).

Thus, we carried out a study about the elements described in *DocBook* model (Figure 1) from which we have obtained a set of terms and relations between them (they were implicit) that we have used in the description of taxonomy of concepts and its constraints. This study along with issues, choices and solutions is better described in Amorim (2007a). For example, in a given article the quotation from authors in other published papers is common. In this case, a paragraph with a quotation and reference in the bibliography is defined according the following way in *DocBook*:




```
<para>Consult <citation>AhoSethiUllman96</citation> for more details on
abstract syntax tree construction.
</para>

<para>…</para>

<bibliolist>
<title>Bibliography</title>

<biblioentry>
  <abbrev>AhoSethiUllman96</abbrev>
  <authorgroup>
    <author><personname>
      <firstname>Alfred V.</firstname><surname>Aho</surname>
    </personname></author>
    <author><personname>
      <firstname>Ravi</firstname><surname>Sethi</surname>
    </personname></author>
    <author><personname>
      <firstname>Jeffrey D.</firstname><surname>Ullman</surname>
    </personname></author>
  </authorgroup>
  <copyright><year>1996</year>
            <holder>Bell Telephone Laboratories, Inc.</holder></copyright>
  <editor>
    <personname>
      <firstname>James T.</firstname><surname>DeWolf</surname>
    </personname>
  </editor>
  <biblioid class="isbn">0-201-10088-6</biblioid>
  <publisher>
    <publishername>Addison-Wesley Publishing Company</publishername>
  </publisher>
  <citetitle>Compilers, Principles, Techniques, and Tools</citetitle>
</biblioentry>
</bibliolist>

</article>
```

Figure 1. *DocBook* tags for a paragraph with quote and bibliography reference.

The IEEE LOM standard has been associated to the *DocBook* terms to allow that the educational context related to a given content can be shown (in an automatic or semi-automatic processing). Figure 2 shows some *DocBook* terms from which we have identified implicit relations such as taxonomic and restrictions relations:

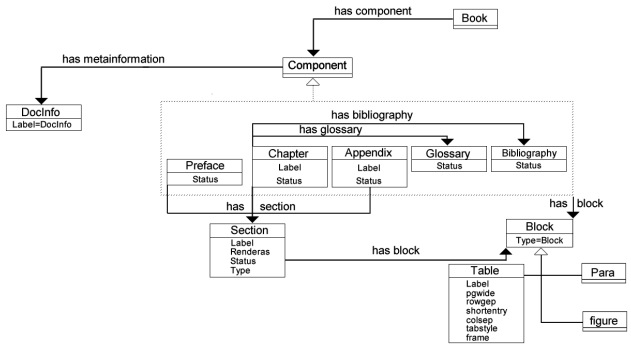


Figure 2. *DocBook* tags and their associated taxonomy [Amorim 2007a].

As shown in Figure 2, the *DocBook* terms and relations were used to specify the components and structures constituting educational learning resources (*Section, Table, Para*, *Figure*, etc). With it, the *DocBook* elements were organized in learning object taxonomy (Figure 3):

