Using Terminology and a Concept Hierarchy for Restricted-Domain Question-Answering

Hai Doan-Nguyen and Leila Kosseim

Department of Computer Science and Software Engineering
Concordia University
Montreal, Quebec, H3G 1M8, Canada
`haidoan@cse.concordia.ca, kosseim@cse.concordia.ca`

Abstract. This paper discusses an important characteristic of restricted-domain question-answering (RDQA): the issue of low precision in finding good answers. We propose methods for improving precision using domain-specific terminology and concept hierarchy to rearrange the candidate list and to better characterize the question-document relevance relationship. Once this relationship has been well established, one can expect to obtain a small set of (almost) all relevant documents for a given question, and use this to guide the information retrieval engine in a two-level search strategy. The methods proposed here have been applied to a real QA system for a large telecommunication company, yielding significant improvements in precision.

1   Introduction

This paper presents our research in the development of a question-answering (QA) system for a restricted domain. The system's goal is to reply to customer's questions on services offered by Bell Canada, a major Canadian telecommunication corporation. Although grounded within a specific context, we try to reveal general problems and develop a general methodology for restricted-domain QA (RDQA).

Although work in RDQA dates back to the early years of Artificial Intelligence, the domain has only recently regained interest in the research community. RDQA performs QA on a specific domain and often uses document collections restricted in subject and volume. Often integrated in real-world applications, RDQA, especially systems working on free text corpora (rather than on structured databases), provides many interesting challenges to natural language engineering. Real questions have no limit on form, style, category, and complexity. In addition, a RDQA system often has to deal with the problem of low precision in finding a correct answer for a given question.

In this paper, we discuss the main characteristics of RDQA, and present a series of methods to improve the precision performance of our system. These methods were not developed specifically for our project at hand but always considered in a general perspective of RDQA.