

# Ontology Lexicalisation and Localisation for the Multilingual Semantic Web

Paul Buitelaar - Monnet Coordinator  
Digital Enterprise Research Institute (DERI)  
National University of Ireland, Galway



Monnet is supported  
by the European Union  
under Grant No. 248458

Business Information query in EN  
fixed assets ; EU wind energy companies ; 2005-2010

Business Information in EN, DE, NL, ES etc.



RELATION	LABEL	LANGUAGE
exactMatch	Fixed assets	en
exactMatch	Vaste activa	nl
exactMatch	Anlagevermögen	de
exactMatch	Actifs immobilisés	fr
narrowMatch	Errichtungs- und Erweiterungsaufwendungen	de
narrowMatch	Formation expenses	en
narrowMatch	Oprichtingskosten	nl
narrowMatch	Frais d'établissement	fr
exactMatch	A) ACTIVO NO CORRIENTE	es

Financial Item

Fixed Assets [Total]

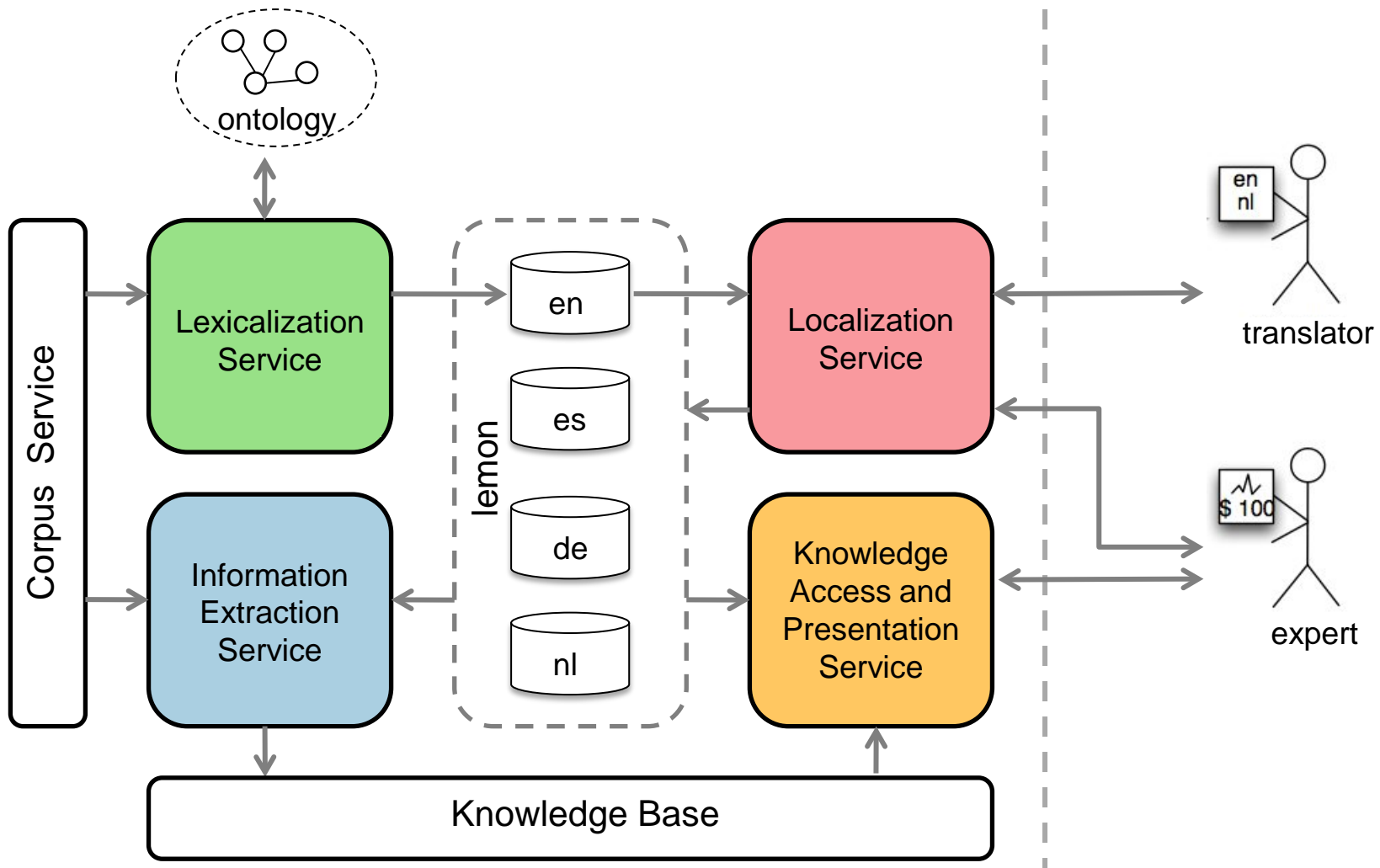
COMPANY

fersa

fersa

3e

3e



## 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 *xEBR taxonomy*

*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

<b>German XBRL term (DE-GAAP)</b>	<i>ausstehende Einlagen, davon eingefordert</i>
<b>English XBRL term (UK-GAAP)</b>	<i>unpaid calls and subscribed capital</i>
<b>Google Translate (German &gt; English)</b>	<i>outstanding deposits, which called for</i>
<b>Monnet MT with domain training</b>	<i>unpaid calls and subscribed capital</i>
<b>German XBRL term (DE-GAAP)</b>	<i>außerordentliches Ergebnis</i>
<b>English XBRL term (UK-GAAP)</b>	<i>extraordinary result</i>
<b>Google Translate (German &gt; English)</b>	<i>extraordinary items</i>
<b>Monnet MT with domain training</b>	<i>extraordinary result</i>

## 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



New to Holland *Dutch government immigration website*

Home Checklist Subjects

## 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

## Subjects

- Education
- Employment
- Permits and visa
- Social security
- Taxes
- Vehicles



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

Type	Canonical Label	Operator(s)	Labeltype	Label
Appointment	Financial health check appointment	scheduled	Postulation scheduled	The Financial check appointment is scheduled
			Question scheduled	Is the Financial check appointment scheduled?
			Negation scheduled	The Financial check appointment is not scheduled
		occured	Postulation occured	The Financial check appointment has occurred
			Question occured	Has the Financial check appointment occurred?
			Negation occured	The Financial check appointment has not occurred
Note	Archive note	available	Postulation available	The Archive note is available
			Question available	Is the Archive note available?
			Negation available	The Archive note is not available
Time limit	Acceptable response time	expired	Postulation expired	The Acceptable response time is expired
			Question expired	Is the Acceptable response time expired?
			Negation expired	The Acceptable response time is not expired
Involvement role	Case handler	involved	Postulation involved	The Case handler is involved
			Question involved	Is the Case handler involved?
			Negation involved	The Case handler is not involved

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/lemonsourc/>



**Thanks for your Attention!**

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

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