

Multilingual access to cultural heritage content on the Semantic Web

Dana Dannélls and Aarne Ranta and Ramona Enache
University of Gothenburg and Chalmers University of Technology
SE-412 96 Gothenburg, Sweden

{dana.dannells, aarne.ranta, ramona.enache}@chalmers.se

Mariana Damova and Maria Mateva

Ontotext

Sofia 1784, Bulgaria

{mariana.damova, maria.mateva}@ontotext.com

Abstract

As the amount of cultural data available on the Semantic Web is expanding, the demand of accessing this data in multiple languages is increasing. Previous work on multilingual access to cultural heritage information has shown that at least two different problems must be dealt with when mapping from ontologies to natural language: (1) mapping multilingual metadata to interoperable knowledge sources; (2) assigning multilingual knowledge to cultural data. This paper presents our effort to deal with these problems. We describe our experiences with processing museum data extracted from two distinct sources, harmonizing this data and making its content accessible in natural language. We extend prior work in two ways. First, we present a grammar-based system that is designed to generate coherent texts from Semantic Web ontologies in 15 languages. Second, we describe how this multilingual system is exploited to form queries using the standard query language SPARQL. The generation and retrieval system builds on W3C standards and is available for further research.

1 Introduction

As the amount of cultural data available on the Semantic Web is expanding (Dekkers et al., 2009; Brugman et al., 2008), the demand of accessing this data in multiple languages is increasing (Stiller and Olensky, 2012).

There have been several applications that applied Natural Language Generation (NLG)

technologies to allow multilingual access to Semantic Web ontologies (Androutsopoulos et al., 2001; O'Donnell et al., 2001; Androutsopoulos and Karkaletsis, 2005; Androutsopoulos and Karkaletsis, 2007; Davies, 2009; Bouayad-Agha et al., 2012). The above authors have shown it is necessary to have an extensive lexical and syntactic knowledge when generating multilingual natural language from Semantic Web ontologies. However, because previous applications are mainly concerned with two or three languages, it is still not clear how to minimize the efforts in assigning lexical and syntactic knowledge for the purpose of enhancing automatic generation of adequate descriptions in multiple languages.

This paper presents our work on making Cultural Heritage (CH) content available on the Semantic Web and accessible in 15 languages using the Grammatical Framework, GF (Ranta, 2011). The objective of our work is both to form queries and to retrieve semantic content in multiple languages. We describe our experiences with processing museum data extracted from two different sources, harmonizing this data and making its content accessible in natural language (NL). The generation and retrieval system builds on the World Wide Web Consortium (W3C) standards and is available for further research.¹

The remainder of this paper is structured as followed. We present the related work in Section 2. We describe the underlying tech-

¹The generation and retrieval system is available online: <http://museum.ontotext.com/>

nology in Section 3. We provide a detailed description of the data and present the approach taken to make this data accessible in the Linked Open Data (LOD) in Section 4. We outline the multilingual approach and discuss the challenges we faced in Section 5. We discuss the results in Section 6. We end with some conclusions and pointers to future work in Section 7.

2 Related work

Lately there has been a lot of interest in enabling multilingual access to cultural heritage content that is available on the Semantic Web. Androutsopoulos et al. (2001) and O'Donnell et al. (2001) have shown that accessing ontology content in multiple languages requires extensive linguistic data associated with the ontology classes and properties. However, they did not attempt to generate descriptions in real time from a large set of ontologies.

Similar to Bouayad-Agha et al. (2012), our system relies on a multi-layered ontology approach for generating multilingual descriptions. In contrast to Dekkers et al. (2009) and Brugman et al. (2008) whose systems make use of Google translation services, which are data driven, our system is grammar driven.

Moreover, we present a multilingual grammar-based approach to SPARQL (SPARQL Protocol and RDF Query Language) (Garlik and Andy, 2013). The method differs from the verbalization methods presented by Ngonga Ngomo et al. (2013) and Ell et al. (2012) in that it realizes the ontology content rather than the ontology axioms. Thus providing a more natural realization of the query language.

