Automatic acquisition of
semantic-based question reformulations
for question answering

Jamileh Yousefi and Leila Kosseim

CLaC laboratory
Department of Computer Science and Software Engineering
1400 de Maisonneuve Blvd. West
Montreal, Quebec, Canada H3G 1M8
j_yousef@cs.concordia.ca, kosseim@cs.concordia.ca

Abstract. In this paper, we present a method for the automatic acquisition of semantic-based reformulations from natural language questions. Our goal is to find useful and generic reformulation patterns, which can be used in our question answering system to find better candidate answers. We used 1343 examples of different types of questions and their corresponding answers from the TREC-8, TREC-9 and TREC-10 collection as training set. The system automatically extracts patterns from sentences retrieved from the Web based on syntactic tags and the semantic relations holding between the main arguments of the question and answer as defined in WordNet. Each extracted pattern is then assigned a weight according to its length, the distance between keywords, the answer sub-phrase score, and the level of semantic similarity between the extracted sentence and the question. The system differs from most other reformulation learning systems in its emphasis on semantic features. To evaluate the generated patterns, we used our own Web QA system and compared its results with manually created patterns and automatically generated ones. The evaluation on about 500 questions from TREC-11 shows comparable results in precision and MRR scores. Hence, no loss of quality was experienced, but no manual work is now necessary.

1 Introduction

Question reformulation deals with identifying possible forms of expressing answers given a natural language question. These reformulations can be used in a QA system to retrieve answers in a large document collection. For example given the question *What is another name for the North Star?*, a reformulation-based QA system will search for formulations like *<NP>, another name for the North Star* or *<NP> is another name for the North Star* in the document collection and will instantiate <NP> with the matching noun phrase. The ideal reformulation should not retrieve incorrect answers but should also identify many candidate answers.