

Scientific Report of Short Term Scientific Mission

COST STSM Reference Number: COST-STSM-IS1305- 160215-055460

Period: 20-02-2015 to 20-5-2015

Duration: 90 working days.

COST Action: IS1305

STSM type: Regular (from Spain to Italy)

STSM Title: Extending the Wikipedia Bitaxonomy

Guest/STSM applicant: Luis Espinosa, Universitat Pompeu Fabra

Host: Dr. Roberto Navigli

Sapienza University of Rome

navigli@di.uniroma1.it

1. Purpose of the STSM

During my STSM at the Linguistic Computing Laboratory, lead by prof. Roberto Navigli, the main goal was to explore potential avenues for improvement in the Wikipedia Bitaxonomy, which is a two-layered taxonomy over Wikipedia Pages and Categories, and which is the *skeleton* of the well-known semantic network BabelNet (Navigli and Ponzetto, 2010).

2. Description of the work carried out during the STSM

I approached this from two perspectives: (1) Enriching an existing general-purpose taxonomy with additional *is-a* or *hypernym* edges, coming from textual definitions. Another goal was to collect collections of unstructured knowledge in the form of *triples*, such as *X-isA Y* or *X-hypernym Y*, to be the first stage in a more ambitious project, where these relations would be used to build a *Bitaxonomy*, therefore, in BabelNet.

The outcome of this project is two submitted papers to EMNLP 2015, in which I describe the experiments carried out in both projects, and which can be summarized as follows: to leverage WiBi for finding new taxonomic relations thanks to our algorithm, which uses a confidence threshold based on the relevance of each *hypernym > hyponym* relation.