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
- given domain was modeled in different ways.

The vision of Semantic Web allowing agents to publish and exchange ontologies requires strong mechanisms supporting ontology merging and alignment (Hendler, 2001). Without them, it may be almost impossible to achieve the semantic interoperability in societies of autonomous Web agents.

However, already developed methods for ontology alignment do not guarantee success. In particular, most of the methods used offer semiautomatic approach which require human assistance (Noy & Musen, 2000; Silva & Rocha, 2003). However, in real-live situations an automatic cooperation between autonomous agents is expected. Moreover, it is hardly possible that all of the agents will use the same method for aligning their ontologies, and this has remarkable consequences on knowledge sharing in Semantic Web. General approach to ontology alignment problem assumes the use of so-called similarity measures which relate concepts from different ontologies. The similarity measure is a function or procedure that analyzes the structure of ontologies to evaluate the level of correspondence between two given concepts (Andrea & Egenhofer, 2003; Lin, 1998; Maedche & Zacharias, 2002; Stuckenschmidt & Timm, 2002).

The similarity functions developed use syntactical, lexical, or structural analysis. Many other techniques were also proposed, for example, Lin (1998) presents an information-theoretic definition of similarity and proposes a measure for computing string and taxonomical similarities. The proposed measure is next compared with several other approaches. Lin's experiments showed that for different similarity measures applied to chosen taxonomy, similarity between certain concepts varies significantly, the same concerns string similarity between chosen pairs of words. Similarity measures may give different results for the same input data. Many require human-controlled tuning of parameters (Noy & Musen, 2000) or additional information (Andrea & Egenhofer, 2003), such as noun frequencies in language corpus or synonyms sets (Resnik, 1999)—in this case, one can never guarantee that such information will be always available.

The issues described create the possibility of a new type of mismatches because it is not straightforward that two agents aligning their ontologies will obtain the same result, which depends on the method used. There is a need of a framework that allows the use of different methods and deals with the uncertainty resulting from differences between similarity measures exploited by the agents. Proposed solutions tend to enforce negotiations between the agents until they reach agreement (or not) over given concepts (Bailin & Truszkowski, 2001) or use a formal mathematical framework to evaluate the uncertainty resulting from the difference in the similarity measures used (Juszczyszyn, 2004).

Although many successful experiments were carried out, the problem of automated ontology alignment is still a research challenge and no ultimate solution has been presented.

CONCLUSION

Emerging Semantic Web offers vast possibilities of knowledge acquisition and sharing across virtual communities. The need for shared ontologies is particularly recognized within area of electronic commerce, virtual organizations, and scientific research. However, the scale of the Web along with the growing need for automation of knowledge management processes imposes serious research challenges. The reliable methods for alignment of Web ontologies are the key enablers for new generation of virtual communities which will use advanced knowledge management techniques to communicate people and software, search for information, and provide Web services on demand.

REFERENCES

Andrea, M., & Egenhofer, M. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, *15*, 442-456.