

# Exploiting Question Concepts for Query Expansion

Hae-Jung Kim, Ki-Dong Bu, Junghyun Kim and Sang-Jo Lee

Department of Computer Engineering, Kyungpook National University,  
Sangyuk-dong, Puk-gu, Daegu, 702-701, Korea  
hjkim325@hanmail.net

**Abstract.** In this paper, we present an efficient semantic query expansion methodology based on a question concept list comprised of terms that are semantically close to concepts represented in a query. The proposed system first constructs a concept list for each question concept and then learns the concept list for each question concept. When a new query is given, the question is classified into the question concept, and the query is expanded using the concept list of the classified concept. In the question answering experiments on 42,654 Wall Street Journal documents of the TREC collection, the traditional system showed in 0.223 in MRR and the proposed system showed 0.50 superior to the traditional question answering system.

## 1 Introduction

Question answering (QA) systems assign relevance degrees to words, paragraphs or clauses based on a given query, and then provide answers ranked according to relevance. However, the efficacy of such systems is limited by the fact that the terms used in a query may be in a syntactic form different to that of the same words in a document. Consider, for example, the following query and sentences:

- Who is the inventor of a paper?
- S1: C is the inventor of knives
- S2: a devised paper in China...

When analyzing this query, the traditional QA system would classify the sample query into “NAME” as a subcategory of “PERSON”, and then keywords such as “inventor” and “paper” would be extracted. In this example, however, S1 contains the keyword “inventor” and S2 contains the keyword “paper”, and hence their relevance degrees for the query will be the same. Moreover, even if we expand the keywords to “inventor”, “discoverer”, and “paper”, the ranking of the sample sentences will remain unchanged because the term “devise” in S2 belongs to a syntactic category different to that of “inventor” in the query. However, if we were to expand the keyword “inventor” to include related words such as “discoverer”, “devise”, “invent”, “develop”, and “creator”, then we could represent the same concept over a range of syntactic and semantic categories, and thereby reduce the number of answer candidates and extract more exact answers.