3 The technological infrastructure

Although the architecture of the Semantic Web and Linked Open Data provides access to distributed data sets,² many of the resources available in these sets are not accessible because of cross-language meta-data. To overcome this limitation, the knowledge representation infrastructure adopted in our approach is designed as a Reason-able View of

²<http://linkeddata.org>

the Web of Data. The Reason-able View is a compound dataset composed of several Resource Description Frameworks (RDFs). To query such a compound dataset, the user has to be intimately familiar with the schemata of each single composing dataset. That is why the Reason-able View approach is extended with the so called ontological reference layer, which introduces a unification ontology, mapped to the schemata of all single datasets from a given Reason-able View and thus provides a mechanism for efficient access and navigation of the data.

3.1 Museum Reason-able View (MRV)

The Museum Reason-able View is an assembly of cultural heritage dominated RDF datasets (Dannélls et al., 2011). It is loaded into OWLIM-SE (Bishop et al., 2011) with inference performed on the data with respect to OWL Horst (ter Horst, 2005).

3.2 The ontological reference layer

The Museum Reason-able View gathers: (a) datasets from LOD, including DBpedia;³ (b) the unification ontology PROTON,⁴ an upper-level ontology, consisting of 542 classes and 183 properties; (c) two cultural heritage specific ontologies: (i) CIDOC-CRM (Crofts et al., 2008),⁵ consisting of 90 classes and 148 properties; (ii) Museum Artifacts Ontology (MAO),⁶ developed for mapping between museum data and the K-samsök schema.⁷ It has 10 classes and 20 properties; (d) the Painting ontology,⁸ an application ontology developed to cover detailed information about painting objects in the framework

³DBpedia, structured information from Wikipedia: <http://dbpedia.org>.

⁴<http://www.ontotext.com/proton-ontology>

⁵<http://www.cidoc-crm.org/>

⁶It is just a coincidence that this ontology has the same name as the Finnish MAO (Hyvonen et al., 2008), which also describes museum artifacts for the Finnish museums.

⁷K-samsök <http://www.ksamsok.se/in-english/>), the Swedish Open Cultural Heritage (SOCH), provides a Web service for applications to retrieve data from cultural heritage institutions or associations with cultural heritage information.

⁸<http://spraakdata.gu.se/svedd/painting-ontology/painting.owl>

of the Semantic Web. It contains 197 classes and 107 properties of which 24 classes are equivalent to classes from the CIDOC-CRM and 17 properties are sub-properties of the CIDOC-CRM properties.

3.3 Grammatical Framework (GF)

The Grammatical Framework (GF) (Ranta, 2004) is a grammar formalism targeted towards parsing and generation. The key feature of GF is the distinction between an abstract syntax, representing the domain, and concrete syntaxes, representing linearizations in various target languages, natural or formal.

GF comes with a resource grammar library (RGL) (Ranta, 2009) which aids the development of new grammars for specific domains by providing syntactic operations for basic grammatical constructions (Ranta, 2011). More than 30 languages are available in the RGL. Our application targets 15 of those, including: Bulgarian, Catalan, Danish, Dutch, English, Finnish, French, Hebrew, Italian, German, Norwegian, Romanian, Russian, Spanish, and Swedish.

4 Cultural heritage data

The data we have been experimenting with to enable multilingual descriptions of museum objects and answering to queries over them is a subset of the Gothenburg City Museum (GCM) database,⁹ and a subset of the DBpedia dataset. Because these two datasets are very different in size and nature, the pre-processing of each set differs substantially. In the following we describe each of the sets and the pre-processing steps in more details.

4.1 Gothenburg City Museum (GCM)

The set from the GCM contains 48 painting records. Its content, both the metadata and data that were originally in Swedish, were translated to English. An example of a record from GCM is shown in Table 1.

