Protégé-2000: A Plug-in Architecture to Support Knowledge Acquisition, Knowledge Visualization, and the Semantic Web

Mark A. Musen, M.D., Ph.D., Ray W. Fergerson, Ph.D, Natalya Fridman Noy, Ph.D., and Monica Crubézy, Ph.D.

Stanford Medical Informatics, Stanford University School of Medicine
Stanford, California 94305-5479

Background. Protégé-2000 is the latest in a series of tools developed in our laboratory to assist developers in the construction of large electronic knowledge bases [1]. The direct-manipulation user interface allows developers to create and edit domain ontologies that represent the salient concepts and relationships among concepts in an application area. From the ontology, the system automatically constructs a graphical knowledge-acquisition system that allows application specialists to enter the content knowledge required for specific applications.

Protégé-2000 is written entirely in the Java programming language, and thus runs under a wide variety of operating systems. Protégé-2000 is component-based, which makes it straightforward for programmers to extend the system by building plug-ins that offer enhanced or applicationspecific functionality. For example, plug-ins allow visualization of knowledge bases and ontologies in two or three dimensions, facilitating browsing and navigation of complex knowledge structures.

System Description. The Protégé-2000 system presents the user with a series of “tabs,” where each tab offers the user access to a different element of the system’s functionality. Standard tabs allow users to edit and browse a domain ontology, to custom-tailor the knowledge-acquisition tool generated from a domain ontology, to enter knowledge into the knowledge-acquisition tool, and to search the knowledge base.

An essential goal of the system is to make knowledge browsing and entry as simple as possible. When the system generates a knowledge-acquisition tool from an ontology, users enter domain information by filling in the blanks of intuitive forms, selecting items from lists, and by drawing diagrams composed of selectable icons and connectors.

Our research group has developed several additional “tabs” that offer Protégé-2000 enhanced functionality. These tabs include a client that can access the National Library of Medicine’s Unified Medical Language System over the Internet and can incorporate concepts from the metathesaurus directly into a domain ontology. Other tabs allow Protégé-2000 to communicate with knowledge servers that comply with the Open Knowledge Base Connectivity (OKBC) protocol and with documents written in XML or RDF.

Protégé-2000 allows knowledge bases to be stored in several formats. Standard options include flat files that use either RDF or a syntax based on NASA’s CLIPS knowledge representation system. It is also possible for Protégé-2000 to work directly with relational databases.

Evaluation. More than 1000 groups have downloaded Protégé-2000 and are using the system to build a wide variety of applications. At Stanford, we rely on Protégé-2000 to build all our intelligent systems. For example, the EON system for automation of protocol-based care relies on Protégé-2000 [2]. Formal studies have measured the advantage of Protégé’s domain-specific user interfaces for knowledge acquisition [3].

Conclusion. Protégé-2000 offers a sophisticated and easy to use development environment for the rapid construction of intelligent systems. Because Protégé-2000 is component-based, the entire Protégé-2000 user community is able to contribute new plug-ins that collectively enhance the system’s capabilities.

Additional Information. Protégé-2000 may be freely downloaded from our Web site under an open-source license (see http://protege.stanford.edu).

References