

Multilingual access to cultural heritage content on the Semantic Web

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 mapping from ontologies to natural language requires at least two different steps: (1) mapping multilingual metadata to interoperable knowledge sources; (2) assigning multilingual knowledge to cultural data. This paper presents our work on making cultural heritage content available on the Semantic Web and accessible in 15 languages. 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.

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 that making cultural data

available across languages requires an extensive lexical and syntactic knowledge to generate from Semantic Web ontologies. However, while these applications mainly are concerned with two or three languages, it is still not clear how to minimize the efforts in assigning lexical and syntactic knowledge to generate adequate multilingual descriptions from ontologies.

This paper presents our work on making cultural heritage content available on the Semantic Web and accessible in 15 languages. 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. Our experiences reveal some of the challenges we must face before multilingual Semantic Web can be reached.

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

mantic Web. (Androutsopoulos et al., 2001; 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 multi-layered ontology approach for generating multilingual descriptions.

In contrast to (Dekkers et al., 2009; Brugman et al., 2008) whose systems make use of Google translation services, which is data driven, our system is grammar driven.

In the context of cultural heritage there have also been some attempts to generate natural language from ontologies using controlled natural language mechanism (Damljanovic and Bontcheva, 2008). Our approach differs from the above approach as it maps from semantic representations to SPARQL (SPARQL Protocol and RDF Query Language) (Garlik and Andy, 2013) by enabling cross-language interaction using GF. In addition, it constructs answers in the form of coherent texts.

3 The technological infrastructure

Although the architecture of the Semantic Web and Linked Open Data provides access to distributed data sets,¹ 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 the Web of Data. The Reason-able View is a compound dataset composed of several 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

¹<http://linkeddata.org>

access and navigation of the data.

We developed a method to access this data using the Grammatical Framework, GF.

3.1 Museum Reason-able View (MRV)

The Museum Reason-able View is an assembly of RDF datasets. 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;² (b) the unification ontology PROTON,³ an upper-level ontology, consisting of 542 classes and 183 properties; (c) two cultural heritage specific ontologies: (i) CIDOC-CRM (Crofts et al., 2008),⁴ consisting of 90 classes and 148 properties; (ii) Museum Artifacts Ontology (MAO),⁵ developed for mapping between museum data and the K-samsök schema.⁶ It has 10 classes and 20 properties; (d) the Painting ontology,⁷ developed to cover detailed information about painting objects in the framework 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. It has been used as a reference unification ontology to support natural language to ontology interoperability and to allow a unified access to the different cultural heritage datasets.

²DBpedia, structured information from Wikipedia: <http://dbpedia.org>.

³<http://www.ontotext.com/proton-ontology>

⁴<http://www.cidoc-crm.org/>

⁵It 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.

⁶K-samsök (<http://www.ksamsok.se/in-english/>), the Swedish Open Cultural Heritage (SOCH), is a Web service for applications to retrieve data from cultural heritage institutions or associations with cultural heritage information.

⁷<http://spraakdata.gu.se/svedd/painting-ontology/painting.owl>

3.3 Grammatical Framework (GF)

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

GF comes with a resource grammar library (RGL) which aids the development of new grammars for specific domains by providing syntactic operations for basic grammatical constructions (Ranta, 2009). Out of the languages that are available in GF, our application supports the following languages: 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,⁸ and a subset of the DBpedia dataset. These two datasets are very different in size and nature. In the following we describe each of the sets 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 are originally were in Swedish, were translated to English. Example of a record from GCM is shown in Table 4.1.

4.2 DBpedia

The set from DBpedia contains 15, 302 painting records, the data covers 97 languages, the metadata is in English. Example of a record from DBpedia is shown in Table 4.2.

4.3 Transition of data to the MRV

The transition of each data set to the Museum Reason-able View was different for each set.

⁸<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

Making the museum data available through the knowledge infrastructure required translation 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* 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. This was necessary because some year strings contained a mixture of literal and numerical data, for example, words such as *around the year* and *approximately*.

For each painter, museum and painting instance we had a single representation in the data. We used a unified function that truncated Uniform Resource Identifiers (URIs)

```

<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 unique identifiers (IDs). For example, `<http://dbpedia.org/resource/A_Burial_At_Ornans>` was truncated to `A_Burial_At_Ornans`, Salvador Dalí became `Salvador_Dalí`. This ID was searched in the rest of the data, to find unique paintings and unify them under the Painting ontology. For different URLs 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 users who wishes to integrate with the cultural data 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 instances.

Following the GF mechanism, lexicalizations is accomplished through linearizations of functions, which can differ for each language.

5.1 Lexicalizations of classes and properties

Most of the ontology classes that are defined in our grammar are linearized with noun phrases in the concrete syntaxes using the RGL. 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 GF constructions: *mkCN* (Common noun) and *mkN* (Noun).