4.2 DBpedia

The set from DBpedia contains 662 painting records, the data covers at least 5 languages,

⁹<http://stadsmuseum.goteborg.se/wps/portal/stadsm/english>

Record field	Value
Field nr.	4063
Prefix	GIM
Object nr.	8364
Search word	painting
Class nr	353532
Classification	Gothenburg portrait
Amount	1
Producer	E.Glud
Produced year	1984
Length cm	106
Width cm	78
Description	oil painting represents a studio indoors
History	Up to 1986 belonged to Datema AB, Flöjelbergsg 8, Gbg
Material	oil paint
Current keeper	2
Location	Polstjärnegatan 4
Package nr.	299
Registration	19930831
Signature	BI
Search field	Bilder:TAVLOR PICT:GIM

Table 1: A painting object representation from the GCM database.

the metadata is in English. An example of a record from DBpedia is shown in Table 2.

4.3 Transition of data to the MRV

Making the museum data available through the knowledge infrastructure required translations of the record fields and values, and mapping to a unified ontology. This process also required pre-processing of the free text fields such as *Description* and *History* (see Table 1) to enrich the data content.

To make the DBpedia data accessible through the knowledge infrastructure, it required some preprocessing, cleaning, and mapping to the Painting ontology for data consistency. This unification was needed to use a consistent SPARQL queries from where NL descriptions could be generated.

Firstly, we attempted to clean data noise and results that would make a single painting reappear in the query results. Then, we transformed year and size strings into only numbers. For each painter, museum and painting literal we had a single representation in the data. All names were normalized, for example, Salvador Dalí was converted

```

<result>
<binding name='painting'>
<uri>http://dbpedia.org/resource/
Virgin_of_the_Rocks</uri> </binding>
<binding name='museum'>
<literal xml:lang='en'>Musée du Louvre
</literal> </binding>
<binding name='author'>
<literal xml:lang='en'>da Vinci, Leonardo
</literal> </binding>
<binding name='height'>
<literal datatype=
'http://www.w3.org/2001/XMLSchema#int'>
190</literal> </binding>
<binding name='width'>
<literal datatype=
'http://www.w3.org/2001/XMLSchema#int'>
120</literal> mateva </binding>
<binding name='title'>
<literal xml:lang='en'>London version
</literal> </binding>
<binding name='type'>
<literal xml:lang='fr'>Huile sur panneau
</literal> </binding>
<binding name='year'>
<literal datatype=
'http://www.w3.org/2001/XMLSchema#int'>
1495</literal> </binding> </result>

```

Table 2: A painting object representation from DBpedia

to *Salvador Dal.*. For different Uniform Resource Identifiers (URIs) pointing to the same painting, we used the OWL (W3C, 2012) construct *owl:sameAs*. With this construct we were able to keep the data linked in the other graphs in the LOD cloud.

5 Multilingual linked data

Our application is targeted towards lay users who wish to formulate queries and retrieve information in any language. Such users do not have any knowledge about ontologies or semantic data processing. For us it was therefore necessary to enable interactions in a simple use.

The work towards making Semantic Web data accessible to different users required lexicalizations of ontology classes, properties and individuals (literal strings associated with a certain class).

Following the GF mechanism, lexicaliza-

tions is accomplished through linearizations of functions. Linearization of functions varies depending on the language.

5.1 Lexicalizations of classes and properties

Most of the ontology classes defined in our grammar are linearized with noun phrases in the concrete syntaxes. These were translated manually by a native speaker of the language. Examples from four languages are shown below. In the examples we find the following RGL constructions: *mkCN* (Common noun) and *mkN* (Noun).

```

Class: Painting
Swe. mkCN (mkN "målning");
Fre. mkCN (mkN "tableau");
Fin. mkCN (mkN "maalaus");
Ger. mkCN mkN "Bild"
      "Bilder" neuter;

Class: Portrait
Swe. mkCN (regGenN "porträtt"
           neutrum);
Fre. mkCN (mkN "portrait");
Fin. mkCN (mkN "muoto"
           (mkN "kuva"));
Ger. mkCN (mkN "Porträt"
           "Porträts" neuter);

