Detecting Expected Answer Relations through Textual Entailment

Matteo Negri, Milen Kouylekov, Bernardo Magnini

Fondazione Bruno Kessler
Via Sommarive, 18 - Povo, Trento, Italy
{negri,kouylekov,magnini}@fbk.eu

Abstract. This paper presents a novel approach to Question Answering over structured data, which is based on Textual Entailment recognition. The main idea is that the QA problem can be recast as an entailment problem, where the text (T) is the question and the hypothesis (H) is a relational pattern, which is associated to "instructions" for retrieving the answer to the question. In this framework, given a question Q and a set of answer patterns P, the basic operation is to select those patterns in P that are entailed by Q. We report on a number of experiments which show the great potentialities of the proposed approach.

1 Introduction

Question Answering (QA) over structured data has been traditionally addressed through a deep analysis of the question in order to reconstruct its logical form, which is then translated in the query language of the target data ([1], [2]). This approach implies a complex mapping between linguistic objects (*e.g.* lexical items, syntactic structures) and data objects (*e.g.* concepts and relations in a knowledge base). Several experiences, however, have shown that such a mapping requires intensive manual work, which represents a bottleneck in the realization of large scale and portable natural language interfaces to structured data.

More recently, Textual Entailment (TE) has been proposed as a unifying framework for applied semantics ([3]), where the need for an explicit representation of a mapping between linguistic objects and data objects can be, at least partially, bypassed through the definition of semantic inferences at the textual level. In this framework, a text (T) is said to entail a hypothesis (H) if the meaning of H can be derived from the meaning of T.

According to the TE framework, in this paper we propose that QA can be approached as an entailment problem, where the text (T) is the question, and the hypothesis (H) is a relational pattern, which is associated to instructions for retrieving the answer to the question. In this framework, given a question Q and a set of relational patterns P=$\{p_1, ..., p_n\}$, the basic operation is to select those patterns in P that are entailed by Q. Instructions associated to patters may be viewed as high precision procedures for answer extraction, which are dependent on the specific data source accessed for answer extraction. In case of QA over