Accepted Manuscript

Language and domain aware lightweight ontology matching

Gábor Bella, Fausto Giunchiglia, Fiona McNeill

PII: S1570-8268(17)30016-1
DOI: http://dx.doi.org/10.1016/j.websem.2017.03.003
Reference: WEBSEM 432

To appear in: *Web Semantics: Science, Services and Agents on the World Wide Web*

Received date : 24 August 2016
Revised date : 15 March 2017
Accepted date : 26 March 2017

Please cite this article as: G. Bella, F. Giunchiglia, F. McNeill, Language and domain aware lightweight ontology matching, *Web Semantics: Science, Services and Agents on the World Wide Web* (2017), http://dx.doi.org/10.1016/j.websem.2017.03.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*Manuscript
Click here to view linked References

ACCEPTED MANUSCRIPT

Language and Domain Aware Lightweight Ontology Matching

Gábor Bella[1], Fausto Giunchiglia[1], Fiona McNeill[2]

Abstract

Concepts and relations in ontologies and in other knowledge organisation systems are usually annotated with natural language labels. Most ontology matchers rely on such labels in element-level matching techniques. State-of-the-art approaches, however, tend to make implicit assumptions about the language used in labels (usually English) and are either domain-agnostic or are built for a specific domain. When faced with labels in different languages, most approaches resort to general-purpose machine translation services to reduce the problem to monolingual English-only matching. We investigate a thoroughly different and highly extensible solution based on *semantic matching* where labels are parsed by multilingual natural language processing and then matched using language-independent and domain-aware background knowledge acting as an interlingua. The method is implemented in NuSM, the language and domain aware evolution of the SMATCH semantic matcher, and is evaluated against a translation-based approach. We also design and evaluate a fusion matcher that combines the outputs of the two techniques in order to boost precision or recall beyond the results produced by either technique alone.

Keywords: cross-lingual matching, multilingual matching, domains, ontology matching, semantic matching, machine translation

1. Introduction

Ontologies and other knowledge organisation systems, while usually serving a purpose of standardisation or generalisation, stem from local needs and practices. By *local* we understand *within an administrative unit* such as a country or a region as well as *within an application domain* such as medicine or transport. Accordingly, ontologies tend to target specific domains and the labels annotating their elements—concepts, relations, metadata—tend to be expressed in the local language. This is especially true for *lightweight ontologies* [20]: classification hierarchies, taxonomies, and other tree-structured data schemas widely used around the world as simple, well-understood, semi-formal resources for knowledge organisation. Such resources often play normative roles on the national level in public services, industry, or commerce, as a means for classification (of documents, books, open data, commercial products, web pages, etc.) as well as being sources of shared vocabularies for actors cooperating in a given domain.

Ontology matching [15] is a process that creates and maintains alignments between elements of two ontologies covering overlapping areas of knowledge. We define *language aware* or *multilingual* matching as a type of ontology matching where a multilingual setting is explicitly assumed, i.e., the matcher is capable of dealing with ontologies expressed in multiple languages. Likewise, we define *domain aware* matching as capable of dealing with domain-specific knowledge and domain terms with specialised meanings.

Activities on supra-national levels such as international trade and mobility need to rely on the interoperability and integration of knowledge organisation resources across countries, languages, and sometimes across domains. *Cross-lingual matching* is a specific case of language aware matching when ontologies in different languages need to be aligned. Likewise, *cross-domain matching* is used to match ontologies pertaining to different domains of knowledge. An example of a simultaneously cross-lingual and cross-domain matching problem is the case of *cross-border emergency response* where

[1]University of Trento, via Sommarive 5, 38123 Trento, Italy.
[2]Heriot-Watt University, Edinburgh EH14 4AS, Scotland.

Download English Version:

https://daneshyari.com/en/article/4973345

Download Persian Version:

https://daneshyari.com/article/4973345

Daneshyari.com