```

Two of the ontology classes that are not linearized with a noun phrase are: *Year* and *Size*. These are linearized with prepositional phrases in which the preposition is language dependent. Below are some examples which show how the *Year* string, i.e. *YInt* function, is realized in six languages. In the examples we find the following RGL constructions: *mkAdv* (Verb Phrase modifying adverb), *Prep* (Preposition) and *symp* (Symbolic).

```

Bul. YInt i = mkAdv prez_Prep
      (symp (i.s ++ year_Str));
Fin. YInt i = mkAdv (prePrep
                    nominative "vuonna") (symp i);
Fre. YInt i = mkAdv en_Prep (symp i);
Ger. YInt i = mkAdv in_Prep (symp i);
Swe. YInt i = mkAdv noPrep
      (symp ("år" ++ i.s));
Rus. YInt i = mkAdv in_Prep
      (symp (i.s ++ godu_Str));

```

The ontology properties are defined with operations in the concrete syntaxes. Because

Table 3: The amount of lexicalized literals in a subset of the MRV

Class	literals
Title	662
Painter	116
Museum	104
Place	22

an ontology property is linearized differently depending on how it is realized in the target language, these operations are of type: verbs (e.g. *paint_V2*), adverbs (e.g. *painted_A*) and prepositions (e.g. *Prep*). Examples from three languages are shown below.

```
Swe. paint_V2 : V2 = mkV2 "måla";
    painted_A : A = mkA "målad";
    at_Prep = mkPrep "på" ;
Fin. paint_V2 = mkV2 "maalata";
    painted_A = mkA "maalattu";
Ger. paint_V2 : V2 = mkV2
    (mkV "malen");
    painted_A : A = mkA "gemalt";
    at_Prep = in_Prep ;
```

The above functions correspond to three ontological properties, namely *painting by*, *painting* and *created in*. This approach to ontology lexicalization permits variations regarding the lexical units the ontology properties should be mapped to. It allows to make principled choices about the different realizations of an ontology property.

5.2 Lexicalizations of literals

The part of the MRV to which we provide translations for consists of 906 individuals, their distribution across four classes is provided in Table 3. The lexical units assigned to painting titles, painters and museum literals are by default the original strings as they appear in the data. The majority of strings are given in English. However, because without translations of the name entities the results can become artificial and for some languages ungrammatical, we run a script that translates museum literals from Wikipedia automatically.

Automatic translation was done by: (1) curling for Web pages for a museum string; (2) extracting the retrieved trans-

Table 4: The number of automatically translated museum names from Wikipedia

Language	Translated names
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

lated entry for each string; (3) reducing the retrieved list by removing duplicated and ambiguous entries. This process was repeated for each language.

As a result of the translation process, a list of lexical pairs was created for each language. Museum literals were then linearized automatically by consulting the created list for each language. In the cases where no translation was found, the original string, as it appears in the dataset was used.

Unfortunately, the amount of the translated museum names was not equal for all languages. The distribution of the translated names is given in Table 4. Below follow some examples of how museum names are represented in the grammar:

```
Swe. MGothenburg_City_Museum =
    mkMuseum "Göteborgs stadsmuseum";
    MMus_e_du_Louvre =
        mkMuseum "Louvren";
Ita. MGothenburg_City_Museum =
    mkMuseum
    "museo municipale di Goteburgo";
    MMus_e_du_Louvre =
        mkMuseum "Museo del Louvre";
Fre. MGothenburg_City_Museum =
    mkMuseum
    "musée municipal de Göteborg";
    MMus_e_du_Louvre =
        mkMuseum "Musée du Louvre";
Cat. MGothenburg_City_Museum =
    mkMuseum "Gothenburg_City_Museum";
    MMus_e_du_Louvre =
```

```

mkMuseum "Museu del Louvre";
Ger. MGothenburg_City_Museum =
mkMuseum "Gothenburg_City_Museum";
MMus_e_du_Louvre =
mkMuseum "Der Louvre ";

```

Where the construct *mkMuseum* has been defined to build a noun phrase from a given string. A special case of *mkMuseum* appears in four languages: Italian, Catalan, Spanish and French, where a masculine gender is assigned to the museum string to get the correct inflection form of the noun.

