DEVELOPING A WEBSERVICE-BASED INDEXING TOOL FOR HEALTH SCIENCE REPOSITORY SYSTEM

Dongming Zhang
Welch Medical Library
Johns Hopkins School of Medicine
Project Goals

- UMLS-based automated MeSH terms assignment;
- WebService-based requests;
- User search query expansion and enhancement;
- Integration with DSpace repository system.
Unified Medical Language System (UMLS)

- Developed by National Library of Medicine;
- Maintain high level standard thesaurus database;
- A compendium of a large number of national and international vocabularies and classifications (over 100) and provides a mapping structure between them:

Examples:
--- Medical Subject Headings (MeSH) (Standard Subject Headings by National Library of Medicine)
--- Systematized Nomenclature of Medicine (SNOMED) (College of American Pathologists)
--- Health Level 7 (HL7) (An international community of healthcare subjects, now HL7, Inc)
Web Service based Service

-- “a software system designed to support interoperable machine-to-machine interaction over a network” defined by “World Wide Web Consortium” (W3C), an main international standard organization;

-- Using XML messages running on SOAP (Simple Object Access Protocol);

-- Web Service Description Language (WSDL) that describes and defines Web services;
Web Services take Web-applications to the next level. Using Web services, your application can publish its function or message to the rest of the world.

Web services use XML to code and to decode data, and SOAP to transport it (using open protocols).

Example: with Web services, your accounting department's Win 2k server's billing system can connect with your IT supplier's UNIX server.
Automated MeSH term Assignment
Process Flow

Input Document
(Title, Abstract, Body)

- **MMTx**
  - UMLS Concepts with Weighting
  - Filters to MeSH Heading
  - MeSH Concepts with Weightings

- **PubMed Similarity Engine**
  - Top 20 Related Citations with Weighting
  - Extracted MeSH Heading
  - MeSH Concepts with Weightings

- **Clustering and Ranking**

- **Ordered List of MeSH Headings**
Web Service Architecture

Service-Oriented Architecture
A completely service-oriented model

User

Services can be used and accessed through any device that hooks up to the web

Platform as a Service
E.g. Integrating with Salesforce.com's CRM

Software as a Service
E.g. Route optimizer software that uses the data to generate quickest routes

Data as a Service
E.g. Addressing data, geodata or personal data (perhaps a list of client information)

Mashups
E.g. Using Google Maps API as front-end

Recurring ongoing cost

Maintained in cloud
Title: A Model of Child Morbidity, Mortality and Health Interventions

Authors: Becker, Stan
Black, Robert

Issue Date: 26-Jul-2006

Abstract: A macro model of morbidity and mortality in children under five years of age is presented. Monthly disease-specific incidence and case fatality rates form the basis of the model and the efficacy and coverage of disease-specific interventions alter these values. In addition, frailty is modeled via relative risks of mortality based on five groups, determined in the first month by the birthweight distribution and at later ages by the proportion of children surviving a given illness who become more frail and the proportion not ill and with adequate nutrition who become less frail. A validation of the model was carried out using data from the comparison and treatment areas of the Demographic Surveillance System in Matlab, Bangladesh. The model very closely predicts the observed mortality level. Scenarios for improvements in coverage of specific interventions in settings in South Asia, West Africa and Andean South America are modeled and their effects on mortality gauged. The model provides a useful tool for those wishing to know the mortality impact of specified mixes of interventions in a given setting. Limitations of the model are considered and possible extensions to address these are discussed.

MeSH Suggested:
mortality
vital statistics
morbidity - disease rate
child
demography
disease
infant mortality
incidence
Search Query Expansion:

[UMLS-based Health Science Repository System]

MeSH Search Results

MeSH Search: All
For: brain injuries

Results 1-1 of 1.

Query Details:

Original Query: brain injuries
Expanded Query: (MeSH: brain injuries * 10.0) OR (MeSH: brain diseases * 9.0) OR (MeSH: craniocerebral trauma * 9.0) OR (MeSH: brain concussion * 7.5) OR (MeSH: epilepsy, post-traumatic * 7.5) OR (MeSH: pneumocephalus * 7.5) OR (MeSH: shaken baby syndrome * 7.5) OR (MeSH: brain hemorrhage, traumatic * 7.5) OR (MeSH: brain injury, chronic * 7.5) OR (MeSH: diffuse axonal injury * 7.5)

Item hits:

<table>
<thead>
<tr>
<th>Date of Issue</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-Oct-2006</td>
<td>Models of Traumatic Brain Injury</td>
<td>Laufer, Helmut; Lenzlinger, Philipp; McIntosh, Tracy</td>
</tr>
</tbody>
</table>
Remote and Plug-in Operation

1. Client submits an item or a search request to IRS (e.g. DSpace) through browser

2. IRS processes the request through regular Keyword Indexes (Lucene based)

3. IRS passes the request to MeSH Index/Search Wrapper

4. MeSH Index/Search Wrapper gets MeSH Heading Suggestions from UMLS-based Indexing Web Service

5. MeSH Index/Search Wrapper processes the request through MeSH Indexes (Lucene based)
Testing and Results

“Gold” Standard from National Library of Medicine

UMLS-based Indexing Tool Results

- Average Recall
- Average Precision

Percentage

Top Suggested MeSH Headings

0 5 10 15 20 25 30 35 40
Future Development

- Develop automated mechanism for SNOMED based term assignment;
- Provide web service to clinical system and applications;
- Ontology work that relate to clinical and translational projects.