M.Sc.Thesis Proposal

Cristiana Araújo

Academic Year 2015/2016

Identification

Title A CIDOC-CRM Ontology for the Museum of the Person.

Student Cristiana Esteves Araújo(*pg27769@alunos.uminho.pt*).

Main Supervisor Professor Pedro Rangel Henriques, from Universidade do Minho/Dep. Informática/Centro Algoritmi;

Co-Supervisor Ricardo Martini, Ph.D Student at Universidade do Minho/Dep. Informática/Centro Algoritmi

Work Location Universidade do Minho, Departamento de Informática.

ECTS 45 ECTS.

Abstract

This document presents a proposal for a Master Thesis in Software Engineering, in the area of Virtual Museums, and Ontologies for knowledge representation and exploration.
The first objective of this thesis is the instantiation of a specific ontology for the document repository of the Museum of the Person (Museu da Pessoa), using a standard for museums, Comité Internacional pour la Documentation - Conceptual Reference Model (CIDOC-CRM). This abstract ontology will be populated with the interviews collected previously for a large project with four sub-projects. A triplestore will be used as database to store all the information that constitutes the Museum assets. At last the virtual museum web page will be created consulting the datastore through SPARQL Protocol and RDF Query Language (SPARQL) queries.

1 Context

The Museum of the Person (Museu da Pessoa[1], MP) was born in Brazil, São Paulo, created by a group of historians who decided to build the country's history using testimonials of ordinary people [SA03, Sta15].

The Museum of the Person aims to gather testimonials from every human being, famous or anonymous, to perpetuate his history [ARH+01, Sta15].

From the life stories of individuals, the objective is to write up the stories of families, communities, or institutions [Sta15]. This museum deals with common people, human beings, not with the physical objects usually composing the traditional museum assets. Thus, it can be said that the Person's Museum contains the intangible[2] cultural heritage assets (immaterial things). In this case, the alive objects are used as informers, reporting the events and emotions they experienced [ARH+01].

The narrators, to report their life stories, remember events and other particular situations they have participated in. These memories, narrated to the interviewer, will act as a basic element for social research, because the set of life stories allows to reconstruct a social universe [ARH+01].

Life stories are evidences in support of facts or statements attested by common people carrying a social and historical character, which must be preserved and processed to become an immeasurable human heritage [ARH+01].

MP resources are documental assets constituted by a collection of documents in eXtensible Markup Language (XML) format. So, there is the possibility of working them automatically through a parser (for instance, to clean them, to extract fragments or to exhibit them in web pages); moreover it is also possible to establish relations among the various documents [ARH+01].

All the arguments presented above motivate our interest on building a *virtual learning space* (in that case, a *virtual museum*) to tell to the world those life stories and to extract knowledge about an epoch and a society connecting and relating them.

On the other hand, if we intend to ingest information to preserve digitally it and to allow the extraction of knowledge from this document repository, it is crucial to research standard formats to archive and process semi-structured

[1] `http://www.museudapessoa.net`

[2] Available at: `http://www.unesco.org/culture/ich/index.php?lg=en&pg=00022#art2`

documents like those marked in XML.

For that purpose, ontologies are well suited and commonly used nowadays for knowledge representation. According to Borst [Bor97], an ontology is a formal and explicit specification of a shared conceptualization. It represents a consensual knowledge in a particular field.

Ontologies are important to represent knowledge. For this reason it was decided to use the Comité Internacional pour la Documentation - Conceptual Reference Model (CIDOC-CRM) as it is an ontology concerning the cultural heritage domain. [OL14, ICO13]. In practice, short versions of CIDOC-CRM are used to deal with specific domains; in order to keep for all them same way of storing and processing CRM-compatible form are used.

In this context, the first objective of this study consists in the deep analysis of the existing MP documental assets and in the instantiation of a specific ontology as a CRM-compatible form. After this, the next step will be the exploitation of the ontology, creating a Web site for exhibition of the stories and their relations.