5.3 Realization of sentences

To generate sentences from a set of classes we had to make different judgements about how to order the different classes. Below we provide an example of a sentence linearization from four languages. The sentence comprises four semantic classes: *Painting*, *Material*, *Painter* and *Year*. In the examples we find following RGL constructors: *mkText* (Text), *mkS* (Sentence), *mkCl* (Clause), *mkNP* (Noun Phrase), and *mkVP* (Verb Phrase).

```

Ita. s1 : Text = mkText (mkS
(mkCl painting (mkVP (mkVP (mkVP
(mkVP dipinto_A) material.s)
(SyntaxIta.mkAdv by8agent_Prep
(title painter.long))) year.s))) ;
Fre. s1 : Text = mkText
(mkS anteriorAnt
(mkCl painting (mkVP (mkVP (mkVP
(passiveVP paint_V2) material.s)
(SyntaxFre.mkAdv by8agent_Prep
(title painter.long))) year.s))) ;
Ger. s1 : Text = mkText
(mkS pastTense
(mkCl painting (mkVP (mkVP
(mkVP (passiveVP paint_V2) year.s)
(SyntaxGer.mkAdv von_Prep
(title painter.long))) material.s))) ;
Rus. s1 : Text = mkText
(mkS pastTense
(mkCl painting (mkVP (mkVP (mkVP
(passiveVP paint_V2)
(SyntaxRus.mkAdv part_Prep
(title painter.long
masculine animate)))
material.s) year.s))) ;

```

Some of the distinguishing differences between the languages are: in Finnish the use of an active voice, in Italian, present tense, in French, past participle, in Spanish, present

simple. The order of the categories is also different. In German the material string appears at the end of the sentence as opposed to the other languages where year is often the last string.

5.4 Realizations of texts

The text grammar has been designed to generate a coherent natural language descriptions from a selected set of the returned triples. More specifically, our grammar covers eight concepts that are most commonly used to describe a painting, including: Title, Painter, Painting type, Material, Colour, Year, Museum and Size. In the grammar module called *TextPainting* they are defined as categories and are captured in one function *DPainting* which has the following representation in the abstract syntax.

```

DPainting :
Painting -> Painter ->
PaintingType -> OptColours ->
OptSize -> OptMaterial ->
OptYear -> OptMuseum -> Description;

```

In the function *DPainting* five arguments have been implemented as optional, i.e. *OptColour*, *OptSize*, *OptMaterial*, *OptYear* and *OptMuseum*. Each of these categories can be left out in a text.

In the current implementation we limited the length of a description to three sentences. A minimal description consists of only one sentences. Below follow some examples of texts generated in English to exemplify the different descriptions we are able to generate from one single function call with a varying number of instantiated parameters.

- 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.
- Interior was painted by Edgar Degas in 1868. It measures 81 by 114 cm. This painting is displayed at the Philadelphia Museum of Art.
- Interior was painted on canvas by Edgar Degas in 1868. It is painted in red and white. This painting is displayed at the Philadelphia Museum of Art.

