

On Ontology Based Abduction For Text Interpretation

Irma Sofia Espinosa Peraldi, Atila Kaya, Sylvia Melzer, Ralf Möller

Hamburg University of Technology,
Institute for Software Systems,
Hamburg, Germany
sofia.espinosa@tuhh.de, at.kaya@tuhh.de,
sylvia.melzer@tuhh.de, r.f.moeller@tuhh.de

Abstract. Text interpretation can be considered as the process of extracting deep-level semantics from unstructured text documents. Deep-level semantics represent abstract index structures that enhance the precision and recall of information retrieval tasks. In this work we discuss the use of ontologies as valuable assets to support the extraction of deep-level semantics in the context of a generic architecture for text interpretation.

1 Introduction

The growing amount of unstructured electronic documents is a problem found in proprietary as well as in public repositories. In this context, the web is a representative example where the need of logic-based information retrieval (IR) to enhance precision and recall is evident. Logic-based IR means the retrieval of unstructured documents with the use of abstract terms that are not directly readable from the surface of the text, but only between its lines. For example, *Chocolate Cake Recipe* is an abstract term for the following text:

Yield: 10 Servings, 5 oz. semisweet chocolate (chopped), 3 oz. unsweetened chocolate (chopped), 1/4 lb. (8 Tbs.) unsalted butter, 1/4 cup all-purpose flour, 4 eggs at room temperature,

Relational index structures are crucial for IR. Therefore, the task of defining the necessary index structures for abstract terms to allow logic-based IR is unavoidable. In our work, the necessary structures for logic-based IR are called *deep-level semantics* and the process of extracting deep-level semantics from unstructured text documents is understood as *text interpretation*. In the course of the work presented here, we will highlight that a feasible architecture (see Figure 1) to enable the automatic extraction of deep-level semantics from large-scale corpora can be achieved through:

- A two phase process of information extraction (IE), where the first phase exploits state-of-the-art shallow text processing mechanisms to extract surface-level structures as input for the second phase. The second phase called deep-level interpretation, exploits reasoning techniques over ontologies to extract deep-level semantics.