SemWeb: integrated access to distributed ontological resources

Summary

We propose to develop a system, dubbed SemWeb, that would revolutionize the way people — from experts to students — interact with conceptual structures and terminology and the way they share such knowledge. We aim at the synergistic exploitation of existing lexical and ontological knowledge bases (ontologies/classifications, taxonomies, thesauri, dictionaries) and their vast intellectual capital through integrated access, allowing a user to consult multiple sources with one search returning one integrated answer that visualizes concept relationships for ease of understanding. SemWeb is intended for a wide variety of users and uses — including education, information retrieval, knowledge-based systems, and natural language processing — and bridge discipline, languages, and cultures. The same environment will support collaborative development and maintenance of ontologies and lexica. We envision knowledge integration through two approaches working together: (1) automating knowledge integration to the extent feasible and (2) providing tools to facilitate knowledge integration by users for their own purposes, but in a format that can be shared with others.

We expect that such a system would have a significant impact on how people think and structure problems; how they retrieve and organize information; how they communicate and collaborate across disciplines, languages, and cultures; and how they learn. By making vast amounts of knowledge accessible, it would also expand the horizons on what can be done in knowledge-based systems and natural language processing. All of this contributes to scientific productivity.

We propose to build such a system, which could then be deployed in many contexts, and install a demonstration prototype Web server. At the core of the system is the SemWeb template for the arrangement of any kind of information about concepts and terms. The template, to be developed collaboratively, brings together existing standards for machine-readable dictionaries, ontologies, subject authority files, and classification data and results from a user needs analysis. It is the basis for the user interfaces, for linking resources into a SemWeb system, and for collaborative development. The system includes a resource inventory, which mediates access to ontological and lexical knowledge bases. It is structured following a metadata schema for resource description to be developed. The SemWeb software will consist of a kernel for resource selection and data integration, plug-ins for accessing individual resources, and a tool for collaborative development of ontologies/classifications, concept maps, thesauri, and dictionaries; this tool will have an interface usable by students at all levels.

We will do research on difficult issues that need to be addressed in the system: for example we will study how ontological and lexical knowledge is used in different disciplines, and we will work on defining measures and methods for the evaluation of ontologies, lexica, and their representations and for correlating and integrating ontologies. We will also study the use and impact of the prototype through its use in a pilot NLP application, through server use and user studies, and through observing the impact on learning by students at all levels.