MULTILINGUAL ACCESS TO CONTENT THROUGH CIDOC CRM ONTOLOGY

Lais Barbudo Carrasco (IESF) - laiscarrasco@hotmail.com

(Instituição - a informar) - manfred.thaller@uni-koeln.de

(Instituição - a informar) - joaquim@uc.pt

Resumo:

The CIDOC CRM ontology can have it vocabulary translated into other languages in order to achieve the multilinguality, i.e. fosters a multilingual access to information as well as provides a higher information retrieval for the user's query.

Palavras-chave: *Multilinguismo, Ontologia, CIDOC CRM.*

Área temática: *Temática I: Tecnologias de informação e comunicação – um passo a frente*

MULTILINGUAL ACCESS TO CONTENT THROUGH CIDOC CRM ONTOLOGY

1 INTRODUCTION

With the recently fast diffusion of distributed document base over the international computer networks of World Wide Web, the question of multilingual access and multilingual information retrieval is becoming increasingly relevant. Because of this reason, it is important to emphasize that Multilingual Access to Content is used to increase and enhance the users' possibilities to access the Cultural Heritages Repositories and their content in their native or preferred language.

The cultural heritage of European countries is usually described in many different national languages. We would therefore need a way, in which we can describe interfaces of computer systems, which can easily be operated in very many different national languages, so they give access to material from all countries.

> Multilinguality in ontologies has become an impending need for institutions world wide with valuable linguistic resources in different natural languages. Since most ontologies are developed in one language, obtaining multilingual ontologies implies to localize or adapt them to a concrete language and culture community. (CEA, 2008, p. 67)

In simple terms, ontology is seen as a controlled vocabulary, glossary or taxonomy.

> The CIDOC CRM promotes a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modeling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (http://www.cidoc-crm.org/)

The main objective of this work is to create a multilingual system from the translated labels of the CIDOC CRM vocabulary.

2 MATERIALS AND METHODS

The CIDOC Conceptual Reference Model (CRM) is an ontology from the

cultural heritage used as the mediated schema to provide multilingual information retrieval in cultural repositories. In order to create multilingual interfaces for digital cultural heritage resources/repositories and contrast an English vocabulary with a Portuguese version, by a mechanism which could be easily generalized to more languages. This process will be handled using ontologies, their interrelationships and the technologies necessary to apply them, not by applying a translation label in the usual sense. Therewith, the fundamental principles of the Multilingualism will be taken into consideration. Hence the research is based on Semantic Web technologies as the RDF/XML.

3 FINAL RESULTS

It was built from the CIDOC CRM vocabulary a correspondent Portuguese vocabulary version for a possible search interface handled by user. Each CIDOC CRM entity was translated to Portuguese, in order to make to make the system multilingual. Multilingualism increases the access to knowledge and achieves more people. The translation process wasn't based in the literal translation, but on equivalence of the terminology and it conceptual meaning.

For example, the CIDOC CRM Entity E5 Event was translated in Portuguese as E5 Evento, the Entity E8 Acquisition was translated in Portuguese as E8 Aquisição, the Entity E16 Measurement was translated in Portuguese as E16 Medida, the Entity E35 Title was translated in Portuguese as E35 Título, the Entity E39 Actor was translated in Portuguese as E39 Autor, the Entity E53 Place was translated in Portuguese as E53 Lugar, the Entity E73 Information Object was translated in Portuguese as E73 Objeto Informacional, the Entity E56 Language was translated in Portuguese as E56 Idioma. Hence, the Portuguese vocabulary was built in order to provide a multilingual system.

4 FINAL CONSIDERATIONS

Given the enormous flow of information available on the Internet in different languages, and regarding the possibility of users to search independently their native language and retrieve relevant information, the question of multilingual access and multilingual information retrieval has become of great significance.

Hence, the World Wide Web should provide systems that are easy to navigate, with flexible tools which help and orient the user in the search for information. And by this the users can access a wide diversity of unrestricted information sources that give them the opportunity to select and discard, retrieving exactly those texts that are of interest.[1]

In summary, multilingual information access allows users to search for information produced in different languages without having to make their search query (question) in each language.

For example, the CRM entity E50 Date – in a Default/English version – can be accessed as E50 Data in a Portuguese version or E50 Datum in a German version or E50 Date in a French version. Another example is the CRM entity E35 Title – as the Default – can be accessed as E35 Título in a Portuguese version, or E35 Titel in a German version, or E35 Titre in a French version.

Therefore, the CIDOC CRM can be translated into other languages in order to achieve the multilinguality, i.e. fosters a multilingual access to information as well as provides a higher information retrieval for the user's query.

REFERENCES

CEA, *et al*. Modelling Multilinguality in Ontologies. *Companion volume* – Posters and Demonstrations, pages 67–70. Manchester, August 2008 .Available at: <http://www.aclweb.org/anthology-new/C/C08/C08-2017.pdf>. Access: 20 out. 2012.

CIDOC CRM – Conceptual Reference Model. Available at: <http://www.cidoc-crm.org/>. Access: 12 jul. 2012.

Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - *A New Framework Strategy for Multilingualism.COM* (2005) 596 Brussels, 22.11.2005. (2005). Available at: <http://ec.europa.eu/education/languages/archive/doc/com596_en.pdf>. Access: 15 jul. 2012.

Crofts, N., Doerr, M., Gill, T., Stead, S. and Stiff, M.. *Definition of the CIDOC Conceptual Reference Model*, *version 4.0*. ICOM/CIDOC CRM Special Interest Group. 2004 . Available at: <http://zeus.ics.forth.gr/cidoc/docs/cidoc_crm_version_3.4.9.pdf>. Access: 20 set. 2012.

Doerr, M. The CIDOC CRM – an Ontological Approach to Semantic Interoperability of Metadata. *AIMagazine*, 24(3). , 2003. Available at: <http://www.ics.forth.gr/isl/publications/paperlink/Doerr_V2.pdf>. Access: 20 set. 2012.

[1] Peters, Carol. Guaranteeing Multilinguality in the Information Society. *Istituto di Elaborazione della Informazione. Consiglio Nazionale delle Ricerche.* Available at: <http://www.ercim.eu/medconf/papers/peters.html>. Access: 23 set. 2012

Doerr, M. Semantic interoperability: theoretical considerations. *Technical Report 345, ICS-FORTH*. , 2004. Available at: <http://www.ics.forth.gr/ftp/tech-reports/2004/2004.TR345_Semantic_Interoperability_Theoretical_Considerations.pfd>. Access: 20 set. 2012.

Gruber, T. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199{220, 1993.

Peters, Carol. Guaranteeing Multilinguality in the Information Society. *Istituto di Elaborazione della Informazione. Consiglio Nazionale delle Ricerche.* Available at: <http://www.ercim.eu/medconf/papers/peters.html>. Access: 23 set. 2012.

UNESCO's Recommendation on multilingualism and access to cyberspace. Available at: <http://unesdoc.unesco.org/images/0013/001329/132965e.pdf>. Access: 17 ago. 2012.