Chapter 4
Ontology Engineering: The "Whats," "Whys," and "Hows" of Data Exchange

J. M. Easton
University of Birmingham, UK

J. R. Davies
University of Birmingham, UK

C. Roberts
University of Birmingham, UK

ABSTRACT

In a challenging financial climate, there is a growing impetus for businesses to use existing process data to support more intelligent decision making. For large-scale complex systems such as railways, electricity grids, and gas distribution networks, this often means combining information from numerous different condition monitoring systems; however, given the vast amounts of data produced every day and the frequently incompatible data models used to represent it, is it possible to be sure that the information generated is being used correctly? This paper provides an introduction to the field of Ontology, an emerging technology that allows the exact "meaning" of an item of data to be described in a way that can be interpreted by computers. Through this retention of meaning, it becomes possible for computers to perform simple reasoning operations, inferring new information about a system from the existing facts, and enabling exciting new Semantic Web technologies.

BACKGROUND

With the increasing availability of condition monitoring capabilities on industrial assets and the squeezing of resources due to the changing financial climate, it is becoming more and more important to extract a business advantage from stored process data. In many cases, the largest potential benefits could be gained by the combination of datasets from different asset types to gain a more holistic view of the system; however, industrial assets often use proprietary data formats

DOI: 10.4018/978-1-4666-4002-3.ch004

making the integration of their data difficult, and the pedigree of the data itself (including factors such as the techniques used for data collection, and any data processing that has taken place) may mean it is unsuitable for particular tasks. As a result of this, the creation and recording of metadata specifying the exact nature of a dataset is as important as the collection of the raw data itself.

"Ontology is gaining popularity and is touted as an emerging technology that has a huge potential to improve information organisation, management and understanding" (Ding & Foo, 2002a). Despite this, "ontology building is still more of a craft than an engineering task" (Pinto & Martins, 2004), with different practitioners using wildly different methodologies for the gathering of domain knowledge and subsequent modelling tasks. This paper aims to provide an overview of ontology for engineers, describing what ontology is and why it is useful, before going on to introduce key technologies and review the engineering methodologies that can be used in ontology development.

An Introduction to Ontology

Ontologies are often described in textbooks and journal papers as being "an explicit specification of a conceptualisation" (Gruber, 1993). That is somewhat inaccessible, so perhaps a more helpful description is that "ontologies are content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge. They provide potential terms for describing our knowledge about the domain" (Chandrasekaran, Josephson, & Benjamins, 1999). While the second description is far easier to read than the first, it suggests that an ontology is very similar to a controlled vocabulary. Ontologies are far more than that; an ontology allows conceptual information about a domain to be captured and stored in a way that can be interpreted by computers, i.e. it reflects not only the terms that are important in the domain, but also some idea of their relationships and meaning.

The Semantic Web

While ontologies have been part of artificial intelligence and knowledge-based systems research for a number of years, recent interest in the field has been driven by the prospect of the Semantic Web, expected to be the next major extension to the World Wide Web. The World Wide Web uses markup in web pages to present information in a way that is easily recognisable by humans, with features such as tables, lists, headings, and underlining for emphasis. The Semantic Web will take this idea a step further by presenting information in ways that can be understood by both humans and machines, paving the way for autonomous pieces of software called software agents to do much of the work of searching and compiling data for us.

In an influential article from 2001, Berners-Lee, Hendler, and Lassila (2001) illustrated one of the potential uses of the Semantic Web by describing a fictional booking of a series of medical appointments. In the scenario they presented, the authors described a software agent on a handheld device that when instructed to book a series of appointments, could contact the doctor's office to identify the treatment prescribed for the patient. The software agent then went on to search for local providers of that treatment, check the patient's health insurance cover, match available appointment times to gaps in the patient's schedule, and present a selection of the best options to the user. Berners-Lee, Hendler, and Lassila's example is pretty stunning, but is actually only the tip of the iceberg when you consider the possibilities for automated data mining and knowledge discovery that the Semantic Web vision presents. It's also very important to industry, as the same technologies could be used securely with other information systems, such as a network of databases distributed across an organisation's sites.

In order to realise the true potential of the Semantic Web, two key components must be in place; firstly, data on the Semantic Web must be expressed in a way that can be understood by any