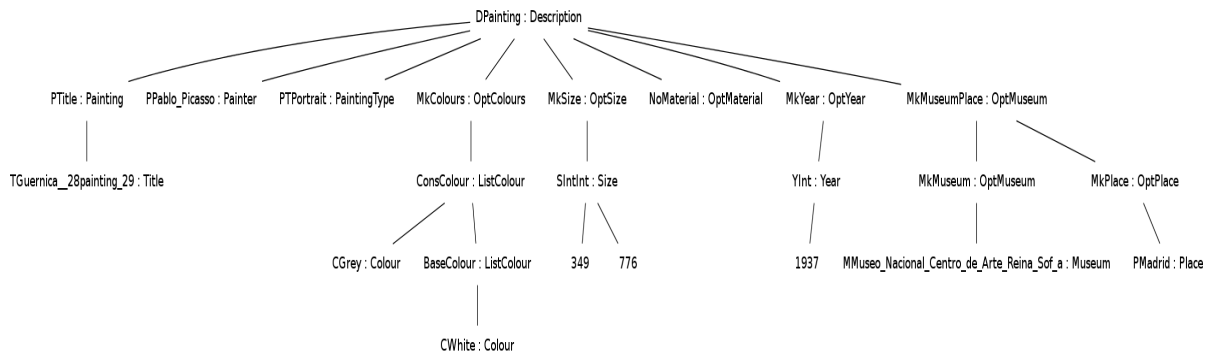


Figure 1: A semantic tree realization of nine ontology classes

- Interior was painted by Edgar Degas. It measures 81 by 114 cm and it is painted in red and white. This painting is displayed at the Philadelphia Museum of Art.
- Interior was painted on canvas by Edgar Degas. It measures 81 by 114 cm and it is painted in red and white.
- Interior was painted by Edgar Degas in 1868. This painting is displayed at the Philadelphia Museum of Art.
- Interior was painted by Edgar Degas.

5.5 Multilingual querying

Semantic Web technologies offer the technological backbone to meet the requirement of integrating heterogeneous data easily, but they are still more adapted to be consumed by computers than by humans. As a consequence, to retrieve semantic content from the knowledge base the user must: 1. master SPARQL, the query language for RDF; 2. have knowledge about each integrated dataset in the knowledge base.

Ngonga Ngomo et al. (2013) have shown that realizations of SPARQL queries in natural language enhance the user understanding of the formulated queries and the retrieved results.

We have implemented an extra SPARQL module that allow us to map from any of the 15 supported languages to SPARQL and from SPARQL to any of the 15 supported languages. The grammar reuses a more generic query module that allows to

form both domain specific and domain independent queries. Some examples of the queries that can be formulated with the multilingual grammar and transformed to SPARQL are:

1. Some X
2. All About X
3. Show everything about X
4. All X painted by Y
5. Some X painted on Y
6. What is the material of X
7. Show everything about all X that are painted on Y

In GF, realization of SPARQL queries is done by introducing new parameters, for example:

```
QPainter p = {
  wh1 = "?author";
  prop = p ;
  wh2 = "painting:createdBy ?painter .
?painter rdfs:label ?author ."} ;
```

The function *QPainter* defined to formulate a query such as *who painted Mona Lisa?* has been added two additional parameters, i.e. *wh1* and *wh2*. With these parameters it is possible to formulate SPARQL queries such as the one below.

```
SELECT ?author
WHERE {
  ?painting rdf:type
    painting:Painting ;
  painting:createdBy ?painter ;
  rdfs:label ?title
  FILTER (str(?title)="Mona_Lisa") .
  ?painter rdfs:label ?author.
}
```

TextPainting: DPainting (PTitle Interior_28Degas_29) PEdgar_Degas PTPainting NoColours (MkSize (SIntInt 50 103)) NoMaterial (MkYear (YInt 1890))
TextPaintingBul: Interior e нарисована от Edgar Degas през 1890 година. Тя е с размер 50 см на 103 см. Този експонат е изложен в Лувър.
TextPaintingCat: Interior fou pintat per Edgar Degas en 1890. Mesura 50 per 103 cm. Aquesta pintura està exposada al Museu del Louvre.
TextPaintingDan: Interior blev malet af Edgar Degas i 1890. Det er 50 ganger 103 cm. Dette maleri er udstillet på Louvre.
TextPaintingDut: Interior werd in 1890 door Edgar Degas geschilderd. Het werk is 50 bij 103 cm. Dit schilderij wordt in Musée du Louvre getoond.
TextPaintingEng: Interior was painted by Edgar Degas in 1890. It measures 50 by 103 cm. This painting is displayed at the Musée du Louvre.
TextPaintingFin: maalaus Interior on maalannut Edgar Degas vuonna 1890. Se on kokoa 50 kertaa 103 cm. Tämä maalaus on esillä Louvressa.
TextPaintingFre: Interior a été peint par Edgar Degas en 1890. Il est de 50 sur 103 cm. Ce tableau est exposé au Musée du Louvre.
TextPaintingGer: Interior wurde in 1890 von Edgar Degas gemalt. Das Werk ist 50 mal 103 cm. Dieses Bild ist ausgestellt im Der Louvre .
TextPaintingHeb: רבולה נא יזמב תבצמו וז הרצי י רתג טס 103 לע 50 לד ובג אה 1890 תנשנ סנד לש רוצ אה מינפ
TextPaintingIta: Interior è dipinto da Edgar Degas in 1890. Misura di 50 su 103 cm. Questo dipinto è esposto al Museo del Louvre.
TextPaintingNor: Interior ble malt av Edgar Degas i 1890. Det er 50 ganger 103 cm. Denne malerien er utstilt på Musée du Louvre.
TextPaintingRon: Interior este pictat de către Edgar Degas în 1890. Este din 50 pe 103 cm. Acest tablou este expus în Musée du Louvre.
TextPaintingRus: Interior нарисовался Edgar Degas в 1890 году. Она с размером 50 см на 103 см. Эта картина видится в Лувр.
TextPaintingSpa: Interior fue pintado por Edgar Degas en 1890. Mide 50 por 103 cm. Esta pintura está expuesta en el Museo del Louvre.
TextPaintingSwe: Interior målades av Edgar Degas år 1890. Den är 50 gånger 103 cm. Den här målningen är utställd på Louvren.