2 Objectives

This master thesis objectives are the following:

- Creating a specific ontology for the document repository of the Museum of the Person, using a standard for museums (CIDOC-CRM);
- Populating the ontology with the interviews data, collected previously for a large project with four sub-projects, to verify if the ontology is appropriate for the document collection;
- Storing in a digital repository (for instance, a triple store) all the information that constitutes the Museum assets;
- Creating the virtual museum web pages to exhibit the collection of life stories, extracting the information from the stored ontology. This last one is the main objective.

Notice that the first and second items are more theoretical, and the two last ones are more technological.

3 Development Approach

To achieve the objectives identified in the Section 2, the methodology that will be followed in this master work is iterative and based on four components: theoretical research (bibliography revision); proposal of a solution for the highlighted problem; implementation of a prototype as a proof of study; and testing.
More specifically the work will be composed of the following steps:

- Analysis of the existing Document Type Definition (DTD);
- Research focusing the concept of Museum of the Person;
- Domain formalization using an ontology;
- Ontology description using a CRM-compatible form;
- Knowledge representation using triples in Resource Description Framework/ eXtensible Markup Language) [BE04, BE00] or similar notation;
- Triple storage in an adequate database;
- Web page design and creation using SPARQL queries[EH, JEN] to retrieve the information from the triple store.

4 Schedule

The duration of this master work is estimated to be one year. It seams acceptable to divide this project into six phases according to the following schedule:

Month 1**:** Literature review about Virtual Museums;

Month $2-3$**:** Literature review about Ontologies, CIDOC-CRM and company;

Month 4**:** Analysis and recovery of all the documentation existing in the Portuguese MP assets;

Month $5-6$**:** Application of CIDOC-CRM to the MP assets to build the desired ontology;

Month $7-10$**:** Web site development;

Month 11^{th} **to** 12^{th}**:** The last two months will be used to evaluate, test and draw conclusions relating the project outcomes.

The thesis report will be written in parallel with all the phases described. In the last phase, conclusions about the work done will be written, and the thesis document will be reviewed.

References

[ARH+01] José João Almeida, Jorge Gustavo Rocha, Pedro Rangel Henriques, Sónia Moreira, and Alberto Simões. Museu da Pessoa – arquitectura. In *Encontro Nacional da Associação de Bibliotecários, Arquivista e Documentalistas, ABAD'01*. BAD, 2001.

[BE00] Dan Brickley and R.V. Guha (Ed.s). Resource Description Framework (RDF) Schema specification. 2000.

[BE04] David Beckett and Brian McBride (Ed.s). Resource Description Framework (RDF/XML) Syntax Specification (Revised). `http://www.w3c.org/TR/REC-rdf-syntax`, 2004. Accessed: 2015-10-06.

[Bor97] Willem Nico Borst. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, 1997.

[EH] Endhe Elias and Olavo Holand. SPARQL: Linguagem de Consulta em Ontologias. `http://egov.ufsc.br/portal/sites/default/files/sparqlrevisado.pdf`. Accessed: 2015-09-30.

[ICO13] ICOM/CIDOC. Definition of the cidoc conceptual reference model. Technical report, ICOM/CIDOC, Oct 2013.

[JEN] APACHE JENA. Getting started with Apache Jena. `https://jena.apache.org/getting_started/index.html`. Accessed: 2015-09-30.

[OL14] Dominic Oldman and CRM Labs. The cidoc conceptual reference model (cidoc-crm): Primer. *International Council of Museums (ICOM)*, 1, July 2014.

[SA03] Alberto Simões and José João Almeida. Histórias de Vida + Processamento Estrutural = Museu da Pessoa. In *XATA 2003 — XML: Aplicações e Tecnologias Associadas*, page 16, Braga, Portugal, 2003. UM.

[Sta15] Philip B. Stafford. Museum of Person. `http://www.museumoftheperson.org`, 2015. Accessed: 2015-09-30.