Multilingualism in Linked Data

Ontology Engineering Group (OEG)
Artificial Intelligence Department
Universidad Politécnica de Madrid (UPM)

W3C Multilingual Web Workshop
Rome, 12-13 March 2013

G. Aguado  J. Gracia  A. Gómez-Pérez  E. Montiel-Ponsoda  D. Vila
Foundations: the model, the data, URIs and links

RDF(S) models (ontologies) and data

Unique identifiers: URI identify or name a resource

Equivalence links to other datasets

Same As

http://iflastandards.info/ns/fr/frbr/frbrer/C1001

http://iflastandards.info/ns/fr/frbr/frbrer/C1005

http://datos.bne.es/resource/XX1718747

Cervantes

http://datos.bne.es/resource/XX3383563

El Quijote

http://viaf.org/viaf/17220427

Same As

Cervantes

http://dbpedia.org/resource/Miguel_de_Cervantes

http://datos.bne.es/resource/XX3383563

Cervantes

Same As

Work

Ontology

Is creator of

Cervantes

Is creator of

El Quijote

Data

Is a

Is creator of

Person

Is creator of

http://datos.bne.es/resource/XX1718747

http://iflastandards.info/ns/fr/frbr/frbrer/C1001
Sources of information in different languages

- Diverse Information
- Geographical Information
- Library and Cultural Heritage
- Sensor Networks data
- Web 2.0
- REST service annotation

RDF Generation and Linking

Visualization

- Geographical Visualization
- Linked Library Data Visualisation
- Sensor Data Visualisation

Observatory of the Multilingual Web of Data

- Analysis of BTC datasets

**2011**
- Analyzed literals: 1,072,386,405
- Total literals with lang tag: 116,058,734
- % Literals with lang tag: 10.822 %
- % Literals tagged as English: 94.68 %

**2012**
- Analyzed literals: 543,933,327
- Total literals with lang tag: 304,115,676
- % Literals with lang tag: 55.91 %
- % Literals tagged as English: 94.44 %
A motivating example for using multilingual LD [1]

"Dame farmacias de guardia en Colonia que tengan Beglan" (*)

(*) Give me the duty chemists in Cologne having Beglan

Multilingualism and the Linked Data Process

- **Specification**
  - Monolingual or multilingual data resources
    - DB, documents, tables, etc.
    - Linguistic resources: Dictionaries, Lexicons, Thesauri, etc.

- **Modelling**

- **RDF Generation**

- **Links Generation**

- **Publication**

- **Exploitation**

- **Ontology (TBox URIs)**
  - http://phenomenontology.linkeddata.es/ontology/Municipio
  - http://iflastandards.info/ns/fr/frbr/frbrer/C1005

- **Data (ABox URIs)**
  - http://geo.linkeddata.es/resource/Municipio/Madrid
  - http://datos.bne.es/resource/XX1718747

How can we adapt and translate the lexical/terminological layer of an existent ontology into other languages?

**Multilingual labeling** approach if languages involved share a single view on a certain domain.

**Cross-lingual linking** approach if independent monolingual ontologies exist that cover same or similar subject domain (Problems: conceptualization mismatches, or granularity and viewpoint differences).
How to represent multilingual Linked Data?

- Traditional annotation properties for most cases

```
lexinfo:dbpedia:Miguel_de_Cervantes
rdfs:label "Miguel de Cervantes"@es.
"ミゲル・デ・セルバンテス"@ja.
"미겔 데 세르반테스"@ko.
```

- Richer models for more demanding applications

**SKOS-XL**

**LIR**

**LexInfo**
Main issues of cross-lingual linking

- How to **discover** cross-lingual links?
- How to **represent** cross-lingual links?
- How to **store** and reuse cross-lingual links?
How to discover correspondences between ontologies and between LD expressed in different natural languages?
1. Projecting lexical content of the ontology into a common language, then applying traditional OM techniques

2. Comparing ontology entities directly by means of cross-lingual semantic measures (see CIDER-CL)
Cross-lingual Link Storage and Reuse

- Links can be discovered:
  - runtime -> need of scalable techniques
  - offline -> need of storage methods

- Storage
  - Following Linked Data principles
  - Links can be stored *jointly* to some of the data sources that they relate (e.g., during LD generation)
  - Links can be stored in *separate repositories* to be accessed by semantic applications (e.g., for CL-Question Answering)
Multilingualism and the Linked Data Process

How can a user pose questions in their own language to be processed against the web of Linked Data?

1. Multilingual query interpretation
2. Query federation, ...

“Colonia”    “farmacia”

Semantic query

How should the results of a semantic query be adapted to the linguistic and cultural background of a user?

1. Adaptation and localization of user interfaces
2. Natural language generation
3. Presentation views to specific linguistic and cultural contexts
Services for the Multilingual Web of Data

Services for cross-lingual access

Users

Linked Data

Multilingual mappings

Multilingual linguistic information

Services for cross-lingual linkage

Services for generating multilingual Linked Data

Data silos

Services for translation and ontology localization
Thanks for your attention!
Research agenda on Multilingual LD at OEG

- **Ontology lexica** representation
  Elena Montiel, Lupe Aguado

- **Lexico-syntactic patterns**
  Elena Montiel, Lupe Aguado

- **Ontology localisation** (translation)
  Elena Montiel, Jorge Gracia, Asun Gomez-Perez

- **Exploratory analysis** of the Multilingual Web of Data
  Daniel Vila, Asun Gómez-Pérez, Jorge Gracia

- **Cross-lingual ontology and Instance matching**
  Jorge Gracia, Daniel Vila

- **Query federation**
  Oscar Corcho