



PERIODIC REPORT: 3rd Annual Report

Grant Agreement number: 248458

Project acronym: MONNET

Project title: Multilingual Ontologies for Networked Knowledge

Funding Scheme: STREP

Date of latest version of Annex I against which the assessment will be made: 20/10/2009

Periodic report: 3rd annual report

Period covered: from: March 1st, 2012 to: February 28th, 2013

Name, title and organisation of the scientific representative of the project's coordinator:

Dr. Paul Buitelaar, Digital Enterprise Research Institute, National University of Ireland, Galway

Tel: +353 91 495007

Fax: +353 91 495541

E-mail: paul.buitelaar@deri.org

Project website address: <http://www.monnet-project.eu/>

Publishable Summary

The Monnet project is concerned mainly with ontology localisation, i.e., the translation of the lexico-terminological level of ontologies (often referred to as the ‘ontology labels’). The project outcomes can be described as a set of software components as follows, all of which can be used in combination as well as stand-alone:

- Ontology Lexicalization
- Ontology Localization
- Cross-lingual Ontology-based Information Extraction
- Cross-lingual Knowledge Access & Presentation

The following picture illustrates the envisioned connections between these components, with their inputs and outputs:

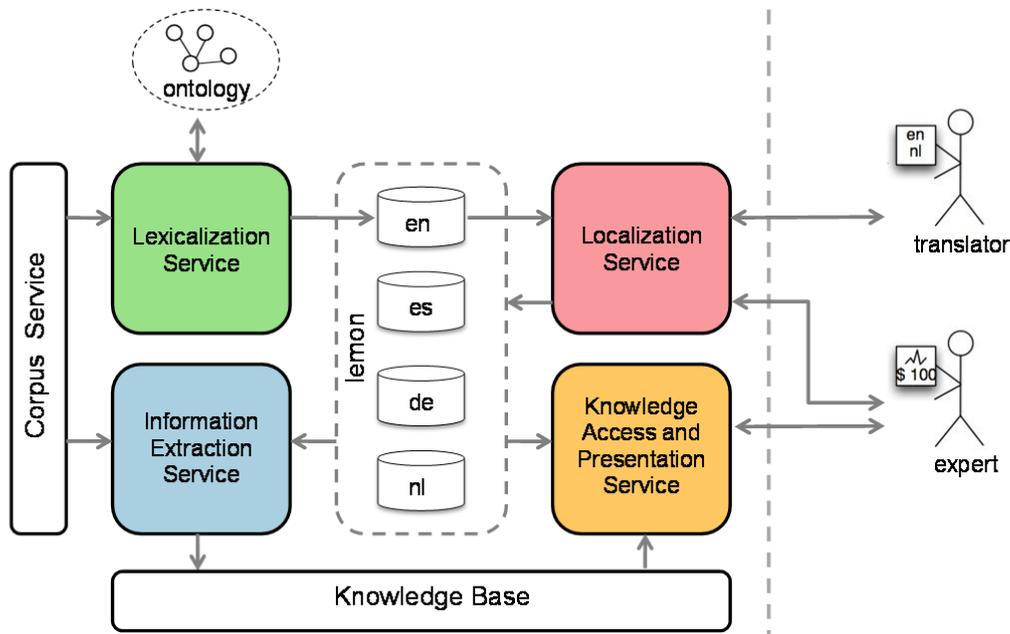


Figure 1: Monnet overview

Ontology Lexicalization The core objective of Monnet is the provision of advanced services for the translation of the lexico-terminological level of ontologies, which will be instantiated by the ‘localization service’. However, as ontologies often have only a very limited representation of lexico-terminological information, a first step will be to analyze a given ontology and enrich it with appropriate information on i) the terminological structure of ontology labels, ii) linguistic information on terminology items, and iii) analysis of implicit semantics where needed. Together we refer to these analysis and enrichment steps as “ontology lexicalisation”, which will be instantiated by the ‘lexicalisation service’ that takes as input an ontology and outputs an ‘ontology-lexicon’ for at least one default language (depending on the language that was used in defining ontology labels). A ‘corpus service’ enables access to external domain corpus evidence for

modelling and analyzing language use in the ontology labels. The ontology-lexicon is represented on the basis of the ‘lemon’ format¹, a lexicon model for ontologies that has been defined by the Monnet project for the appropriate integration of lexical, linguistic and terminological information in ontologies.

Ontology Localization After the generation of an appropriate ontology-lexicon by the ‘lexicalisation service’, the ‘localisation service’ can now operate on this for the translation process, i.e., note that Monnet thus implements ontology localisation by translation of the ontology-lexicon. A hybrid machine translation (MT) approach has been explored that builds on state of the art statistical MT as provided by the open source Moses² system, and knowledge-based approaches using domain knowledge provided by ontologies and associated domain terminologies. The ontology localisation service has been developed and tested in the context of two real-life use cases on the translation of i) financial services taxonomies such as FINREP defined in XBRL (eXtensible Business Reporting Language) developed by Monnet partner XBRL Europe, and ii) public services ontologies developed by Monnet partner Be Informed for the Dutch government.

Cross-lingual Ontology-based Information Extraction To provide a further proof of concept in a real-life use case, the Monnet project has implemented a cross-lingual ontology-based information extraction component that exploits ontology-lexica (for several of the project languages English, German, Spanish and Dutch) in financial fact extraction with the XBRL-based xEBR vocabulary for European-wide harmonization of business register information. The ‘information extraction service’ instantiates this functionality by use of the lemon API (for accessing lemon-based ontology lexica) and the corpus service (for accessing linguistically/semantically annotated domain corpora) to generate a knowledge base (‘KB’) of financial facts.

Cross-lingual Knowledge Access & Presentation Finally, the ‘knowledge access & presentation service’ delivers a number of components and frameworks for localization, supporting multilingual querying and visualisation of the KB in the context of a real-life use case on cross-lingual business intelligence. These components provide cross-lingual, ontology-based access to financial information by use of XBRL semantics in combination with multilingual ontology-lexica.

¹ <http://lemon-model.net/>

² <http://www.statmt.org/moses/>