Figure 2: Multilingual generation results

5.6 Multilingual text generation

Our approach allows different texts to be generated depending on the information that is available in the ontology. A minimal description consists of three classes: a title, a painter and a painting type. A complete description consists of nine classes, as illustrated in Figure 1. With only one function *DPainting* our system is able to generate 16 different text variants. Figure 2 illustrates a generation results in 15 languages.

6 Discussion

The majority of the challenges in the production of the CH data pool stemmed from the very nature of the Linked Open Data. The data in the LOD cloud are notoriously noisy and inconsistent.

The multilingual labels from the FactForge datasets and more precisely from DBpedia, are not always available in all the supported languages. Although DBpedia in its large pool of data provides access to multilingual content, it is inconsistent. Many of the entries it contains are missing translations. There is a mixture of numeric and string literals. There are many duplications, most of them occur because the same ID appears in different languages. The content of the data is verbose, for example place-names and museum-names are represented with one string, for example: “Rijksmuseum, Amsterdam”, instead of two different strings linked by two separate concepts, i.e. *Museum* and *Place*. This kind of inconsistent data representation had an impact on the translation of museum names.

Another problem was that not all art objects are uniformly described with the same set of characteristics. For instance, some paintings were missing a title or a painter name. Because we constructed the grammar in such a way that disallows absence of this information, we had to replace titles with id numbers and empty painter names with the string *unknown*. Moreover, the data contained many duplications. This occurred because some of the property assertions were presented with different strings and triggered many RDF triples.

We also faced many linguistic challenges on different levels. Lexicalizations of ontology classes and properties regarding use of compounds, variations of verbs, adverbs and prepositions. On sentence level, order of classes, variations of tense and voice. On both sentence and discourse levels, aggregation variations and use of coreference elements.

7 Conclusions

We presented an ontology-based multilingual application developed in the Grammatical Framework and a cross-language retrieval system that uses this application for generating museum object descriptions in the Semantic Web.

The generation and retrieval system builds on W3C standards. It covers semantic data from the Gothenburg City Museum database and DBpedia. The grammar enables descriptions of paintings and answering to queries over them, covering 15 languages for baseline functionality.

Acknowledgment

This research has been supported by MOLTO, the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-ICT-247914.

