

The Application of Terminology Services across an Academic Medical Center: The LexGrid Project

Christopher G. Chute, MD DrPH
Biomedical Informatics, Mayo Clinic College of Medicine

The Lexical Grid (<http://informatics.Mayo.edu>) is a comprehensive model of terminologies that can accommodate simple term lists, controlled content, complex vocabularies, and description-logic-based ontologies into a common model. Associated with this content coordination is a suite of services to enable access to terminological content as a data-grid resource on the network. Access methods have been written in Java, SOAP, COM, and other common mechanisms to permit terminological access within applications, across related projects, and most pertinently as a shared and scalable resource within a complex information enterprise.

The LexGrid project has been adopted by the NCI's caBIG initiative to provide the infrastructural framework for terminology access. Similarly, it underpins the HL7 terminology content model in close association with RIM artifacts. Additionally, the CDC has recently adopted the LexGrid environment to be the basis of its PHIN vocabulary services. The LexGrid project is a fully open-source, collaborative engagement, components of which are now ANSI and ISO standards (specifically, the common terminology services (CTS) component).

The LexGrid is also a core technology in the NIH roadmap/BISTI project which forms the National Center for Biomedical Ontology, led by Mark Musen at Stanford University. LexGrid content presently includes the entire UMLS, specialized versions of SNOMED, entire HL7 table structure, foundational model of anatomy, the gene ontology, and scores of other common terminologies. It is locally extensible to accommodate site- and application-specific content. Cross-linkage across terminologies is also supported, though presently manifest only in some HL7 value set definitions.

LexGrid resources are being prototypically applied in novel ways, not typically envisioned for terminology services. Some of those applications spaces at Mayo include:

- Data Modeling

Mayo Clinic has engaged in a comprehensive data modeling effort, to operate across its clinical, research, and educational spaces. While the top-level model is unconstrained by existing terminologies, subsequent layers adhere to control components within the Lexical Grid.

- Semantic metadata annotation

As data across Mayo Clinic is formalized with respect to data provenance, intended applications, and access privileges, an additional layer of semantic metadata is being formulated to enable broad-based access, intelligent query formulation, and informed retrieval strategies. Mayo Clinic is exploring with IBM practical strategies to add contextual semantics to this annotation, which would include relevant information

associated with disease work-up, clinical guidelines, co-morbidities, and episode-related semantics. For example, a finding of an elevated white blood count carries additional contextual semantics in the setting of hematologic malignancies.

- Retrieval portals

Mayo Clinic has been collaborating with IBM in the creation of a centralized data warehouse of clinical and genomic data. Leveraging the semantic annotations of this data, query formulation requires access to similar terminological discipline. Hence, Common Terminology Services (a component of LexGrid) is incorporated into the query formulation and the NLP (Natural Language Processing) of textual, clinical content within the warehouse.

Discussion:

While the importance of comparable and consistent terminological foundations in health care and informatics has been written about for decades, the practical tools and publicly available terminologies are now making this practical. The task of data integration, beyond the scope of a single parochial application, must eventually confront the semantic integration problem. This is particularly acute in a multidisciplinary, team-science environment, seeking progress across the spectrum of genotype to phenotype interpretation. Mayo Clinic manifests several projects to achieve this, with demonstrable prototypes within its Comprehensive Cancer Center Informatics Core and its newly formed CTSA Informatics group.

A presentation at the workshop would focus on the core technologies within the LexGrid suite that are presently available, and illustrate some of their applications to information integration across the Mayo enterprise.