Ontology Lexicalisation and Localisation for the Multilingual Semantic Web

Paul Buitelaar - Monnet Coordinator
Digital Enterprise Research Institute (DERI)
National University of Ireland, Galway
Business Information query in EN
fixed assets; EU wind energy companies; 2005-2010

Business Information in EN, DE, NL, ES etc.
<table>
<thead>
<tr>
<th>RELATION</th>
<th>LABEL</th>
<th>LANGUAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>exactMatch</td>
<td>Fixed assets</td>
<td>en</td>
</tr>
<tr>
<td>exactMatch</td>
<td>Vaste activa</td>
<td>nl</td>
</tr>
<tr>
<td>exactMatch</td>
<td>Anlagevermögen</td>
<td>de</td>
</tr>
<tr>
<td>exactMatch</td>
<td>Actifs immobilisés</td>
<td>fr</td>
</tr>
<tr>
<td>narrowMatch</td>
<td>Errichtungs- und Erweiterungsaufwendungen</td>
<td>de</td>
</tr>
<tr>
<td>narrowMatch</td>
<td>Formation expenses</td>
<td>en</td>
</tr>
<tr>
<td>narrowMatch</td>
<td>Oprichtingskosten</td>
<td>nl</td>
</tr>
<tr>
<td>narrowMatch</td>
<td>Frais d'établissement</td>
<td>fr</td>
</tr>
<tr>
<td>exactMatch</td>
<td>A) ACTIVO NO CORRIENTE</td>
<td>es</td>
</tr>
</tbody>
</table>
Monnet in a Nutshell

Lexicalization Service

Localization Service

Knowledge Access and Presentation Service

Knowledge Base

Corpus Service

Information Extraction Service

Ontology

en
es
de
nl

translator

expert

Corpus Service

Knowledge Base
Research Objectives

– Development and use of ‘multilingual ontologies’
  • ontologies with rich multilingual descriptors

– Exploit ‘domain semantics’ to improve Machine Translation
  • use of ontological, terminological, linguistic knowledge

Use Cases

– Financial Use Case
  • Cross-lingual Business Intelligence

– Public Services Use Case
  • Multilingual Access to Government Information
Harmonizing Business Registration across Europe

XBRL (eXtensible Business Reporting Language) Europe Working Group works with Monnet on the \textit{xEBR taxonomy}.

\textit{xEBR} (XBRL European Business Register) taxonomy defines common concepts with mappings to country/language specific taxonomies.

- National Bank of Belgium (Belgium)
- Eogs / DCCA (Denmark)
- Registrite ja infosüsteemide Keskus eRik (Estonia)
- Bilans Service - Infogreffe (France)
- Bundesanzeiger (Germany)
- Infocamere (Italy)
- RSCL (Luxembourg)
- Kamer van Koophandel (Netherlands)
- Informa DB – Colegio de Registradores (Spain)
- Bolagsverket (Sweden)
- Companies House (United Kingdom)
- EBR (Europe)
- GBR (Global)
- IASCF
- Bank of Spain
- Software – Audit – Consulting
| German XBRL term (DE-GAAP)               | ausstehende Einlagen, davon eingefordert |
| English XBRL term (UK-GAAP)             | unpaid calls and subscribed capital      |
| Google Translate (German > English)     | outstanding deposits, which called for   |
| Monnet MT with domain training          | unpaid calls and subscribed capital      |
| German XBRL term (DE-GAAP)              | außerordentliches Ergebnis               |
| English XBRL term (UK-GAAP)             | extraordinary result                    |
| Google Translate (German > English)     | extraordinary items                     |
| Monnet MT with domain training          | extraordinary result                    |

Domain training with hybrid methods:

- Domain lexicon generation from Wikipedia & domain parallel corpora
- LDA topic modelling with features (words) mixed-in from the ontology
- Alignment and disambiguation across web ontologies for translation mining
Welcome

If you come to the Netherlands to live, work or study, you are likely to have some questions about the arrangements you need to make. This site will guide you to the government organizations you may have to deal with.

What can you do?

If you fill in your profile, you will see a list of subjects that are relevant for you, with links to information from the organizations concerned. You can also choose a subject directly.

Your profile

What is your age?
[ ] Make your choice

What is your nationality?
[ ] Make your choice

Length of your stay?
[ ] Make your choice

The purpose of your stay?
(You may give more than one answer)

[ ] work in Holland
[ ] study in Holland
[ ] live in Holland

Create my checklist

Translation of Dutch regulation (legal ontology) into several EU languages:
⇒ Immigration law
⇒ Tax law
⇒ Student benefit law
⇒ Health care benefit law
⇒ Social security law
⇒ Law on higher education
Different Requirements in Public Services Use Case
Complex Semantics (Modal, Procedural) in Ontology Label
  – Analyze, Translate & Generate
  – Multilingual Generation (combined with) Machine Translation

<table>
<thead>
<tr>
<th>Type</th>
<th>Canonical Label</th>
<th>Operator(s)</th>
<th>Labeltype</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appointment</td>
<td>Financial health check appointment</td>
<td>scheduled</td>
<td>Postulation scheduled</td>
<td>The Financial check appointment is scheduled</td>
</tr>
<tr>
<td></td>
<td></td>
<td>occured</td>
<td>Question scheduled</td>
<td>Is the Financial check appointment scheduled?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negation scheduled</td>
<td>The Financial check appointment is not scheduled</td>
</tr>
<tr>
<td>Note</td>
<td>Archive note</td>
<td>available</td>
<td>Postulation available</td>
<td>The Archive note is available</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Question available</td>
<td>Is the Archive note available?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negation available</td>
<td>The Archive note is not available</td>
</tr>
<tr>
<td>Time limit</td>
<td>Acceptable response time</td>
<td>expired</td>
<td>Postulation expired</td>
<td>The Acceptable response time is expired</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Question expired</td>
<td>Is the Acceptable response time expired?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negation expired</td>
<td>The Acceptable response time is not expired</td>
</tr>
<tr>
<td>Involvement role</td>
<td>Case handler</td>
<td>involved</td>
<td>Postulation involved</td>
<td>The Case handler is involved</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Question involved</td>
<td>Is the Case handler involved?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Negation involved</td>
<td>The Case handler is not involved</td>
</tr>
</tbody>
</table>

GELATO (Generation of LAnguage and Text from Ontologies)
  – Label > Lexicalize > Translate + Operators > Multilingual Generation
  – Explore joint research with the MOLTO project
Motivation

– Lexical layer to represent internal linguistic structure of ontology labels (terms, statements)

Use Cases

– Ontology Localisation & Verbalisation, Ontology-based Information Extraction, Ontology Learning, etc.

W3C Ontology-Lexicon Community Group

– http://www.w3.org/community/ontolex/
– Monnet proposed format: *lexicon model for ontologies* http://monnetproject.deri.ie/lemonsourcet/
Thanks for your Attention!

http://www.monnet-project.eu/

http://www.w3.org/community/ontolex/