# Chapter 5.9
# Virtual Communities and the Alignment of Web Ontologies

**Krzysztof Juszczyszyn**
*Wroclaw University of Technology, Poland*

## EMERGING SEMANTIC WEB

The World Wide Web (WWW) is a global, ubiquitous, and fundamentally dynamic environment for information exchange and processing. By connecting vast numbers of individuals, the Web enables creation of virtual communities, and during the last 10 years, became a universal collaboration infrastructure.

The so-called Semantic Web, a concept proposed by Tim Berners-Lee, is a new WWW architecture that enhances content with formal semantics (Berners-Lee, Hendler, & Lassila, 2001). Hence, the Web content is made suitable for machine processing (i.e., it is described by the associated metadata), as opposed to HTML documents available only for human consumption. Languages such as Resource Description Framework (RDF) and Ontology Web Language (OWL) along with well-known XML are used for description of Web resources. In other words, the Semantic Web is a vision of the future Web in which information is given explicit meaning. This will enable autonomous software agents to reason about Web content and produce intelligent responses to events (Staab, 2002).

The ultimate goal of the next generation's Web is to support the creation of virtual communities which will be composed of software agents and humans cooperating within the same environment. Sharing knowledge within such a community requires a shared conceptual vocabularies—ontologies, which represent the formal common agreement about the meaning of data (Gomez-Perez & Corcho, 2002). Artificial intelligence defines ontologies as explicit, formal specification of a shared conceptualization (Studer, Benjamins, & Fensel, 1998). In this case, a *conceptualization* stands for an abstract model of some concept from the real world; *explicit* means that the type of concept used is explicitly defined. *Formal* refers to the fact that an ontology should be machine readable; and finally *shared* means that ontology expresses knowledge that is accepted by all the subjects. In short, an ontology defines the terms used to describe and represent an area of knowledge.

However, the shared ontologies must be first constructed by using information from many sources which may be of arbitrary quality. Thus, it is necessary to find a way to seamlessly combine the knowledge from many sources, maybe diverse and heterogeneous. The resultant ontologies enable virtual communities and teams to manage and exchange their knowledge.

It should be noted, that the word *ontology* has been used to describe notions with different degrees of structure—from taxonomies (e.g., Yahoo hierarchy), metadata schemes (e.g., Dublin Core), to logical theories. The Semantic Web needs ontologies with a significant degree of structure. These should allow the specification of at least the following kinds of things:

- Concepts (which identify the classes of things like *cars* or *birds*) from many domains of interest
- The relationships that can exist among concepts
- The properties (or attributes) those concepts may have

## ONTOLOGY ALIGNMENT

The straightforward way to achieve semantic interoperability is to provide a global ontology used by all the agents. A WordNet database is a good example (Fellbaum, 1998). Unfortunately, many experiments have shown that upper-level ontologies (analogous to language conventions) can be hardly applied in dynamic communities; moreover, modeling large domains is time-consuming, difficult, and expensive (Bailin & Truszkowski, 2001). On the other hand, interaction between agents from diverse communities cultivating their domain ontologies seems inevitable. Integration of virtual communities and organizations seems also impossible without integrating their knowledge, in the form of Web ontologies.

In the absence of domain ontology, software agents acting within the Semantic Web environment must relate concepts that are semantically close or identical (via equivalence or subsumption relations) to achieve mutual understanding of processed data. The operation of identifying such concepts is called *ontology alignment*. Ontology alignment is a mapping between concepts defined in a source ontology and concepts defined in a target ontology. To align ontologies, one must specify the concept from the target ontology that represents as closely as possible the meaning of the source ontology concept. For each concept in the source ontology, we try to identify a corresponding concept in the target ontology; however, it may be impossible for all concepts (Klein, 2001; McGuinness, Fikes, Rice, & Wilder, 2000). Then the corresponding concepts may be mapped onto each other, allowing communication between agents using source and target ontologies (they reach semantic agreement about the meaning of the given concepts). A review of ontology-mapping methods was recently presented by Kalfoglou and Schorlemmer (2003).

Because ontologies are developed and managed independently the semantic mismatches between two or more ontologies are inevitable. Practical applications show that fully shared vocabularies are rather exceptional—a number of possible different semantic conflicts was identified by Shaw and Gaines (1989); other classifications were proposed by Hameed et al. (2001). Most of ambiguities emerge during agreement upon concepts and terms, when

- the same term (homonym) is used for different concepts;
- different terms (synonyms) are used for the same concept;
- different ontology representation languages were used;
- given concept was represented in different ontologies at different levels of detail; and