Learning the Query Generation Patterns

Marcin Skowron and Kenji Araki

Graduate School of Information Science and Technology, Hokkaido University
Kita-ku, Kita 14, Nishi 9
Sapporo, Japan 060–0814
{ms,araki}@media.eng.hokudai.ac.jp

Abstract. With the current method of query formation, a Question Answering system retrieves a set of documents that are similar to a question, while what is mostly required is a set where an answer occurs frequently. This paper addresses this problem by presenting the Query Generation Pattern method. The aim of the method is to automatically learn an optimal combination, modifications of question words and possible extension of a query with non-question words, which for a given question category and syntax, form reliable queries that retrieve an answer-rich set of documents.

1 Introduction

In a Question Answering (QA) system, locating answers requires text analysis at a level of details that cannot be performed at a satisfactory retrieval time for large text collections[1]. As a result, most of the current QA systems employ a two-stage approach. The aim of the first stage is to select a set of documents relevant to a query from the whole document collection. In the second stage, a detailed analysis of a selected set is performed to find answers. Although, the performance of the second stage, and consequently of a whole QA system depends heavily on the quality of documents retrieved in the first stage, to date the research in this area drew relatively little attention.

The commonly used keyword based and similar approaches to a query formation do not retrieve an optimal set of documents. With this approach, a QA system retrieves a set of documents that are similar to a question, while what a user requests is an answer. In order to provide an answer, a QA system needs to retrieve a set of documents, where such answers occur frequently. In our opinion, for a given question category and question syntax, patterns that transform a given question into a reliable query can be automatically learned in a training process using question-answer pairs. Below, we introduce the idea and learning process of the Query Generation Pattern method. The preliminary test demonstrated a significant improvement in the results, compared to the commonly used keyword based and similar methods of query formation.