Using UMLS to map from a Library to a Clinical Classification: Improving the Functionality of a Digital Library

Judas Robinson,
Postgraduate Student,
St. George’s, University of London,
Cranmer Terrace, Tooting,
London, SW17 0RE
jsrobins@sgul.ac.uk
Primary Care Electronic Library
http://www.pcel.info
Digital Libraries Produced by the PCI Group at St. Georges

DrsDesk (1998-present)

http://drsdesk.sgul.ac.uk


http://www.pcel.org.uk

PCEL (2003-present)

http://www.pcel.info
Medical Coding and Classification Systems

Library Classifications
Exemplified by Medical Subject Headings (MeSH). Used by the National Library of Medicine for indexing articles from 4,800 of the world's leading biomedical journals for the MEDLINE/PubMED database. 20 000 terms organised in hierarchies.

Clinical Classifications
In current use by UK General Practitioner are READ codes. It is planned that these will be superseded by the SNOMED CT classification, which is to be used in the electronic patient record in the UK. 700 000 terms arranged in hierarchies.

Epidemiological Classifications
The International Classification of Diseases (ICD), published by the World Health Organisation (WHO) has become the international standard diagnostic classification for all general epidemiological and many health management purposes.

Different classifications for different purposes
Unified Medical Language System (UMLS)

UMLS consists of three components, the Metathesaurus, the Semantic Network and the SPECIALIST Lexicon. To map between library and clinical classifications we have used the Metathesaurus.

"The Metathesaurus is a very large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health related concepts, their various names, and the relationships among them."

The relationships recorded in the Metathesaurus database have enabled us to map from one classification system to another. It is important to note that not all classifications are represented in the Metathesaurus. READ codes, for example, are not. Some of the classifications which are represented are MeSH, SNOMED CT, ICPC, ICD, and HL7 amongst 50 or so others.
### The End Result

**Primary Care Electronic Library (PCEL)**

**Community Health Sciences**
**St. George’s, University of London**
**Cranmer Terrace, London, SW17 0RE**
**Email: pcel@sgul.ac.uk**

<table>
<thead>
<tr>
<th>SNOMED CT Concept</th>
<th>Disorder of cardiovascular system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
<td>Cardiovascular finding</td>
</tr>
</tbody>
</table>

- **Heart disease** (76)
- **Hypoxemia** (0)
- **Syncope** (1)
- Circulatory system disease NOS (0)
- Disorder of cardiovascular prostheses and implants (0)
- Other specified diseases of circulatory system (0)
- Peripheral vascular communications of care (0)
- Vein, lymphatic and circulatory diseases NOS (0)
- Additional circulatory system disease classification terms (0)
- Cardiovascular disease, unspecified (14)
- Other and unspecified disorders of the circulatory system classification terms (0)

- Atrial Fibrillation
- British Heart Foundation (BHF) - Factfiles
- British Journal of Cardiology (BJC)
- Cardiac Arrhythmia
- Cardiac Risk in the Young (CRY)
- ECG Library
- European Heart Network (EHN)
- Facts About Cardiomyopathy
- Heart UK
- Intermittent claudication
- National Heart Forum (NHF)
- Primary Care Cardiovascular Society
- Primary Care Management of Atrial Fibrillation
- Risk Score for Cardiovascular Disease
Two tasks are necessary to achieve the end result:

• Produce a browsable version of SNOMED CT. Able to identify parent and children of a given node.

• Map MeSH terms to SNOMED CT terms. Assign resources a position in the hierarchy of SNOMED CT concepts.
BROWSABLE VERSION OF SNOMED CT

UMLS: MRHIER table: "Computable" Hierarchies

“This file contains one row for each hierarchy or context in which each atom appears. If a source vocabulary does not contain hierarchies, its atoms will have no rows in this file. If a source vocabulary is multi-hierarchical (allows the same atom to appear in more than one hierarchy), some of its atoms will have more than one row in this file. MRHIER.RRF provides a complete and compact representation of all hierarchies present in all Metathesaurus source vocabularies. Hierarchical displays can be computed by combining data in this file with data in MRCONSO.RRF.”