```

Class: Painting
Swe.
mkCN (mkN "målning");
Fre.
mkCN (mkN "tableau");
Fin,
mkCN (mkN "maalaus");
Ger
mkCN painting_N;

```

```

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 that show how the *Year* function, i.e. *YInt* is lexicalized in six languages. In the examples we find the following GF constructions: *mkAdv* (Verb Phrase modifying adverb), *Prep* (Preposition) and *symp* (Symbolic).

```

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

```

The ontology properties are defined with operations in the concrete syntaxes. Because 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 realization of an ontology property.

5.2 Lexicalizations of instances

The part of the MRV to which we provide translations for consists of 906 instances, their distribution across four classes is provided in Table 3. The lexical units assigned

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

Class	Instances
Title	662
Painter	116
Museum	104
Place	22

to painting titles, painters and museum instances are by default the original strings as they appear in the data. The majority of strings is 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 instances from Wikipedia automatically.

Automatic translation was done by: (1) curling for Web pages for a museum string; (2) extracting the retrieved translated entry for each string; (3) reducing the retrieved list by removing duplicated and ambiguous entries. The process was repeated for each language.

As a result of this process, a list of lexical pairs were created for each language. Museum instances 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 entities was not equal for all languages. The distribution of the amount of translated museum instances is given in Table 4.

Examples of how they are presented in the grammar are:

```

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.

