

University of Macau

Abstract

WEB QUERY EXPANSION BY WORDNET

by Cheang Chan Wa

Thesis Supervisor: Associate Professor Gong Zhiguo
E-Commerce Technology Master Degree Program

In this thesis, we address a novel method of Web query expansion by using WordNet and TSN. WordNet is an online lexical dictionary which describes word relationships in three dimensions of Hypernym, Hyponym and Synonym. And their impacts to the expansion are different. We provide quantitative descriptions of the query expansion impact along each dimension. However, WordNet may bring many noises for the expansion since its collection independent characteristic. Furthermore, it may not catch current state of words and their relationships since the explosive increase of the Web.

To overcome those problems, firstly, we use WordNet::Similarity, which is a variety of semantic similarity measures that can be used in conjunction with WordNet, to reduce our web query expansion. Then a collection-based TSN (Term Semantic Network) is created with respect to word co-occurrence in the collection in our system. We use TSN both as a filter and a supplement for WordNet in expanding queries. We also provide a quantitatively study as what is the best way for the expansion with TSN. In our system, we combine the query expansions along each semantic dimension as our overall solution. Our experiments reveal that the combined expansion can provide a satisfied result for the Web query performance.