WP8 Case study: Cultural Heritage

Dana Dannélls, Aarne Ranta, Ramona Enache, Mariana Damova, Maria Mateva

University of Gothenburg and Ontotext

MOLTO
Final presentation, May 2013
Motivation and aim

As the amount of cultural data available on the Semantic Web (SW) is expanding, the demand of accessing this data in multiple languages is increasing.

If we want to enable multilingual retrieval and generation of cultural heritage (CH) content we must:
1. align multilingual metadata to interoperable knowledge bases;
2. add multilingual knowledge to this domain specific data.

Aim: To build an ontology-based application for communication of museum content on the Semantic Web and make it accessible in 15 languages.
Outline

- Museum Reason-able View: Interoperable cultural heritage knowledge bases
- Ontology-based multilingual grammar for retrieving and generating museum content:
  - RDF to NL – well-formed descriptions
  - NL to RDF – SPARQL queries linearization
- Cross-language retrieval and representation system using Semantic Web technology
The Museum Reason-able View (MRV)

Upper-level ontology: Proton

Domain ontology: CIDOC Conceptual Reference Model (CRM) v. 5.0.1

Application ontologies: Painting ontology and Museum Artifacts Ontology (MAO)

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Classes</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proton</td>
<td>542</td>
<td>183</td>
</tr>
<tr>
<td>CIDOC-CRM</td>
<td>87</td>
<td>130</td>
</tr>
<tr>
<td>Painting (SUMO, SOCH)</td>
<td>197</td>
<td>107</td>
</tr>
<tr>
<td>MAO</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>836</td>
<td>440</td>
</tr>
</tbody>
</table>
The museum Linked Open Data (LOD)

Gothenburg City Museum (GCM): Only two collections (GSM and GIM)

DBpedia: Large amount of painting instances

<table>
<thead>
<tr>
<th></th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCM</td>
<td>48</td>
</tr>
<tr>
<td>DBpedia</td>
<td>15,302</td>
</tr>
<tr>
<td>Total</td>
<td>15,350</td>
</tr>
</tbody>
</table>
The application grammar overview

Lexicon → Text Query Answer → Data

RGL → Text Query Answer → YAQL
Supported languages

Bulgarian, Catalan, Danish, Dutch, English, Finnish, French, Hebrew, Italian, German, Norwegian, Romanian, Russian, Spanish, Swedish.
The lexicon grammar

Covers a subset of the ontology classes and properties

Classes and properties were manually translated

Input:

\textbf{createdBy} (Ophelia, Brynolf Wennerberg)

\textbf{isA} (Brynolf Wennerberg, Painter).

\textbf{isA} (Ophelia, Painting)

Direct verbalization:

Ophelia is a painting. Ophelia was created by Brynolf Wennerberg. Brynolf Wennerberg is a painter.

Our approach:

Ophelia was painted by Brynolf Wennerberg.
The data grammar

Contains ontology instances that were extracted from GCM and DBpedia

No adequate translations from DBpedia

Instances of Museum are automatically translated from Wikipedia

<table>
<thead>
<tr>
<th>Class</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>662</td>
</tr>
<tr>
<td>Painter</td>
<td>116</td>
</tr>
<tr>
<td>Museum</td>
<td>104</td>
</tr>
<tr>
<td>Place</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>904</td>
</tr>
</tbody>
</table>
Automatic translation process overview
Successfully translated instances (out of 104)

<table>
<thead>
<tr>
<th>Language</th>
<th>Translated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian</td>
<td>26</td>
</tr>
<tr>
<td>Catalan</td>
<td>63</td>
</tr>
<tr>
<td>Danish</td>
<td>33</td>
</tr>
<tr>
<td>Dutch</td>
<td>81</td>
</tr>
<tr>
<td>Finnish</td>
<td>40</td>
</tr>
<tr>
<td>French</td>
<td>94</td>
</tr>
<tr>
<td>Hebrew</td>
<td>46</td>
</tr>
<tr>
<td>Italian</td>
<td>94</td>
</tr>
<tr>
<td>German</td>
<td>99</td>
</tr>
<tr>
<td>Norwegian</td>
<td>50</td>
</tr>
<tr>
<td>Romanian</td>
<td>27</td>
</tr>
<tr>
<td>Russian</td>
<td>87</td>
</tr>
<tr>
<td>Spanish</td>
<td>89</td>
</tr>
<tr>
<td>Swedish</td>
<td>58</td>
</tr>
</tbody>
</table>
Text grammar

Nine classes: Title, Painter, Type, Colour, Size, Year, Material, Museum, Place

Three are obligatory

   Interior was painted by Edgar Degas.

