Building the Museum of the Person

Based on a combined CIDOC-CRM/ FOAF/ DBpedia Ontology

Cristiana Esteves Araújo PG27769
MEI 15/16

Context

- The society is more and more concerned with the preservation and the dissemination of Cultural Heritage;
- This can be achieved resorting to the information and communication technologies that allow physical objects to become accessible to anyone and not deteriorated.
- Increasing the capability of information extraction, storage and visualization Museums have taken advantage to expand its field of action, as well as their own concept:
  - Expanding their geographical borders by providing information on the web to exhibit their collections;
  - Giving birth to completely virtual environments, called Virtual Museums.
Motivation

- Our main research challenge is concerned with the creation of Virtual Museums using directly the digital repositories (the documents or databases) that constitute the Museums’ assets.

- Apply this research to a concrete case study:
  - the Museum of the Person (MP) that contains citizen’s interviews that narrate their life stories.

Goals

- Create a specific ontology for the document repository of the Museum of the Person, based on:
  - CIDOC-CRM (Comité Internacional pour la Documentation - Conceptual Reference Model);
  - enriched with concepts and relations defined in FOAF (Friend-of-a-Friend) and Dbpedia;
  - populate the ontology with the interviews, to verify if the ontology is appropriate for the document collection.

- Store in a digital repository (a TripleStore and a Relational database) the information that constitutes the Museum assets
  - automatizing this process (data extraction from XML documents and ontology population);

- Create automatically the Virtual Museum Web pages to exhibit the collection of life stories, extracting the information from the stored ontology, using SPARQL queries.
Museum of the Person, an overview

- Museum of the Person was born in Brazil, São Paulo, in 1991
  - created by a group of historians who decided to build the country’s history using testimonials of ordinary people.

- Museum of the Person aims at gathering testimonials from every human being, famous or anonymous, to perpetuate his history.

- Persons are used as informers, reporting the events and emotions they experienced and other particular situations they have participated in;

- Its “art collection” is made up of intangible or immaterial things;

- These memories will act as a basic element for social research, because the set of life stories allows to rebuild a social universe.

Museum of the Person Documents

The Museum of the Person’s collection consists of XML documents connected to each participant:

- a mini-biography and personal data, such as name, date and place of birth, and job (BI);

- two versions of the interview:
  - the original interview refers to the raw interview; it contains all the questions asked and the narrator’s answers;
  - the edited document is a plain text, structured by themes that define small portions of a person’s life story;

- photographs and their legend:
  - The legend document contains a section for each photo or scanned document as file name and a caption;
  - This legend includes a description of the image and the date, and wherever possible the name of the stakeholder;

Additionally a Thesaurus that includes key concepts mentioned in the stories.
OntoMP: an ontology for MP

- The first step was the design on a specific **ontology** for the document repository of the Museum of the Person, and its specification based on
  - **CIDOC-CRM**
  - enriched with concepts and relations defined in **FOAF** and **DBpedia**.
- Then the ontology was populated with documents content from the collection.
A fragment of Maria Cacheira’s life

A person (E21 Person) has education Sabe ler e escrever (4ª classe) professes the religion Católica and has profession Peixeira e Empregada de limpeza.

This person (E21 Person) is depicted in the photo (E38 Image). This photo is identified by (P100 000-02-09) (E41 Appellation) has note (P3) Maria Alice Rodrigues Cacheira refers to (P67) Maria Alice Rodrigues Cacheira (E55 Type Description) and was taken in a (E52 Time Span) — that is identified by (P78) 2001-12-07 and at a (E53 Place) — that is identified by (P87) Junta de Freguesia da Afurada.
MP architecture

- We need an Information System
  - to extract the data from the stored ontology,
  - and build the Virtual Museum Web Pages to exhibit the collection of life stories.

- We propose a **generic architecture** and **two possible instances**

---

Proposal – General Approach
General Approach

Approach 1

- Supported on annotated documents and a direct representation of the Ontology (TriplesStore)

Approach 2

- Supported on a Relational Database and a mapping to the Ontology

Approach 1 - Module 1

[Diagram showing the process from Museum of the Person Repository to Ontology through TriplesStore]
<?xml version="1.0" encoding="ISO-8859-1"?>
<fotos>
  <foto ficheiro="090-F-01.jpg">
    <quem>Maria Alice Rodrigues Cacheira</quem>
    <onde>Junta de Freguesia da Afurada</onde>
    <quando data="2001-07-12"/>
  </foto>
</fotos>

Data Extraction and Storage
XML Photo

Data Extraction and Storage
RDF Photo
Translation System

XML -> RDF

Translation System

Print RDF
Approach 1 - Module 2

Generating the Learning Spaces (Exhibition Rooms)

RDF -> HTML
Querying and Displaying

npMP Project site
http://npmp.epl.uminho.pt
npMP Entrance Hall

http://npmp.epl.uminho.pt

npMP Life Stories Room (1 person)

http://npmp.epl.uminho.pt
Conclusion

The main contributions of this masters’ work are:

- The Documents collection of npMP (interviews, Thesaurus, DTD’s and photographies) was analyzed.

- An Ontology for the Museum of the Person, called OntoMP, was designed
  - OntoMP was described in a standard ontology format used for museums, CIDOC-CRM enriched with pertinent FOAF and DBpedia concepts and properties.

- A General Architecture was proposed to create the museum’s virtual exhibition rooms, as web pages, extracting information from the museum’s digital repository.
  - Two instantiation Approaches were designed, depending on the Data Storage (TripleStore or Relational Database).
  - Approach 1 was fully implemented

Conclusion

- The Ingestion Function [M1], to automatically extract data from the XML documents and generate the RDF triples, was implemented by a specially tailored text filter.

- The Generator [M2], to configure the Virtual Learning Spaces according to the exhibition requirements, was implemented:
  - using a Python script, complemented with HTML and CSS code.
  - reusing the SPARQL queries manually created, those needed for a concrete request are selected and sent to the SPARQL endpoint.
  - processing the returned data to be displayed in the [VLS].
Future Work

Some directions for future work are:

- Improve the visual interface of the npMP increasing the interactivity and exploiting geographical information applications following what is being done in Museu Marítimo de Sesimbra;
- Improve the system performance, resorting to other technologies or adapting the system’s architecture;
- Implement Approach 2;
- Combine the Ingestion Function [M1: XML2RDF] of Approach 1 with the Generator [M2: CaVaGen] of Approach 2;

Future Work (cont.):

- Investigate new standards and other notations for ontology description;
- Integrate and develop new tools to automatize the complete life cycle of npMP:
  - integrate in our platform a mobile application that was built to record the interviews;
  - create an application to transcribe audio interviews to text, and to markup the important concepts.
- Compare the approach developed by us with other approaches (either at the level of the museum specification, or at the level of the output produced), such as OdA (Objetos de Aprendizaje), a software system created at Universidad Complutense de Madrid (UCM) to build Virtual Museums, Repositories of Learning Objects, etc.;
- Apply all the knowledge acquired and implemented on Approach 1 to another case study, such as the Museu do Brinquedo Português or the Centro de Interpretação do Território de Ponte de Lima.