

MONNET: Multilingual Ontologies for Networked Knowledge

Mihael Arcan Paul Buitelaar

Unit for Natural Language Processing, DERI, NUI Galway, Ireland
{mihael.arcan , paul.buitelaar}@deri.org

Abstract

The Monnet project is concerned mainly with ontology localisation, i.e., the translation of the lexico-terminological level of ontologies. 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 Lexicalisation and Localization, Cross-lingual Ontology-based Information Extraction, Cross-lingual Knowledge Access & Presentation

Ontology Lexicalisation 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 analyse 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.¹ A "corpus service" enables access to external domain corpus evidence for modelling and analysing language use in the ontology labels. The ontology-lexicon is represented on the basis of the "lemon" format,² 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 Moses system, and knowledge-based approaches using domain knowledge provided by ontologies and associated domain terminologies. The ontology localisation service³ 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 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 and the corpus service 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.⁴ These components provide cross-lingual, ontology-based access to financial information by use of XBRL semantics in combination with multilingual ontology-lexica.

¹<http://monnetproject.deri.ie/lemonsources/>

²<http://lemon-model.net/>

Sima'an, K., Forcada, M.L., Grasmick, D., Depraetere, H., Way, A. (eds.)

Proceedings of the XIV Machine Translation Summit (Nice, September 2-6, 2013), p. 431.

©2013 The authors. This article is licensed under a Creative Commons 3.0 licence, no derivative works, attribution, CC-BY-ND.

³<http://monnet01.sindice.net/term-translation/>

⁴<http://monnet01.sindice.net:8080/financial/>