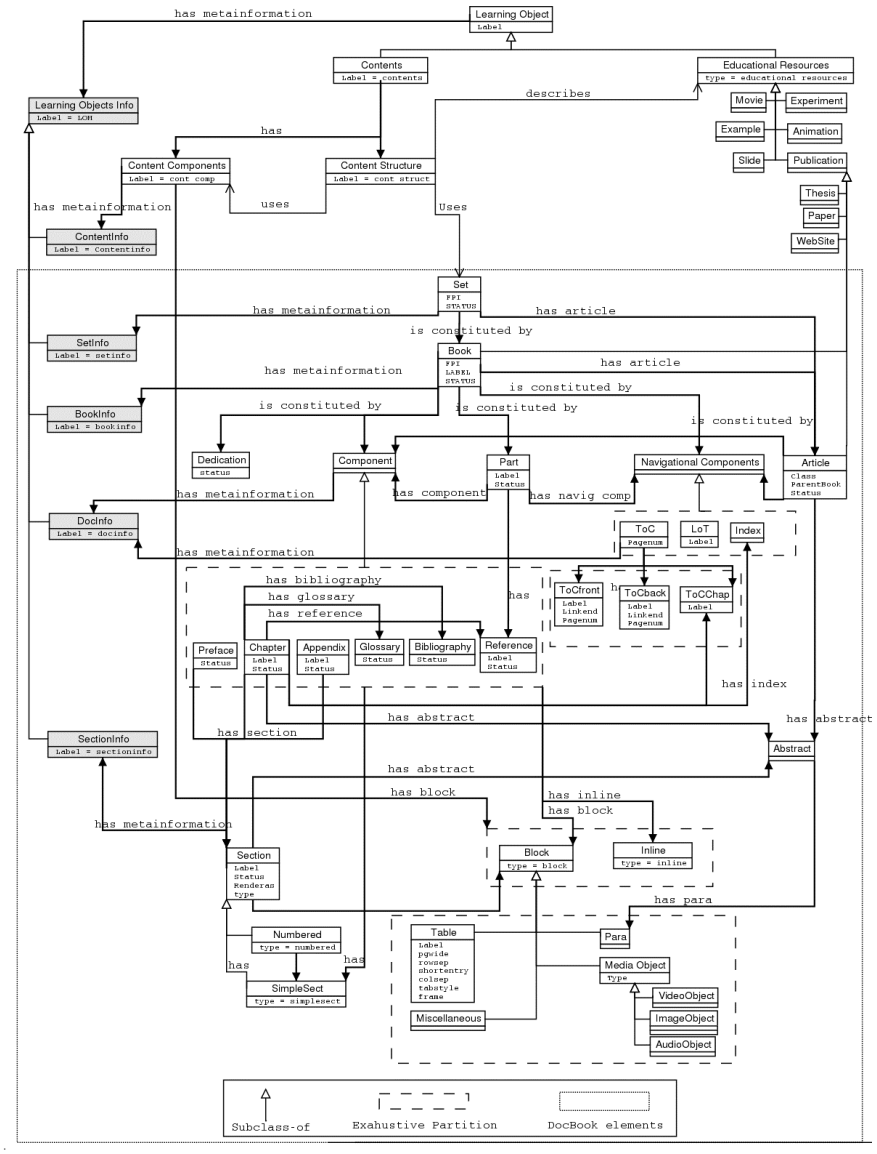


Figura 3. DocBook terms in learning object taxonomy [Amorim 2007a].

In Figure 3, for the description of learning objects we use the element *Docinfo* from *DocBook* associated with LOM metadata elements as a standardized way of describing this kind of educational resource; the *Blocks* are fragments or atomic units that can't be decomposed. For example: an *Article* can be composed by *Figures* and *Tables,* which can belong to a *Website*, and *Paragraphs*, which can belong to a *Chapter Section* of a *Book*. The granularity of a fragment is a nontrivial problem, as it depends on the author's perception [Amorim 2007a].

5.1 Axioms

In the description of tags in *DocBook*, there are many implicit relations between terms that were used in the composition of axioms. It's because there some restrictions that are complex to be defined with other ontology components. These restrictions can be used as rules to be followed or recommendations in an educational resource edition. For example, in a same document a *Para* instance should be unique (in first order logic):

$$\forall\ p, c1, c2, c3 \ldots cn \mid p \in Para \wedge c1, c2, c3, \ldots cn \in Chapter \rightarrow \neg\exists\ p \mid [p \in c1 \wedge (p \in c2 \vee p \in c3 \vee \ldots p \in cn) \vee p \in c2 \wedge (p \in c1 \vee p \in c3 \vee \ldots p \in cn) \vee p \in c3 \wedge (p \in c1 \vee p \in c2 \vee \ldots p \in cn) \vee p \in c1 \wedge \ldots p \in cn \wedge (p \in c1 \vee p \in c2 \wedge p \in c3) \quad (1)$$

And, all *Citation* in a given document must find one and only one corresponding input in Bibliography:

$$\forall\ p, ci, b \mid p \in Para \wedge ci \in Citation \wedge b \in Bibliography \wedge \exists\ ci(p) \wedge \exists\ ci(b) \rightarrow ci(p, b) \quad (2)$$

With the axioms, classes and the related relations and constraints of *DocBook* are described in a more precise way in the ontology. For limitations of space reasons, other defined axioms are not mentioned in this paper.

