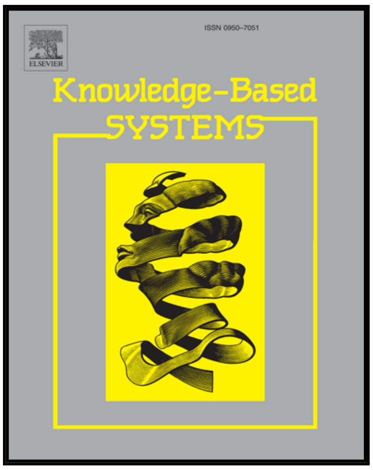

# Bilingual Embeddings with Random Walks over Multilingual Wordnets

Josu Goikoetxea, Aitor Soroa, Eneko Agirre

PII: S0950-7051(18)30141-2
DOI: 10.1016/j.knosys.2018.03.017
Reference: KNOSYS 4266

To appear in: *Knowledge-Based Systems*

Received date: 23 June 2017
Revised date: 8 March 2018
Accepted date: 10 March 2018

Please cite this article as: Josu Goikoetxea, Aitor Soroa, Eneko Agirre, Bilingual Embeddings with Random Walks over Multilingual Wordnets, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.03.017

# Bilingual Embeddings with Random Walks over Multilingual Wordnets

Josu Goikoetxea[†], Aitor Soroa, Eneko Agirre

*IXA NLP group, Faculty of Informatics, UPV/EHU,*
*Manuel Lardizabal 1 (20018), Donostia, Basque Country*

**Abstract**

Bilingual word embeddings represent words of two languages in the same space, and allow to transfer knowledge from one language to the other without machine translation. The main approach is to train monolingual embeddings first and then map them using bilingual dictionaries. In this work, we present a novel method to learn bilingual embeddings based on multilingual knowledge bases (KB) such as WordNet. Our method extracts bilingual information from multilingual wordnets via random walks and learns a joint embedding space in one go. We further reinforce cross-lingual equivalence adding bilingual constraints in the loss function of the popular Skip-gram model. Our experiments on twelve cross-lingual word similarity and relatedness datasets in six language pairs covering four languages show that: 1) our method outperforms the state-of-the-art mapping method using dictionaries; 2) multilingual wordnets on their own improve over text-based systems in similarity datasets; 3) the combination of wordnet-generated information and text is key for good results. Our method can be applied to richer KBs like DBpedia or BabelNet, and can be easily extended to multilingual embeddings. All our software and resources are open source.

*Keywords:* multilinguality, distributional semantics, embeddings, random walks, WordNet

[†] Corresponding author. *Email adress:* josu.goikoetxea@ehu.eus