A Java-based Interface for Medical Research Project Classification Using Metaphrase

Qiang Liu, James D. Buntrock, Douglas L. Crowson, Mark S. Tuttle*, Peter L. Elkin MD, CG Chute MD, DrPH
Mayo Clinic/Foundation, Rochester MN; * Lexical Technology, Inc., Alameda CA

The indexing and retrieval of ongoing clinical research projects, including those overseen by the IRB (Institutional Review Board), is a valuable resource in large medical research organizations. It enables the identification of like projects in the proposal development stage by investigators, and supports the description and enumeration of ongoing research within an institution. The prevalence of active research projects at Mayo numbers several thousand at any one time.

We have substantial experience developing indexing and classification tools for clinical patient description1. We sought to establish whether similar tools, based upon the Metaphrase™ terminology server2, could be generalized for biomedical project description.

Methods

While the Metaphrase terminology server constitutes the underlying tool, we chose MeSH (NLM Medical Subject Headings) as the target classification system. The semantics describing Biomedical research projects parallel the published literature much more closely than they do clinical information, making our patient oriented terminologies such as SNOMED-RT less relevant.

Patient classification systems often have an advantage of being used in a predictable environment, on a defined hardware platform. No such assumption was possible for classification by disparate researchers. Thus, we adopted Java as the development environment, invoking JDBC to access the existing IRB project “demographics” in a Sybase® database. However, even using a commercial implementation of JDK 1.1 (Visual Café 1.1) there were insufficient tools to build a fully robust application. We augmented this with the TeaSet Widgets developed by InetSoft Technology3.

We used two TextGrids as a foundation for a “user-friendly” interface for the application. One was used to display retrieval results from IRB database, and as a data update environment for inserting, modifying and deleting descriptive adjuncts to the core IRB database. Of course, no federally mandated data fields could be altered by this application. The second interacted with the terminology server. Short phrase descriptions of the project are entered, matched against the most appropriate concepts in the UMLS Metathesaurus, and returned as the closest MeSH match possible. Users can then browse the semantic neighborhood of a candidate term, iteratively refining the selection of descriptive terms. When the user is satisfied with the controlled terminology identified, both the descriptive phrase and the MeSH terms are returned to the project description files. The database functions RowState and Vector are invoked to “remember” all changes during an edit session.

Evaluation focused upon system response times, enabling optimizations for useful real-time term assignment.

Discussion

The re-use of middleware object resources in a large networked environment is a goal of all system developers. We demonstrate that a resource intended for patient classification can function for the related but different task of Biomedical project indexing. The Figure illustrates the simple three tiered architecture of our system. Not shown, are the myriad of additional applications which invoke the same middleware, albeit with different interface definitions. Further, we show that the client development environment, in this case Java, is not a significant factor when designing three tiered applications.

3 www.inetsoftcorp.com