6. Discussion and future work

In last years, several studies were developed aiming to facilitate the production of educational resources using metadata standards. From these, we highlight some software applications mentioned in Griffiths et al. (2005) and Silva and Santanchè (2009): (i) the Komposer Suite, for authoring of educational materials using Microsoft Word along with an application that converts the document produced in a package of IMS Content Packaging [IMS CP 2007]; (ii) the semantic Word that proposes a way to facilitate the semantic annotation of content in order to reduce the effort made by the author; and, (iii) ARARA is a proposal of an optimized process for authoring of complex digital objects from the semantic annotation using educational standards.

However, considering the scope of this study (section 1), such proposals do not address some questions referring to OER (section 4) or the education domain: for example, Komposer Suite and Semantic Word are platform dependent; the ARARA produces documents from the language XML Schema of the specification IMS LD [IMS-LD 2003] which is inappropriate for representing most of the semantics described as human language text in the Information Model document of this specification [Amorim et al. 2006]. Thus, this paper proposes a solution from consensual knowledge in an ontology based on *DocBook* and LOM, described with the language OWL that is recommended by the W3C consortium [WWW-CONSORTIUM 2004].

With the ontology, during edition, the resources are structured according this consensual knowledge where there are syntactic rules and associated semantic: for example, usually is not admissible instance of the same paragraph into two different sections, likewise, the same section in separate chapters. In these situations, relations of disjunction between such concepts and axioms can be applied.

Moreover, implementing of an application dedicated to select and content aggregation can be facilitated, taking into account that part of the necessary code may be obtained from a translation of OWL to the development language used. For this purpose (Figure 4), we have used the Protégé tool [Noy et al. 2000]:

