Chapter IX
Semantics Enhancing Knowledge Discovery and Ontology Engineering Using Mining Techniques:
A Crossover Review

Elena Irina Neaga
Loughborough University Leicestershire, UK

ABSTRACT

This chapter deals with a roadmap on the bidirectional interaction and support between knowledge discovery (KD) processes and ontology engineering (ONTO) mainly directed to provide refined models using common methodologies. This approach provides a holistic literature review required for the further definition of a comprehensive framework and an associated meta-methodology (KD4ONTO4DM) based on the existing theories, paradigms, and practices regarding knowledge discovery and ontology engineering as well as closely related areas such as knowledge engineering, machine/ontology learning, standardization issues and architectural models. The suggested framework may adhere to the ISO-reference model for open distributed processing and OMG-model-driven architecture, and associated dedicated software architectures should be defined.

INTRODUCTION AND MOTIVATION

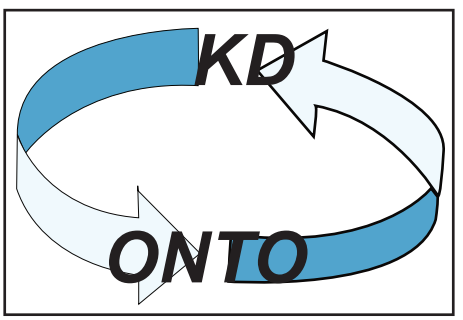
Generally, the role of ontologies in knowledge engineering has been deeply investigated since early 1990 such as the research reported by van Heijst (1995) in his thesis mainly dealing with the modalities of defining explicit ontologies and using them in order to facilitate the knowledge engineering processes. However, these earlier approaches have been achieved in the context of

building knowledge based systems and enhancing the agents' knowledge base. These approaches are not contradictory to the current emergence of Semantic Web technologies using as a core concept the ***ONTOLOGY*** as "*an explicit specification of a conceptualization*" (Gruber, 1994, p. 200). Recently, ontologies have provided the development of common frameworks for dissimilar systems supporting disparate applications related to engineering, business and medicine. Generally, these applications have a high-level common foundation based on generic concepts and theories. Moreover, generic knowledge and ontology models capture the common aspects of different domains.

Currently, the problems related to *ontology engineering* workbenches, methodologies and tools are quite similar of those that knowledge engineers have approached in order to define knowledge bases. Besides the machine learning is a common artificial intelligence technique used in ontology development as well as in knowledge intensive systems including data mining. Although the ontology and knowledge bases can be independently defined, used, processed and maintained there is not a strict separation between them (Maedche, 2002), and the approach enclosed in this chapter emphasizes the similarities and the differences.

There are not many crossover approaches regarding both knowledge discovery process and ontology engineering, but Gottgtroy, Kasabov, and MacDonell (2003, 2004) have explored the bidirectional interaction and support between the related processes. This interaction is graphically represented in Figure 1.

Figure 1. KD4ONTO4KD



The knowledge discovery process includes several phases, such as data preparation, cleaning and transformation, and each of these phases or steps in the life-cycle might benefit from an *ontology-driven approach* which underpins the semantics in order to enhance the knowledge discovery processes. Applying intelligent data analysis, visualization, and mining techniques to discover and identify meaningful relationships, missing and clustering concepts may contribute to refine ontology models and related processes (Gottgtroy et al., 2003, 2004). On the other hand, as well as Maedche (2002) and Staab (2000, 2001) have introduced a novel approach of applying knowledge discovery to multiple data sources to support the development and maintenance of ontologies and techniques for ontology learning mainly from text.

Nowadays, a very important aspect is the explosive growth of information available on the World Wide Web which makes very difficult to deal with information overload and relevance. These issues can be approached using Web personalization based on Web usage mining which applies mining algorithms such as association rules, sequential patterns and clustering in the context of Web mining. De Moor (2005) suggested an approach based on context dependent ontologies by using so-called pragmatic patterns which are defined as meta- patterns including the ontology models and the context description. It then becomes possible to better deal with information overload and relevance as well as partial, contradicting, evolving ontologies and meaning negotiation.

An ontological approach of data mining systems may also provide the interoperability functionalities of knowledge discovery distributed systems. Furthermore given the present lack of a generally accepted framework for knowledge