

WIR Project Proposal

Members: *Angelo Catalani, Valerio Colitta, Alessandro Lo Presti*

1.1 Proposal

The paper describes a new approach to compute similarity between documents and queries. It not only takes into account tf-idf, but also SIMILARITY among terms via a new formula.

Terms are identified through WordNet, where they are linked to each other in different ways. Given a particular synset (sense) for a term, you can traverse the whole network to find related terms (superclasses, subclasses, etc).

The paper introduces a specific metric for quantifying the similarity of two words by measuring shortest path between their senses in the WordNet graph.

We implement a new metric, using the WordNet graph. In particular we establish that the similarity depends on the least common ancestor in the term-hypernym path (the least the better).