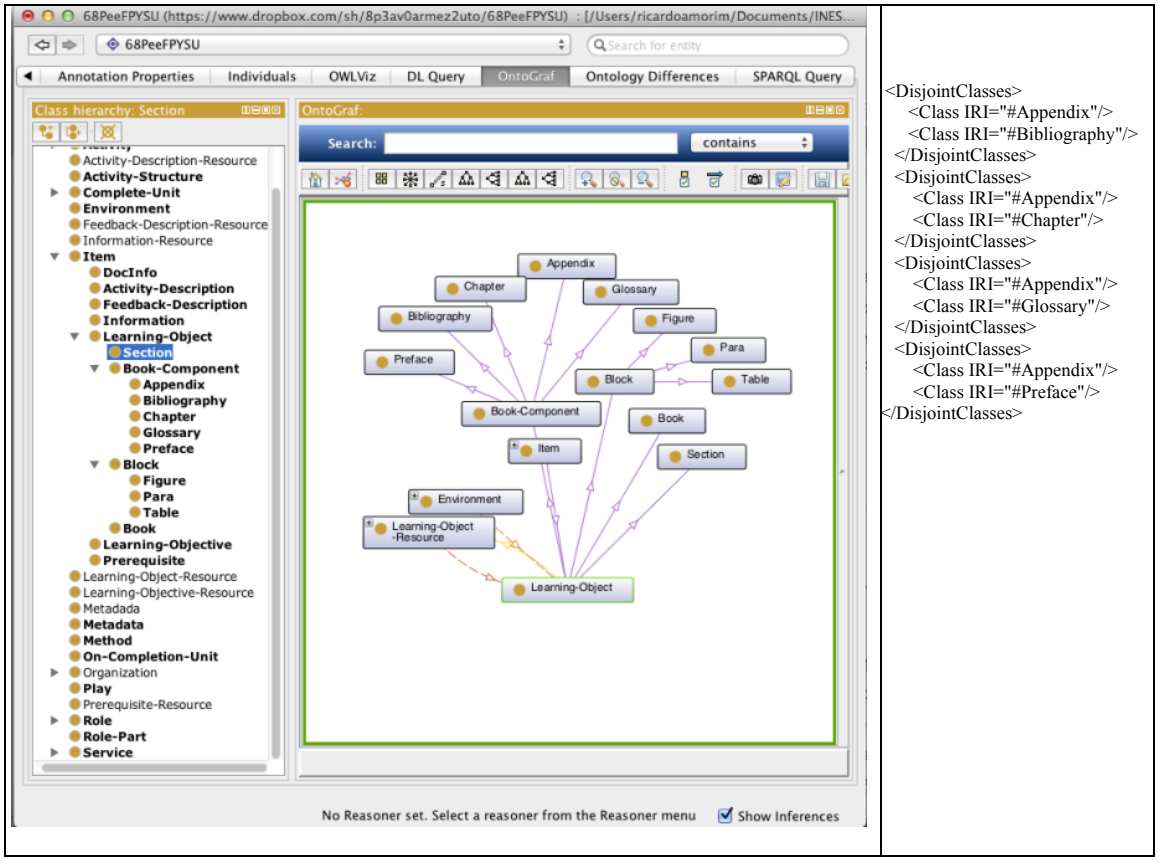


Figure 4. ONTOER edition and OWL code of Appendix disjoint classes.

This ontology could serve as a basis for some editing tool of open materials, including Komposer Suite, Semantic Word and ARARA. In this case, during the edition of a given educational material with MS Word, using a search engine (on the right and below in Figure 5), the ontology would be used both to facilitate their location as in structuring this as a learning object in accordance with *DocBook* and LOM standard.

Figure 5 on the left shows a hierarchy of elements that are instances of *Section, Para* and *Figure* whose content is edited in a natural way on a sheet in the MS Word (center). These instances are organized according the taxonomy presented from a Protégé application with a graphical view of key concepts and relations (right).

