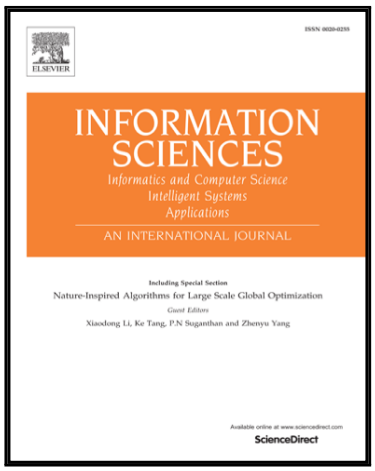

Cross-lingual link discovery with TR-ESA

Fedelucio Narducci, Matteo Palmonari, Giovanni Semeraro

PII: S0020-0255(17)30490-5
DOI: 10.1016/j.ins.2017.02.019
Reference: INS 12742

To appear in: *Information Sciences*

Received date: 20 May 2016
Revised date: 18 October 2016
Accepted date: 8 February 2017

Please cite this article as: Fedelucio Narducci, Matteo Palmonari, Giovanni Semeraro, Cross-lingual link discovery with TR-ESA, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.02.019

Cross-lingual link discovery with TR-ESA

Fedelucio Narducci[a], Matteo Palmonari[b], Giovanni Semeraro[a]

[a] *Department of Computer Science, University of Bari Aldo Moro*
Via E. Orabona 4, I-70125 Bari, Italy

[b] *Department of Informatics, Systems and Communication, University of Milano Bicocca*
Viale Sarca, 336, I-20126 Milano, Italy

Abstract

Cross-lingual data linking is the problem of establishing links between *resources*, such as places, services, or movies, which are described in different languages. In cross-lingual data linking it is often the case that very short descriptions have to be matched, which makes the problem even more challenging. This work presents a method named TRanslation-based Explicit Semantic Analysis (TR-ESA) to represent and match short textual descriptions available in different languages. TR-ESA translates short descriptions in any given language into a pivot language by exploiting a machine translation tool. Then, it generates a Wikipedia-based representation of the translated text by using the Explicit Semantic Analysis technique. The resulting representations are used to match short descriptions in different languages. The method is incorporated in CroSeR (Cross-lingual Service Retrieval), an interactive data linking tool that recommends potential matches to users. We compared results coming from an in-vitro evaluation on a gold standard consisting of five datasets in different languages, with an in-vivo experiment that involved human experts supported by CroSeR. The in-vivo evaluation confirmed the results of the in-vitro evaluation and the overall effectiveness of the proposed method.

Keywords: Cross-lingual Matching, Cross-lingual Data Linking, Wikipedia

Email addresses: fedelucio.narducci@uniba.it (Fedelucio Narducci), palmonari@disco.unimib.it (Matteo Palmonari), giovanni.semeraro@uniba.it (Giovanni Semeraro)