

A Common Terminology Services (CTS) Back-end to Protégé

Harold R Solbrig

Chrisopher G Chute, M.D., Dr. P.H.

Division of Biomedical Informatics

Mayo Clinic

Coding and classification schemes have been used by the medical community to classify clinical events for several centuries. Recently, the medical community has focused on the ability to share the content of clinical records across institutions and between healthcare providers and clinical researchers. Part of this effort has involved the creation of various “ontological” or “terminological” resources in the clinical domain, including the Unified Medical Language System (UMLS) – a thesaurus that cross-references medical coding and classifications schemes, the Medical Subject Headings (MeSH) – a hierarchical “controlled vocabulary thesaurus” used to index articles from over 4,600 different biomedical journals and the Systemized Nomenclature of Medicine Clinical Terms – a comprehensive “reference ontology” of medicine.

This rich set of terminological resources needs to be accessible to future development efforts, and any new ontology or information model that overlaps the biomedical domain needs to be able to draw upon this existing body. The obvious reason to do this is that it reduces the need to duplicate effort recreating identical or similar things. Another important reason, however, is the use of these terminologies makes it possible to draw on the vast body of already existing clinical information to answer new research questions.

If these resources are to be used in new bioontology projects, they need to:

- Share a common model
- Be readily searched and navigated
- Be downloadable and accessible in a variety of well-understood and useful formats.

The Division of Biomedical Informatics at the Mayo Clinic has been actively pursuing this goal of shared and integrated clinical terminology content – the lexical grid (LexGrid). The Lexical Grid, built on an LDAP backbone, publishes clinical terminologies in a common, shared model. The Lexical Grid project includes search and navigation tools, authoring tools and a variety of services that allow content to be used directly online or to be downloaded in different formats for localized integration.

One of the services that available through the Lexical Grid is an implementation of the Common Terminology Services (CTS) standard. CTS is an interface specification that was developed by Health Level Seven (HL7), an ANSI accredited standards developer, for the purpose of querying and accessing terminological content. The CTS services have been used to create new back end for Protégé that allows Protégé to import terminology content in a format familiar to Protégé users.

The CTS Backend provides the ability to:

- Render any CTS accessible terminology in a native Protégé format. Since all of the LexGrid resources are available via CTS, this means that any LexGrid resource, such as ICD-9, MeSH, SNOMED-CT, NCI Thesaurus, LOINC, NDF-RT, ISO Country Codes, ISO Language codes, UNSPSC, etc. are directly available in Protégé.
- Include to all or part of single or multiple terminologies.
- Include content either by reference, where the actual content is maintained on the web and retrieved dynamically – or by value, where a copy of the web content is made available locally.
- Have classes, slots and facets in a Protégé model to reference a terminological resource. This reference can either be by name if the actual terminological content is not currently available or by value if it is. This ability allows new ontologies to selectively include resources only when and where they are needed.
- Locally extend and augment existing terminology resources – this allows development to continue even though the source terminology may be ambiguous or incomplete. Obviously, the authors would want to push the local changes back to the source terminology stewards for inclusion in future releases.
- Publish Protégé-based terminological content as a CTS service – Terminologies created directly in terminology can be published as a CTS service and be made available to other CTS clients.

The Common Terminology Services defines a common model for representing terminology content. This model is represented by the CTS Backend in Protégé via the following classes.

- Code System – a cohesive collection of concept codes managed in a single namespace
- Coded Concept – a unique code that represents a class or concept within the code system namespace, along with various designations, descriptions, comments and other properties that help to identify the concept meaning
- Designation – represents the concept in at least one universe of discourse
- Definition – describes or defines the concept
- Annotation – used to add meta-information about the history or use of the concept

The CTS Backend comes with a CTSBackend Tab that allows Protégé users to connect and log on to CTS Services and select the terminological resources to be imported.