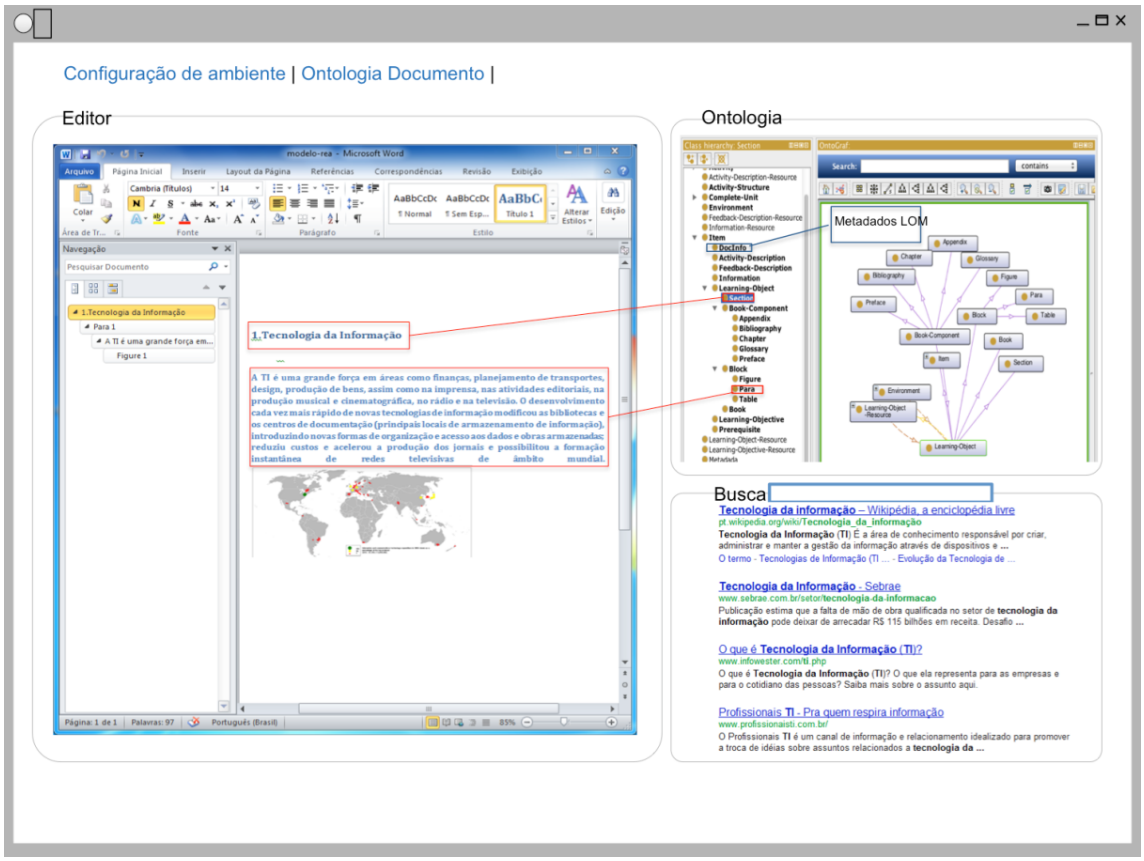


Figure 5. Prototype application for OER editing with MS Word and ONTOER.

According Figure 5, the subtitle "1. Tecnologia da Informação "(left and center) is configured as an instance of *Section*, whose content is an instance of *Para* and the map, an instance of *Figure* (right). For each of these concepts there is a *Docinfo* associated with the same structure of the LOM metadata to be used in search engines and in the aggregation with other learning objects.

As future work, we intend to proceed with the study of *DocBook* in order to identify more relations, constraints, and axioms that allow the description of OER in a more precise way; and, to carry out a study with the aim of evaluating techniques of metadata annotation, in an automatic or semi-automatic way applicable to OER; and, finally, to identify the requirements and use cases for building an OER editing application based on ONTOER and user centered techniques.

Acknowledgements

This work was supported by Instituto Nacional de ciência e tecnologia para Engenharia de Software (INES), funded by CNPq and FACEPE, **processes 573964/2008-4** and **APQ-1037-1.03/08**.

References

Amorim, R.(2007a).“Modelo del conocimiento y arquitectura para la integración de metodologias educativas y nuevas tecnologias en el aula”. Thesis (Ph.D, in Eletrical and Computer) – University of Santiago de Compostela, Santiago de Compostela, Spain.

Amorim, R. and, Lama, M. and, Sánchez, E. and, Amorim, D. (2007b). "Modelling and executing units of learning using an IMS LD ontology". In: *First Brazilian Workshop on Semantic Web and Education (WSWEd)*, *XX Simpósio Brasileiro de Informática na Educação (SBIE)*, São Paulo, Brazil, November 28-30.

Amorim, R. and, Lama, M. and, Sánchez, E. and, Riera, A. and, Vila, X.A.(2006), "A Learning Design Ontology based on the IMS Specification". Educational Technology & Society, 9(1):38-57.

Barros, F. and, Gonçalves, P. and Santos, T. (1998). "Providing context to web searches: the use of ontologies to enhance search engine's accuracy". *Journal of the Brazilian Computer Society,* 5(2), pp. 45-55.

Baumgartner, P. and Naust, V. and Canals, A.(2007), et al. "Open Educational Practices and Resources: OLCOS Roadmap 2012". OLCOS Project.

Brown, J.S. and Adler, R.P.(2008). 'Open education, the long tail, and learning 2.0". Educause review, 43, number 1, pages 16-20.

Caswell, T. and Henson, S. and Jensen, M. and Wiley, D.(2008). "Open content and open educational resources: Enabling universal education". The International Review of Research in Open and Distance Learning, volume 9, number 1.

Crowston, K. and Howison, J.(2005). "The social structure of free and open source software development". First Monday, vol. 10, number 2-7.

Greenberg, J. and Robertson, W. (2002). "Semantic web construction: an inquiry of authors' views on collaborative metadata generation". In: *Proceedings of the 2002 international Conference on Dublin Core and Metadata applications (DCMI '02)*.

Griffiths, D. and Blat J. and Garcia R. and Vogten H. and Kwong KL (2005). "Learning Design Tools". Learning Design: modelling and implementing network-based education & training, Springer Verlag. p. 109-135.

