Fuzzy Constraint Based Answer Validation

Steven Schockaert, Martine De Cock, and Etienne E. Kerre

Ghent University, Department of Applied Mathematics and Computer Science,
Fuzziness and Uncertainty Research Unit,
Krijgslaan 281 - S9, B-9000 Gent, Belgium
{Steven.Schockaert, Martine.DeCock, Etienne.Kerre}@ugent.be

Abstract. Answer validation is an important component of any question answering system. In this paper we show how the formalism of prioritized fuzzy constraint satisfaction allows to unify and generalize some common validation strategies. Moreover, answer candidates are represented by fuzzy sets, which allows to handle imprecise answers.

1 Introduction

Question answering systems try to improve the functionality of search engines by providing an exact answer to a user's question, rather than a list of documents. A typical question answering system consists of a question analysis module, a search engine, an answer extraction module and an answer validation module. At least two fundamentally different ways to handle answer validation are used by current systems. Corpus–based methods (e.g. [5]) rely on a deep linguistic analysis of the question and the answer candidates, while redundancy–based methods (e.g. [2],[3],[6]) rely on the massive amount of information available on the web. This paper will focus on the latter kind of methods.

Since it is reasonable to assume that on the web, the answer to most questions is stated in a lot of documents, we can assume that there will be documents in which the answer is formulated in a simple way. As a consequence, simple answer extraction algorithms often suffice. However, simplicity comes with a price; a lot of web pages contain incorrect information, so the answer validation process used in corpus–based methods is not appropriate. Most redundancy–based methods apply some kind of voting: the answer which occurs most often is considered the most likely answer to be correct. This approach has the disadvantage of favouring short, unspecific, answers (e.g. "1928" over "July 26, 1928"). Some systems (e.g. [2],[6]) therefore apply heuristics to boost the scores of specific answers. These heuristics would treat an occurrence of "1928" as evidence for "July 26, 1928" which, in our opinion, is not a fully satisfactory approach.

In this paper we propose an alternative voting scheme, which separates positive and negative information about the feasibility of the answer candidates. To this end, we represent answer candidates as fuzzy sets and define a degree of inconsistency and a degree of inclusion between answer candidates. We show how this scheme can be further refined by asking additional questions and enforcing fuzzy constraints on the results.

P.S. Szczepaniak et al. (Eds.): AWIC 2005, LNAI 3528, pp. 394–400, 2005.