UMLS documentation

<table>
<thead>
<tr>
<th>AUI</th>
<th>Atom Unique Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAB</td>
<td>Source Abbreviation</td>
</tr>
<tr>
<td>PTR</td>
<td>Path to Root</td>
</tr>
</tbody>
</table>
Sample records

NODE
Disorder of cardiovascular system

PARENT
Cardiovascular finding
A3684559.A3886745.A3456474.A3340519

CHILD
Heart disease
Preparing for the web

• INDEXING HUI COLUMN
  Maximum length of HUI is 278 characters
  TEXT rather than VARCHAR.
  CREATE INDEX x_snomedct_hui ON snomedct(hui(300))

• SUBSETTING TABLES
  No problem finding parent but execution time rises to five to ten minutes when looking for children.

  SELECT hui, str
  FROM snomedct
  WHERE hui LIKE '{$_GET["hui"]}.__________'}}
Mapping from MeSH to SNOMED CT

• For historical reasons a NLM distribution of MeSH separate to, but identical to, UMLS was used.

• Directory indexed according to MeSH Tree Number (MN)
  Cardiomyopathy, Hypertrophic
  C14.280.238.100

• MN maps to Unique Identifier (UI) in MeSH distribution
  Cardiomyopathy, Hypertrophic
  D002312

• UMLS contains the MeSH UI in the CODE column of the MRCONSO table.
UMLS: MRCONSO table: Concept Names and Sources

“There is exactly one row in this file for each atom (each occurrence of each unique string or concept name within each source vocabulary) in the Metathesaurus, i.e., there is exactly one row for each unique AUI in the Metathesaurus. Every string or concept name in the Metathesaurus appears in this file, connected to its language, source vocabularies, and its concept identifier.”

UMLS documentation

<table>
<thead>
<tr>
<th>CODE</th>
<th>Source Asserted Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAB</td>
<td>Source Abbreviation</td>
</tr>
<tr>
<td>CUI</td>
<td>Concept Unique Identifier</td>
</tr>
<tr>
<td>AUI</td>
<td>Atom Unique Identifier</td>
</tr>
</tbody>
</table>
From MeSH UI to SNOMED CT HUI

• Select distinct CUIs corresponding to a given MeSH UI (limiting SAB to MSH):
  D002312
  C0007194
  C0205700
  C0700053

• Select AUIs corresponding to the given CUIs limiting the SAB to SNOMEDCT
  C0007194
  A2872579
  A2889219
  A2978478
  A2974495
  A3027399
  A3027616
  A3501792
  A3501820
UMLS: MRHIER table: "Computable" Hierarchies

“This file contains one row for each hierarchy or context in which each atom appears. If a source vocabulary does not contain hierarchies, its atoms will have no rows in this file. If a source vocabulary is multi-hierarchical (allows the same atom to appear in more than one hierarchy), some of its atoms will have more than one row in this file. MRHIER.RRF provides a complete and compact representation of all hierarchies present in all Metathesaurus source vocabularies. Hierarchical displays can be computed by combining data in this file with data in MRCONSO.RRF.”

UMLS documentation

<table>
<thead>
<tr>
<th>AUI</th>
<th>Atom Unique Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAB</td>
<td>Source Abbreviation</td>
</tr>
<tr>
<td>PTR</td>
<td>Path to Root</td>
</tr>
</tbody>
</table>
## Final database structure

**snomedct**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUI</td>
<td>Hierarchical Unique Identifier</td>
</tr>
<tr>
<td>STR</td>
<td>String</td>
</tr>
</tbody>
</table>

**snomedct_short**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUI</td>
<td>Hierarchical Unique Identifier</td>
</tr>
<tr>
<td>STR</td>
<td>String</td>
</tr>
</tbody>
</table>

**directory_snomedct**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUI</td>
<td>Hierarchical Unique Identifier</td>
</tr>
<tr>
<td>RES_ID</td>
<td>Resource ID</td>
</tr>
</tbody>
</table>
Implications for Digital Libraries

**Improved data retrieval**
Clinicians are now able to browse the hierarchical structure of SNOMED CT to identify relevant resources.

**Integration with the clinical record**
There is an auto-completion search tool online for SNOMED-CT terms included in the digital library. This permits users to type a series of letters and be provided with matching SNOMED CT terms. The next step, which is not technically challenging, is to allow clinicians to link to SNOMED CT terms automatically from the electronic patient record.
Using UMLS to map from a Library to a Clinical Classification: Improving the Functionality of a Digital Library

Primary Care Electronic Library:
http://www.pcel.info/

Power Point Presentation:
http://www.gpinformatics.org/meetings.htm

Judas Robinson,
Postgraduate Student,
St. George’s, University of London,
Cranmer Terrace, Tooting,
London, SW17 0RE
jsrobins@sgul.ac.uk