Gómez-Pérez, A. and, Fernando-López, M. and, Corcho, O.(2004). "Ontological Engineering". Springer Verlag: Berlin.

Gruber, T. R. (1993). "A translation approach to portable ontologies", Knowledge Acquisition, Vol. 5, No. 2, pp. 199-220.

Hodgins, H. W.(2002). "The Future of Learning Objects", *In: e-Technologies in Engineering Education: Learning Outcomes Providing Future Possibilties*, Jack R. Lohmann and Michael L. Corradini Ed., ECI Symposium Series, Vol. P01.

HP(2012). Hewlett-Packard Development Company. Available in: <http://www.hp.com/>, accessed on 03/09/2012.

IEEE-LOM (2002). Draft Standard for Learning Object Metadata. Available in: < http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>, accessed on 03/09/2012.

IMS CP. IMS Content Packaging Information Model (2007). IMS Global Learning Consortium, Inc. Available in: <http://www.imsglobal.org/content/packaging/>, accessed on 03/09/2012.

IMS-LD (2003). IMS Global Learning Consortium, IMS Learning Design Information Model. Versão 1.0 Final Specification. Available in: <http://www.imsglobal.org/learningdesign/ldv1p0/imsld-infov1p0.html>, accessed on 03/09/2012.

Linuxdoc(2012). Available in: <http://www.linuxdoc.org/>, accessed on 03/09/2012.

Litto, F.(2006). "A nova ecologia do conhecimento: conteúdo aberto, aprendizagem e desenvolvimento". Inclusão Social, Brasília, v.1, n.2, 73-78, abr/set.

Microsoft (2012). Available in: <http://www.microsoft.com/>, accessed on 03/09/2012.

Lu, S. and Dong, M. and Fotouhi, F.(2002) "The Semantic Web: opportunities and challenges for next-generation Web applications". *Information Research* 7(4). Available in: <http://InformationR.net/ir/7-4/paper134.html>

Novell(2012). Available in <http://www.novell.com/>, accessed on 03/09/2012.

Noy, N. and Fergerson, R., and Musen, M. (2000). "The knowledge model of Protege-2000: Combining interoperability and flexibility". In: 12th *International Conference in Knowledge Engineering and Knowledge Management* (EKAW), October 2-6, Juan Les Pins, France.

Oracle (2012). Available in: <http://www.oracle.com/> accessed on 03/09/2012.

Redhat(2012). Available in: <http://www.redhat.com/>, accessed on 03/09/2012.

Santana, B. and Rossini, C. and Pretto, N.(2012). "Recursos Educacionais Abertos: práticas colaborativas políticas públicas". Editora da Universidade Federal da Bahia - Edufba: Salvador.

Sco(2012). UnXis Inc., Available in <http://www.sco.com/>, accessed on 03/09/2012.

Sicilia, M.A.(2006), "Metadata, semantics, and ontology: providing meaning to information resources", *International Journal of Metadata, Semantics and ontologies*, volume 1, number 1, pages 83—86.

Silva, L. and Santanchè, A. (2009). "ARARA: Autoria de Objetos Digitais Complexos Baseada em Documentos". In: *XX Simpósio Brasileiro de Informática na Educação (SBIE)*, Florianópolis – SC, ISSN: 2176-4301.

Walsh, N. and Muellner, L.(1999). "Docbook: The Definitive Guide". O'Reilly, Sebastopol, C.A, USA, 1999. Available in <http://www.xml.com/pub/a/1999/10/docbook /docbook-making.html>.

Weller, M.(2011). "A Pedagogy of abundance". *Spanish Journal of Pedagogy*.

WWW-CONSORTIUM(2004). "OWL Web Ontology Language Reference". W3C Recommendation 10 Feb, 2004. Available in <http://www.w3.org/TR/owl-ref/>. Accessed on 18/09/2012.

Zhong, L. and Wang, H. and Li, R. and Song, H.(2007). "Implementation of Ontology in E-learning Resource Repository Based on Semantic Web". Jisuanji Gongcheng/ Computer Engineering, vol.e 33, num. 8, pages 282-284.