Cross-language Linking of eGov Services to the LOD Cloud

Fedelucio Narducci*, Matteo Palmonari*, Giovanni Semeraro°

*DISCO, University of Milan-Bicocca, Italy
°Department of Computer Science, University of Bari Aldo Moro, Italy

AI for Smart Cities Workshop
AI*IA 2013 - 25th Year Anniversary
XIII Conference
Turin (Italy), December 5, 2013
EC 6 axes
to evaluate ‘city smartness’

http://www.smart-cities.eu/model.html
Cross-language Linking of Open Government Data

- A large amount of Open Government Data in many languages*:
  - 1,000,000+ datasets published online (February 2013)
  - 40 different countries
  - 24 different languages

*http://logd.tw.rpi.edu/iogds_data_analytics
Government service catalogs are part of the LOD cloud

- Effective Service Delivery (ESD)-toolkit
- European Local Government Service List (LGSL)
  - 2000+ interlinked public services in 6 languages
Cross-language Linking of eGov Services

Why is it useful?

• Advantages for PAs
  - Compare local service offerings with best practices in other countries
  - Support interoperability among PAs of different countries and other service providers
  - Enrich service descriptions with additional information via links to LGSL (e.g., link to life event ontologies)

• Advantages for citizens
  - Find eGov services when in a foreign country
  - Towards cross-language service access

Costly and Error Prone Activity
Catalogs of several hundreds of services
Why is it challenging?

- Challenging cross-language matching problem
- Most of the approaches:
  - or report problems when automatic translation returns descriptions with heterogeneous vocabulary [Hertling & Paulheim 2012]

Semantic heterogeneity
- not a mere “translation” problem
- cultural bias

≈ sameAs links

Ultra-short descriptions
CroSeR
Cross-language Service Retriever

CROSER: Cross-language Service Retriever
connects your service catalogs

Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English

Based on Machine Translation and Explicit Semantic Analysis (ESA)

TRY IT @ http://siti-rack.siti.disco.unimib.it:8080/croser/
CROSER: Cross-language Service Retriever

Web tool to **support the linkage** of a source eGov service catalog represented in **any language** to a target catalog represented in **English**

Based on **Machine Translation** and **Explicit Semantic Analysis (ESA)**
CroSeR
Cross-language Service Retriever

CROSER: Cross-language Service Retriever
connects your service catalogs

Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English

Based on Machine Translation and Explicit Semantic Analysis (ESA)
Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English.

Based on Machine Translation and Explicit Semantic Analysis (ESA)
Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English.

Based on Machine Translation and Explicit Semantic Analysis (ESA).
CroSeR
Cross-language Service Retriever

CROSER: Cross-language Service Retriever
connects your service catalogs

- Load a catalog

- Select a source service

Service Catalog

<table>
<thead>
<tr>
<th>Service</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>Straßen- und Brückenunterhaltung, Straßenverkehrstechnik</td>
</tr>
<tr>
<td>63</td>
<td>Hausumringer Bremerhaven</td>
</tr>
<tr>
<td>65</td>
<td>Fäll- und Schnittgenehmigung nach der bremischen Baumschutzverordnung</td>
</tr>
<tr>
<td>79</td>
<td>Spritzen - gebrauchte - Entsorgung</td>
</tr>
<tr>
<td>133</td>
<td>Briefwahl</td>
</tr>
</tbody>
</table>

Service Matched from ESD

1. Voting
2. Proxy voting
3. Postal voting
4. Members - elections - polling stations
5. Election results
6. Members - elections - results
7. Electoral register

Service Info

- Look at the retrieved services (link recommendations)
- Link SKOS broader / exact / narrower match

Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English

Based on Machine Translation and Explicit Semantic Analysis (ESA)
CroSeR
Cross-language Service Retriever

CROSER: Cross-language Service Retriever
connects your service catalogs

Web tool to support the linkage of a source eGov service catalog represented in any language to a target catalog represented in English

Based on Machine Translation and Explicit Semantic Analysis (ESA)
CroSeR: Matching Approach

Machine Translation of Service Descriptions

Extraction of ESA-based representations and indexing (Vector Space Model)

Top-k Service Retrieval by Cosine Similarity
Explicit Semantic Analysis (ESA)

Technique able to provide a fine-grained semantic representation of natural language texts in a high-dimensional space of comprehensible concepts derived from Wikipedia [GM06]

Wikipedia viewed as an ontology = a collection of ~1M concepts

Experiments: Discussion

• CroSeR finds matchings that cannot be discovered by machine translation + keyword comparison

• CroSeR’s recommendations can support the users to refine the links

GT: “Absentee Ballot”
References (for details)

- Model and Experimental Evaluation @ISWC 2013

- System Demo @ECIR 2014
Thanks, questions?