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)

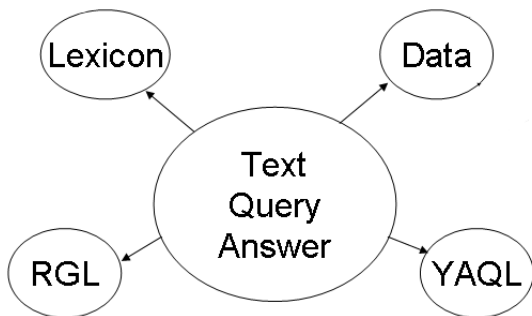| Ontology	Classes	Properties
Proton	542	183
CIDOC-CRM	87	130
Painting (SUMO, SOCH)	197	107
MAO	10	20
Total	836	440

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

DBpedia: Large amount of painting instances

	Instances
GCM	48
DBpedia	15,302
Total	15,350

The application grammar overview

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:
  createdBy (Ophelia, Brynolf Wennerberg)
  isA (Brynolf Wennerberg, **Painter**).
  isA (Ophelia, **Painting**)

Direct verbalization:
  Ophelia is a painting. Ophelia was created by Brynolf Wennerberg. Brynolf Wennerberg is a painter.

Our approach:
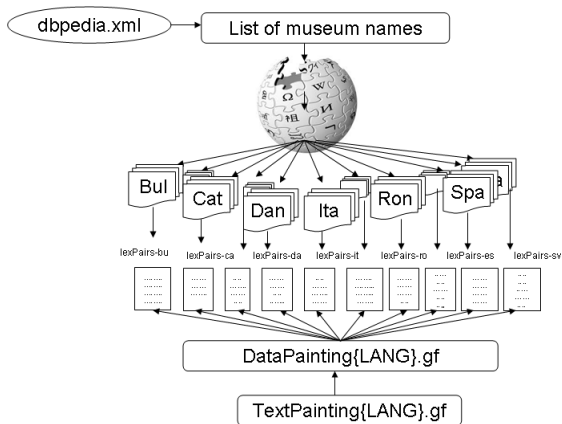 Ophelia was painted by Brynolf Wennerberg.

The data grammar

Contains ontology instances that were extracted from GCM and DBpedia

No adequate translations from DBpedia

Instances of *M*useum are automatically translated from Wikipedia

Class	Instances
Title	662
Painter	116
Museum	104
Place	22
Total	904

Successfully translated instances (out of 104)

Language	Translated
Bulgarian	26
Catalan	63
Danish	33
Dutch	81
Finnish	40
French	94
Hebrew	46
Italian	94
German	99
Norwegian	50
Romanian	27
Russian	87
Spanish	89
Swedish	58

Text grammar

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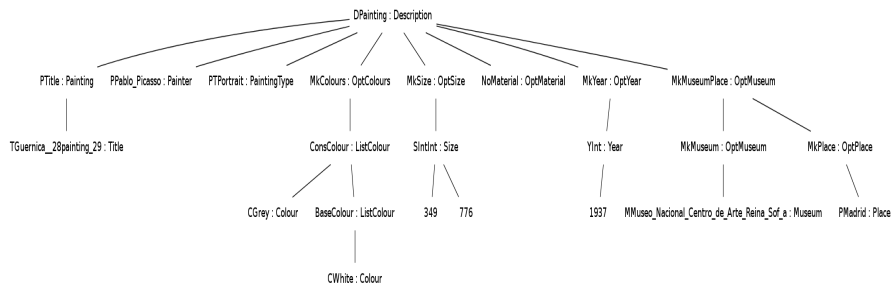
Three are obligatory

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

DPainting : Description

PTitle : Painting PPablo_Picasso : Painter PTPortrait : PaintingType MkColours : OptColours MkSize : OptSize NoMaterial : OptMaterial MkYear : OptYear MkMuseumPlace : OptMuseum

TGuernica__28painting_29 : Title

ConsColour : ListColour

SIntInt : Size

YInt : Year

MkMuseum : OptMuseum   MkPlace : OptPlace

CGrey : Colour   BaseColour : ListColour   349   776

1937   MMuseo_Nacional_Centro_de_Arte_Reina_Sof_a : Museum   PMadrid : Place

CWhite : Colour

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 Alof de Wignacourt е нарисувана от Caravaggio през 1607 година. Тя е с размер 195 см на 134 см. Този експонат е изложен в Лувър.

Cat: Portrait of Alof de Wignacourt està pintat per Caravaggio en 1607. Ell és de 195 sobre 134 cm. Aquesta pintura està exposada al Museu del Louvre.

Dan: Portrait of Alof de Wignacourt blev malet af Caravaggio i 1607. Det er 195 ganger 134 cm. Dette maleri er udstillet på Louvre.

Dut: Portrait of Alof 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 Alof 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 Alof de Wignacourt on maalannut Caravaggio vuonna 1607. Se on kokoa 195 kertaa 134 cm. Tämä maalaus on esillä Louvressa.

Fre: Portrait of Alof de Wignacourt a été peint par Caravaggio en 1607. Il est de 195 sur 134 cm. Ce tableau est exposé au Musée du Louvre.

Ger: Portrait of Alof de Wignacourt wurde in 1607 von Caravaggio gemalt. Das Werk ist 195 mal 134 cm. Dieses Bild ist ausgestellt in der Der Louvre.

