Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment

AMIA 2003 Open Source Expo

Natalya F. Noy, PhD, Monica Crubézy, PhD, Ray W. Fergerson, PhD, Holger Knublauch, PhD, Samson W. Tu, MS, Jennifer Vendetti, BS Mark A. Musen, MD, PhD

Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA 94305

http://protege.stanford.edu

Abstract Protégé-2000 is an open-source tool that assists users in the construction of large electronic knowledge bases. It has an intuitive user interface that enables developers to create and edit domain ontologies. Numerous plugins provide alternative visualization mechanisms, enable management of multiple ontologies, allow the use of inference engines and problem solvers with Protégé ontologies, and provide other functionality. The Protégé user community has more than 7000 members.

Background Protégé-2000 is the latest in a series of tools developed in our laboratory to assist users 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 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 in Java, and thus runs under a wide variety of operating systems. Protégé-2000 is available under the open-source license and can be downloaded from http://protege.stanford.edu.

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 for users 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.

Furthermore, perhaps, one of the biggest values of the Protégé system is the availability of dozens of **plugins**[2] (developed by the Protégé group and by contributors from all over the world). These plugins provide alternative visualization mechanisms, enable management of multiple ontologies, including merging and version management, allow the use of various inference engines and problem solvers with Protégé ontologies, and provide other functionalities.

Protégé provides a Java API for application developers to access and modify all aspects of Protégé knowledge bases and its user interface.

Protégé stores ontologies in many different formats including relational databases, UML, XML, and RDF. We are currently working on providing support for the OWL language, which is a language designed for the next generation of the World-Wide Web—the Semantic Web.[3]

User community Protégé has a wide and active user community, which has more than 7000 registered users, and continues to grow rapidly. There is an active discussion list with more than 1200 subscribers. Because Protégé-2000 is a component-based system, the entire Protégé user community is able to contribute new plugins that collectively enhance the system's capability. Currently the Protégé plugin library contains around 40 plugins and applications. Many of these plugins are also available under an open-source license (see http://protege.stanford.edu/plugins).

Funding Protégé was developed with support from federal agencies (Defense Advanced Research Projects Agency, National Cancer Institute, National Institute of Standards and Technology, National Library of Medicine, National Science Foundation) and private companies.

References

[1] J. Gennari et al. The evolution of Protégé: An environment for knowledge-based systems development. *Int. Journal of Human-Computer Interaction*, 58(1), 2003.

[2] M. A. Musen et al. Component-based support for building knowledge-acquisition systems. In *IFIP WCC Conference on Intelligent Information Processing*, Beijing, China, 2000.

[3] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May, 2001.