

Enhancing Text-Based Relatedness Measures with Semantic Web Data

Ana Gjorgjevikj, Riste Stojanov, and Dimitar Trajanov

Faculty of Computer Science and Engineering, Ss. Cyril and Methodius in Skopje,
Macedonia

ana.gorgevic@gmail.com, riste.stojanov@finki.ukim.mk,
dimitar.trajanov@finki.ukim.mk

Abstract. Entity relatedness measures quantify the amount of association between two entities, such as people, places or events, and are fundamental part of many Natural Language Processing and Information Retrieval applications. Calculating entity relatedness requires access to entity specific information, so a very common practice is to use Wikipedia or its Semantic Web representations as source of knowledge. This paper explores which of the different semantic relationships that associate two entities in DBpedia are good indicators of their relatedness and could be used to enhance some of the standard text-based relatedness measures. The ultimate goal is learning a well performing relatedness calculation method that does not require vast amount of preprocessing, but is applicable in cases when entities lack either textual context or semantic relationships. The KORE entity relatedness dataset was used for learning a convenient and well performing method for measuring relatedness and its evaluation.

Keywords: Entity relatedness, Semantic relatedness, Entity ranking, DBpedia