```

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

Language	Translated instances
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

```

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 GF 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. 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 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, especially non-engineers or developers. The main obstacle for this is: 1. the need to master SPARQL, a query language for RDF (Resource Description Framework) (Garlik and Andy, 2013) in order to retrieve semantic content from the knowledge base; 2. holding knowledge about each integrated dataset in the knowledge base.

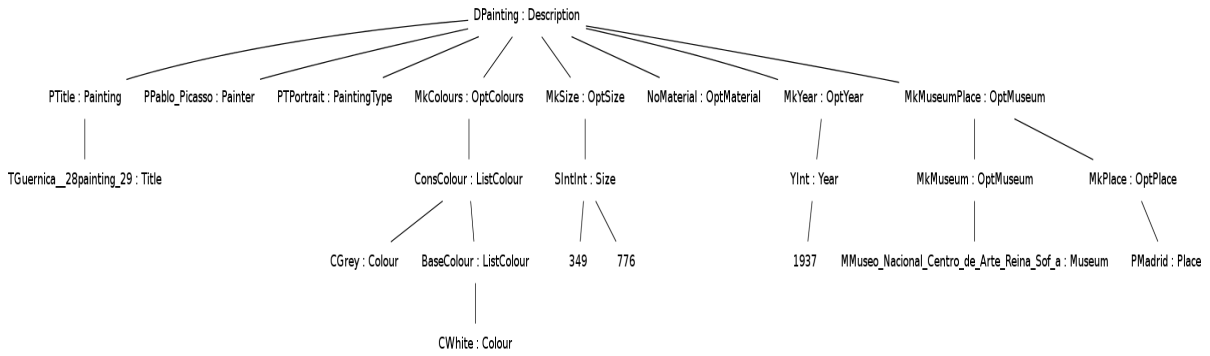


Figure 1: A semantic tree realization of nine ontology classes

Our grammar provides solution to this. We have implemented an extra SPARQL layer that maps from NL to SPARQL and from SPARQL to NL. Some examples of the queries that can be formulated with the multilingual grammar and transformed to SPARQL are:

1. All About X
2. Show everything about X
3. How many X
4. Who is X
5. What is X
6. Some X
7. All X painted by Y
8. Some X painted on Y
9. What is the material of X
10. Show everything about all X that are painted on Y
11. X is by Y
12. X is made of Y

5.5 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 2. With only one function *DPainting* our system is able to generate 16 different text variants. Figure 2, exemplifies two of the text variants.

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.

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 two RDF triples.

To summarize, even though 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 made the translation of museum entries harder because there was no match of those strings in the Wikipedia pages.

TextPaintingBul:	The Potato Eaters е нарисувана от Vincent van Gogh през 1885 година. Тя е с размер 82 см на 115 см. Този експонат е изложен в Музей ван Гог.
TextPaintingCat:	The Potato Eaters està pintat per Vincent van Gogh en 1885. Ell és de 82 sobre 115 cm. Aquesta pintura està exposada al Museu Van Gogh.
TextPaintingDan:	The Potato Eaters blev malet af Vincent van Gogh i 1885. Det er 82 ganger 115 cm. Dette maleri er udstillet på Van Gogh-museet.
TextPaintingDut:	The Potato Eaters werd in 1885 door Vincent van Gogh geschilderd. Het werk is 82 bij 115 cm. Dit schilderij wordt in Van Gogh Museum getoond.
TextPaintingEng:	The Potato Eaters was painted by Vincent van Gogh in 1885. It measures 82 by 115 cm. This painting is displayed at the Van Gogh Museum.
TextPaintingFin:	Maalauksen The Potato Eaters on maalannut Vincent van Gogh vuonna 1885. Se on kokoa 82 kertaa 115 cm. Tämä maalaus on esillä Van Gogh Museumissa.
TextPaintingFre:	The Potato Eaters a été peint par Vincent van Gogh en 1885. Il est de 82 sur 115 cm. Ce tableau est exposé au Musée van Gogh.
TextPaintingGer:	The Potato Eaters wurde in 1885 von Vincent van Gogh gemalt. Das Werk ist 82 mal 115 cm. Dieses Bild ist ausgestellt im Van Gogh Museum.
TextPaintingHeb:	אובלי הבולבוסיים היא תמונה של וינסנט ואן גוך משנת 1885. היא בגודל 82 על 115 סנטימטר. יצירה זו מוצגת במוזיאון ואן גוך.
TextPaintingIta:	The Potato Eaters è dipinto da Vincent van Gogh in 1885. Misura di 82 su 115 cm. Questo dipinto è esposto al Van Gogh Museum.
TextPaintingNor:	The Potato Eaters ble malt av Vincent van Gogh i 1885. Det er 82 ganger 115 cm. Denne malerien er utstilt på Van Gogh Museum.
TextPaintingRon:	The Potato Eaters este pictat de către Vincent van Gogh în 1885. Este din 82 pe 115 cm. Acest tablou este expus în Van Gogh Museum.
TextPaintingRus:	The Potato Eaters нарисованося Vincent van Gogh в 1885 году. Она с размером 82 см на 115 см. Эта картина видится в Музей Винсента ван Гога.
TextPaintingSpa:	The Potato Eaters está pintado por Vincent van Gogh en 1885. Mide 82 por 115 cm. Esta pintura está expuesta en el Museo van Gogh.
TextPaintingSwe:	The Potato Eaters målades av Vincent van Gogh år 1885. Den är 82 gånger 115 cm. Den här målningen är utställd på Van Gogh-museet.
TextPaintingBul:	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.
TextPaintingCat:	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.
TextPaintingDan:	Portrait of Alof de Wignacourt blev malet af Caravaggio i 1607. Det er 195 ganger 134 cm. Dette maleri er udstillet på Louvre.
TextPaintingDut:	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.
TextPaintingEng:	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.
TextPaintingFin:	Maalauksen Portrait of Alof de Wignacourt on maalannut Caravaggio vuonna 1607. Se on kokoa 195 kertaa 134 cm. Tämä maalaus on esillä Louvressa.
TextPaintingFre:	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.
TextPaintingGer:	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.
TextPaintingHeb:	דיוקן של אלוף דה ויגנאקורט היא תמונה של קארבאגיו מ-1607. היא בגודל 195 על 134 סנטימטר. יצירה זו מוצגת במוזיאון הלובר.
TextPaintingIta:	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.
TextPaintingNor:	Portrait of Alof de Wignacourt ble malt av Caravaggio i 1607. Det er 195 ganger 134 cm. Denne malerien er utstilt på Musée du Louvre.
TextPaintingRon:	Portrait of Alof de Wignacourt este pictat de către Caravaggio în 1607. Este din 195 pe 134 cm. Acest tablou este expus în Musée du Louvre.
TextPaintingRus:	Portrait of Alof de Wignacourt нарисованося Caravaggio в 1607 году. Она с размером 195 см на 134 см. Эта картина видится в Лувр.
TextPaintingSpa:	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.
TextPaintingSwe:	Portrait of Alof de Wignacourt målades av Caravaggio år 1607. Den är 195 gånger 134 cm. Den här målningen är utställd på Louvren.

Figure 2: Multilingual generation results

7 Conclusions

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

The presented application 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.

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 Rospoche, 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 (LREC)*.
- Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff, 2008. *Definition of the CIDOC Conceptual Reference Model*.
- D. Damjanovic and K. Bontcheva. 2008. Enhanced semantic access to software artefacts. In *Workshop on Semantic Web Enabled Software Engineering (SWESE) held in conjunction with ISWC*.
- 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. Date: 01 March 2009, Version: 1.7.
- Steve Harris Garlik and Seaborne Andy, 2013. *SPARQL 1.1 Query Language*, March. <http://www.w3.org/TR/sparql11-query/>.
- E Hyvyonen, E. Maekelae, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Ketutula. 2008. Museum finland. In *Finnish Museum on the Semantic Web*.
- 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>.
- 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/>.