One function, three sentences

   Interior was painted on canvas by Edgar Degas in 1868. It measures 81 by 114 cm and it is painted in red and white. This painting is displayed at the Philadelphia Museum of Art.
Text grammar: RDF to NL
Word alignment example
Multilingual challenges

- Lexicalizations
  Classes: compounds, multiword expressions
  Properties: verbs, adverbs, prepositions

- Order of semantic elements
  Material, Year

- Tense and voice
  Past, past participle, present, active/passive

- Aggregation
  Conjunction, relative clause, punctuations

- Coreference
  Pronoun, noun, empty reference
Multilingual texts

Bul: Portrait of Aof de Wignacourt е нарисувана от Caravaggio през 1607 година. Тя е с размер 195 см на 134 см. Този експонат е изложен в Лувър.
Dut: Portrait of Aof de Wignacourt werd in 1607 door Caravaggio geschilderd. Het werk is 195 bij 134 cm. Dit schilderij wordt in Musée du Louvre getoond.
Eng: Portrait of Aof de Wignacourt was painted by Caravaggio in 1607. It measures 195 by 134 cm. This painting is displayed at the Musée du Louvre.
Fin: Maalauksen Portrait of Aof de Wignacourt on maalannut Caravaggio vuonna 1607. Se on kokoa 195 kertaa 134 cm. Tämä maalaus on esillä Louvressa.
Fre: Portrait de Aof de Wignacourt a été peint par Caravaggio en 1607. Il est de 195 sur 134 cm. Ce tableau est exposé au Musée du Louvre.
Heb: ציור של אוף דה ויגנкурט בצייר של קראבג'יו משנת 1607. הוא בגודל 195 על 134 סמ'. הציור משודרג במוזיאון הלובר.
Nor: Portrait av Aof de Wignacourt ble malt av Caravaggio i 1607. Det er 195 ganger 134 cm. Denne malerien er utstilt på Musée du Louvre.
Por: Portrait of Aof de Wignacourt é pintado de Caravaggio no 1607. Este dipinto é exposto no Museu do Louvre.
Rus: Portrait of Aof de Wignacourt нарисовано Caravaggio в 1607 году. Она с размером 195 см на 134 см. Эта картина видится в Лувр.
Spa: Portrait of Aof de Wignacourt está pintado por Caravaggio en 1607. Mide 195 por 134 cm. Esta pintura está expuesta en el Museo del Louvre.
Query grammar: NL to SPARQL
The SPARQL approach

Input:

```
SELECT distinct ?painting
WHERE ?painting rdf:type painting:OilPainting
```

Direct verbalization:
Select distinct Painting such that their type is OilPainting.

Our approach:
Show everything about all oil paintings.
Multilingual queries

Eng: show everything about all oil paintings at the Museu Calouste Gulbenkian

Swe: visa allting om alla oljemålningar i Gulbenkianmuseet

SPARQL: PREFIX painting:<http://spraakbanken.gu.se/rdf/owl/painting.owl#> $n
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> $n
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#> $n
WHERE $n {
  ?painting rdf:type painting:OilPainting ; $n
  rdfs:label ?title ; $n
  painting:hasMaterial ?material ; $n
  painting:hasCurrentLocation?museum ; $n
  painting:hasCreationDate ?date ; $n
  painting:hasDimension?dim . $n
  painting:lengthValue ?length ; $n
  painting:heightValue ?height . $museum rdfs:label ?loc . $n $n
  $n FILTER (regex(?loc, "Museu_Calouste_Gulbenkian", "i")) } LIMIT 200
http://museum.ontotext.com/
Grammar advantage/disadvantages

+ Modular grammar design
+ The lexicon and the data are shared by all grammars
+ 16 text patterns from one function
+ Natural realization of the ontology content
+ Takes 30 minutes to implement a new language if the language is in the RGL

- The texts can become artificial if instances are missing translations
- It may take two to five days to implement if the language is not in the RGL
Future plans

- Improve the lexicalization methodology
- Expand the amount of data and explore new classes and properties
- Increase multilingual coverage through semantic enrichment, e.g. place names from Geonames
- Collaboration with the National Museum in Stockholm, Swedish Open Cultural Heritage (SOCH) and with partners from Europeana
Dissemination

Dana Dannélls, Mariana Damova, Ramona Enache, Maria Mateva and Aarne Ranta (2013). Multilingual access to cultural heritage content on the Semantic Web. Submitted to the ACL workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH).


Dannélls Dana (2012). On generating coherent multilingual descriptions of museum objects from Semantic Web ontologies. Proceedings of the 7th International Natural Language Generation Conference (INLG), Utica, USA.

Dissemination


