Chapter XXII
A Language and Algorithm for Automatic Merging of Ontologies

Alma-Delia Cuevas-Rasgado
Instituto Politécnico Nacional, Mexico

Adolfo Guzman-Arenas
Instituto Politécnico Nacional, Mexico

ABSTRACT

Ontologies are becoming important repositories of information useful for business transactions and operations since they are amenable to knowledge processing using artificial intelligence techniques. They offer the potential of amassing large contents of relevant information, but until now the fusion or merging of ontologies, needed for knowledge buildup and its exploitation by machine, was done manually or through computer-aided ontology editors. Thus, attaining large ontologies was expensive and slow. This chapter offers a new, automatic method of joining two ontologies to obtain a third one. The method works well in spite of inconsistencies, redundancies, and different granularity of information.

INTRODUCTION

Computers are no longer isolated devices but they are important to the world-wide network that interchanges knowledge for business transactions. Nowadays, using the Internet to get data, information, and knowledge is a business need.

Most of the important information resources that businessmen require are available through the Internet. Here, machines face the problem of heterogeneous sources. The computer has a hard time finding whether two data representations refer to the same object (a *bill* can be a bank tender or an invoice)[1] because there are no suitable standards in knowledge representation. This chapter addresses this need of businesses and academia.

When businessmen demand answers that require access to several Internet data sources, they have to manually or mentally merge the acquired information in a reasonable way. It would be nice if a computer program helped in this very useful but tedious task. This chapter solves this problem, which has important implications (see the section on "Commercial Areas Ready to Exploit OM").

The Problem to Solve[2]

To merge two data sources in such a way that its common knowledge could be represented and more easily used in further tasks.

Computers represent the information in files, databases, text documents, lists, and so forth. Computer merging of information in databases or in semistructured data, has its own challenges, and will not be addressed here. Merging information stored in documents is done manually, since the computer does not "understand" what a document says. If the information is stored in spreadsheets, merging can be done by a computer-aided person who understands the contents of different cells and their units. Information can also be stored in ontologies and thus be subject to merging. So far, merging of ontologies has been done manually (see the section on "Ontology Merging") using an ontology editor.

Ontology

An ontology is a data structure where information is stored as nodes (representing concepts such as `hammer`, `printer`, `document`, appearing in this chapter in `Courier font`) and relations (representing restrictions among nodes, such as cuts, transcribes, or hair color, appearing in this chapter in Arial Narrow font, as in (`hammer` cuts `wood`), (`printer` transcribes `document`), Figure 9. Usually, the information it stores is "high level" and it is known as *knowledge.* For working purposes, we further restrict this definition to those data structures compliant with ontology merging (OM) notation (*quo vide*).

Ontologies are useful when arbitrary relations need to be represented; one has more freedom to represent different types of concepts.

Current notations to represent ontologies are DAML+OIL (Connoly et al., 2001), RDF (Manola & Miller, 2004; Asunción & Suárez, 2004) and OWL (Bechnofer et al., 2004). These languages are a notable accomplishment, but some lack certain features:

- A relation can not be a concept. For instance, if color is a relation, it is difficult to relate color to other concepts (such as shape) by using other relations.
- Partitions (subsets with additional properties, see the section on "Contributions of OM Notation") can not be represented.

This chapter offers the *OM notation* to represent ontologies that solves above problems and better represents the semantics involved.

Ontology Merging

Realizing the importance of the problem to solve, different scientists have approached it. Previous works incudes CYC (Lenat & Guha, 1989), whose goal was to represent common sense knowledge in a gigantic hand-built ontology. CYC does not do merging. Prompt (Noy & Musen, 2000), Chimaera (McGuinness, Fikes, Rice, & Wilder, 2000), OntoMerge (Stumme & Maedche, 2001) and ISI (Loom) rely on the user to solve the most important problems found in the process, and are considered non automatic methods. FCA-Merge (Dou, McDermott, & Qi, 2002) and IF-Map (Kalfoglou & Schorlemmer, 2002) require consistent ontologies that are expressed in a formal notation employed in Formal Concept Analysis (Ganter, Stumme, & Wille, 2005) which limits their use. Hcone (Kotis, Vouros, & Stergiou, 2006) uses WordNet and a formal approach to ontology merging. Cuevas-Rasgado (2006) mentions additional previous works.