Chapter 2
An Overview of Shallow and Deep Natural Language Processing for Ontology Learning

Amal Zouaq
Simon Fraser University - Athabasca University, Canada

ABSTRACT
This chapter gives an overview over the state-of-the-art in natural language processing for ontology learning. It presents two main NLP techniques for knowledge extraction from text, namely shallow techniques and deep techniques, and explains their usefulness for each step of the ontology learning process. The chapter also advocates the interest of deeper semantic analysis methods for ontology learning. In fact, there have been very few attempts to create ontologies using deep NLP. After a brief introduction to the main semantic analysis approaches, the chapter focuses on lexico-syntactic patterns based on dependency grammars and explains how these patterns can be considered as a step towards deeper semantic analysis. Finally, the chapter addresses the “ontologization” task that is the ability to filter important concepts and relationships among the mass of extracted knowledge.

1. INTRODUCTION
Given the large amount of textual data in almost all the aspects of our everyday lives and the fact that natural language is our primary medium for communicating knowledge, there is no doubt that natural language processing (NLP) technologies are of tremendous importance for analyzing textual resources and extracting their meaning. One of the current research avenues where NLP should play a leading role is the Semantic Web. In fact, NLP should be considered as one of the pillars of the Semantic Web (Wilks & Brewster, 2009) for its role in the acquisition of domain ontologies. Despite the vast majority of works dedicated to ontology learning based on NLP (Buitelaar & Cimiano, 2008) (Cimiano & Volker, 2005) (Buitelaar et al., 2005), it is clear that the whole potential of the available techniques and representations has not been fully exploited. More precisely, the works from the computational semantics community (Bos, 2008c) have been

DOI: 10.4018/978-1-60960-625-1.ch002
largely neglected, to my knowledge, in the ontology learning field until now. Advances from this community are of particular relevance since they address important aspects of text understanding that are not available in more shallow techniques. I believe that these deep aspects are essential for building an accurate domain ontology reflecting the content of its source data. This chapter is a tentative to bring this issue to the attention of the research community. This chapter also provides a quick overview over techniques from the computational semantics community that may be of interest for ontology learning. We will look at some of the shallow NLP techniques as well as deeper methods based on dependency grammars and lexico-syntactic patterns which have already been employed for ontology learning. Many of these deeper methods used for ontology learning have not yet reached the required depth that is advocated in computational semantics. They nevertheless can be considered as early attempts towards this goal.

The chapter is organized as follows. After the introduction, Section 2 provides the definition of ontologies and the ontology learning task. Section 3 discusses the various natural language processing techniques that may be used in an ontology learning process. It also provides a quick overview over the available techniques in the computational linguistic and semantics communities including shallow and deep analysis, both at the syntactic level and the semantic level. Section 4 presents a set of projects for ontology learning with a special emphasis on the dependency grammar formalism and the use of patterns based on this formalism. This section emphasizes the links between the presented projects and the various NLP techniques that were used. Section 5 explains an important stage of the ontology learning process which is the ontologization task. We end this chapter in Section 6 with a discussion on a number of issues that require the attention of the research community.

2. BACKGROUND

There are a number of resources that describe what an ontology is, with the most cited definition being the one presented by (Gruber, 93): “An ontology is a formal specification of a conceptualization”. Although this definition may seem too broad, we can extract from it two keywords that are essential for our understanding of ontologies: formal and conceptualization.

- **The formal characteristic:** In the domain of computer science and formal logic, a formal system designates a system using a **formal language**, a grammar that indicates the well-expressed formulas according to the language and a set of axioms or inference rules to reason over this language. A formal language is defined using a set of symbols.

- **The conceptual characteristic:** Having its root in philosophy, the notion of concept has been widely used in the Artificial Intelligence community. According to (Guarino, 98), a conceptualization must be defined on an intentional level and an extensional level. The intentional level deals with the meaning of what is being defined (the domain of interest), while the extensional level describes the instances of that domain.

As it can be seen, an ontology is grounded in the domain of mathematical logic, reasoning and theorem-proving. In fact, it is the main knowledge structure of the Semantic Web, whose aim is to provide a set of machine understandable semantics. These semantics are generally organized in a structure called a domain ontology, which is used to express the conceptualization of that domain.

Formally, a domain ontology is represented by a tuple \(<C, H, R, A, I>\), where:
Related Content

Genetic Learning: Initialization and Representation Issues
[www.igi-global.com/chapter/genetic-learning-initialization-representation-issues/24215?camid=4v1a](www.igi-global.com/chapter/genetic-learning-initialization-representation-issues/24215?camid=4v1a)

Creation and Integration of Reference Ontologies for Efficient LOD Management
[www.igi-global.com/chapter/creation-integration-reference-ontologies-efficient/63902?camid=4v1a](www.igi-global.com/chapter/creation-integration-reference-ontologies-efficient/63902?camid=4v1a)

Rare Class Association Rule Mining with Multiple Imbalanced Attributes
Huaifeng Zhang, Yanchang Zhao, Longbing Cao, Chengqi Zhang and Hans Bohlscheid (2010). *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection* (pp. 66-75).
[www.igi-global.com/chapter/rare-class-association-rule-mining/36900?camid=4v1a](www.igi-global.com/chapter/rare-class-association-rule-mining/36900?camid=4v1a)

Image Mining for the Construction of Semantic-Inference Rules and for the Development of Automatic Image Diagnosis Systems
Petra Perner (2007). *Knowledge Discovery and Data Mining: Challenges and Realities* (pp. 75-97).
[www.igi-global.com/chapter/image-mining-construction-semantic-inference/24902?camid=4v1a](www.igi-global.com/chapter/image-mining-construction-semantic-inference/24902?camid=4v1a)