References

- S. Kallonis Androutsopoulos and V. Karkaletsis. 2005. Exploiting OWL ontologies in the multilingual generation of object descriptions. In *The 10th European Workshop on NLG*, pages 150–155, Aberdeen, UK.
- J. Oberlander Androutsopoulos and V. Karkaletsis. 2007. Source authoring for multilingual generation of personalised object descriptions. *Natural Language Engineering*, 13(3):191–233.
- Ion Androutsopoulos, Vassiliki Kokkinaki, Aggeliki Dimitromanolaki, Jo Calder, Jon Oberl, and Elena Not. 2001. Generating multilingual personalized descriptions of museum exhibits: the M-PIRO project. In *Proceedings of the International Conference on Computer Applications and Quantitative Methods in Archaeology*.
- B. Bishop, A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, and R. Velkov. 2011. OWLIM: A family of scalable semantic repositories. *Semantic Web Journal, Special Issue: Real-World Applications of OWL*.
- Nadjet Bouayad-Agha, Gerard Casamayor, Simon Mille, Marco Rospocher, Horacio Saggion, Luciano Serafini, and Leo Wanner. 2012. From Ontology to NL: Generation of multilingual user-oriented environmental reports. *Lecture Notes in Computer Science*, 7337.
- Hennie Brugman, Véronique Malaisé, and Laura Hollink. 2008. A common multimedia annotation framework for cross linking cultural heritage digital collections. In *International Conference on Language Resources and Evaluation*.
- Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff, 2008. *Definition of the CIDOC Conceptual Reference Model*.
- Dana Dannélls, Mariana Damova, Ramona Enache, and Milen Chehev. 2011. A Framework for Improved Access to Museum Databases in the Semantic Web. In *Recent Advances in Natural Language Processing (RANLP). Language Technologies for Digital Humanities and Cultural Heritage (LaTeCH)*.
- Rob Davies. 2009. EuropeanaLocal – its role in improving access to Europe's cultural heritage through the European digital library. In *Proceedings of IACH workshop at ECDL2009 (European Conference on Digital Libraries)*, Aarhus, September.
- Makx Dekkers, Stefan Gradmann, and Carlo Meghini. 2009. Europeana outline functional specification for development of an operational european digital library. Technical report. Europeana Thematic Network Deliverables 2.5. Contributors and peer reviewers: Europeana.net WP2 Working Group members, Europeana office.
- Basil Ell, Denny Vrandečić, and Elena Simperl. 2012. SPARTIQUATION – Verbalizing SPARQL queries. In *Proceedings of ILD Workshop, ESWC 2012*.
- Steve Harris Garlik and Seaborne Andy, 2013. *SPARQL 1.1 Query Language*, March. <http://www.w3.org/TR/sparql11-query/>.
- E Hyvonen, E. Maekelae, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula. 2008. Museum finland. In *Finnish Museum on the Semantic Web*.
- Axel-Cyrille Ngonga Ngomo, Lorenz Bühmann, Christina Unger, Jens Lehmann, and Daniel Gerber. 2013. Sorry, i don't speak sparql: translating sparql queries into natural language. In *Proceedings of the 22nd international conference on World Wide Web, WWW '13*, pages 977–988, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Michael J. O'Donnell, Chris Mellish, Jon Oberlander, and Alistair Knott. 2001. ILEX: An architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3):225–250.
- Aarne Ranta. 2004. Grammatical Framework, a type-theoretical grammar formalism. *Journal of Functional Programming*, 14(2):145–189.
- Aarne Ranta. 2009. The GF resource grammar library. *The on-line journal Linguistics in Language Technology (LiLT)*, 2(2). <http://elanguage.net/journals/index.php/lilt/article/viewFile/214/158>.
- Aarne Ranta. 2011. *Grammatical Framework: Programming with Multilingual Grammars*. CSLI Publications, Stanford. ISBN-10: 1-57586-626-9 (Paper), 1-57586-627-7 (Cloth).
- J. Stiller and M. Olensky. 2012. Europeana: A multilingual trailblazer. In *The W3C Workshop: The Multilingual Web - Linked Open Data and Multilingual Web-LT Requirements*, Dublin.
- H. J. ter Horst. 2005. Combining RDF and Part of OWL with Rules: Semantics, Decidability, Complexity. In *Proceedings of The Semantic Web ISWC*, volume 3729 of LNCS, pages 668–684, Heidelberg. Springer Berlin.
- W3C, 2012. *OWL Web Ontology Language Overview*, December. <http://www.w3.org/TR/owl2-overview/>.