Heb: ריוקן של אלוף דה וויגנקורט היא תמונה של קאראוואג'י משנת 1607. היא בגודל 195 על 134 סנטימטר. יצירה זו מוצגת במוזיאון הלובר.

Ita: Portrait of Alof de Wignacourt è dipinto da Caravaggio in 1607. Misura di 195 su 134 cm. Questo dipinto è esposto al Museo del Louvre.

Nor: Portrait of Alof de Wignacourt ble malt av Caravaggio i 1607. Det er 195 ganger 134 cm. Denne malerien er utstillt på Musée du Louvre.

Ron: Portrait of Alof de Wignacourt este pictat de cãtre Caravaggio în 1607. Este din 195 pe 134 cm. Acest tablou este expus în Musée du Louvre.

Rus: Portrait of Alof de Wignacourt нарисовалося Caravaggio в 1607 году. Она с размером 195 см на 134 см. Эта картина видится в Лувр.

Spa: Portrait of Alof de Wignacourt está pintado por Caravaggio en 1607. Mide 195 por 134 cm. Esta pintura está expuesta en el Museo del Louvre.

Swe: Portrait of Alof de Wignacourt målades av Caravaggio år 1607. Den är 195 gånger 134 cm. Den här målningen är utställd på Louvren.

Query grammar: NL to SPARQL
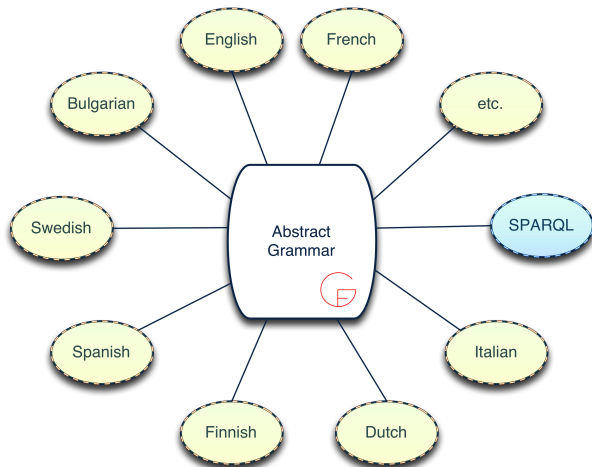
The SPARQL approach

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
SELECT distinct ?painting ?title ?material ?author ?year ?length ?height ?museum $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
?author rdfs:label ?painter . $n ?date painting:toTimePeriodValue ?year . ?dim
painting:lengthValue ?length; $n
painting:heightValue ?height . ?museum rdfs:label ?loc . $n $n
$n FILTER (regex(?loc, "Museu_Calouste_Gulbenkian" , "i")) } LIMIT 200
```

Demo

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)*.

Mariana Damova, Dana Dannélls, Ramona Enache, Maria Mateva and Aarne Ranta (2013). *Towards multilingual Semantic Web*, chapter. Natural Language Interaction with Semantic Web Knowledge Bases and LOD. Berlin: Springer.

Dana Dannélls, Aarne Ranta, Ramona Enache, Mariana Damova, and Maria Mateva (2013). *D8.3 Multilingual grammar for museum object descriptions.* Deliverable of EU Project MOLTO Multilingual Online Translation.

Dannélls Dana (2012). *Multilingual text generation from structured formal representations.* PhD thesis, Department of Swedish, University of Gothenburg. Sweden.

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.

Dannélls Dana, Enache Ramona, Mariana Damova, Milen Chechev (2012). Multilingual Online Generation from Semantic Web Ontologies. *Proceedings of the World Wide Web conference (WWW12),* Lyon, France. Dana Dannélls, Aarne Ranta, and Ramona Enache (2012). *D8.2 Multilingual grammar for museum object descriptions.* Deliverable of EU Project MOLTO Multilingual Online Translation.

Dissemination

Olga Caprotti, Aarne Ranta, Krasimir Angelov, Ramona Enache, John Camilleri, Grégoire Détrez Dana Dannélls, Thomas Hallgren, K.V.S. Prasad, and Shafqat Virk (2012). High-quality translation: Molto tools and applications. *In Proceedings of the fourth Swedish Language Technology Conference (SLTC).*

Damova, M., Kiryakov, A., Grinberg, M., Bergman, M.K., Giasson, F. and Simov, K. (2012). Creation and integration of reference ontologies for efficient LOD management. *In: Semi-Automatic Ontology Development: Processes and Resources, IGI Global*, Hershey PA, USA.

Dana Dannélls (2011). *D8.1 Ontology and corpus study of the cultural heritage domain.* Deliverable of EU Project MOLTO Multilingual Online Translation.

Dannélls Dana and Damova Mariana (2011). Reason-able View of Linked Data for Cultural Heritage. *Proceedings of the Third International Conference on Software, Services and Semantic Technologies (S3T),* Bourgas, Bulgaria.

Dannélls Dana, Enache Ramona, Damova Mariana and Chechev Milen (2011). A Framework for Improved Access to Museum Databases in the Semantic Web. *Proceedings of Language Technologies for Digital Humanities and Cultural Heritage. Workshop associated with the RANLP 2011 Conference*, Hissar, Bulgaria.

Dannélls Dana (2010). Discourse Generation from Formal Specifications Using the Grammatical Framework, GF. *Special issue of the journal Research in Computing Science (RCS)*, volume 46. pages 167–178.