Using AMT & SNOMED CT-AU to support clinical research

Simon J. McBRIDE, Michael J. LAWLEY, Hugo LEROUX and Simon GIBSON
CSIRO Australian E-Health Research Centre

2 August 2012
Overview

• Context
• Problem
• Solution options
• Method
• Results
• Limitations & Future work
Australian Imaging, Biomarkers & Lifestyle Study of Ageing (AIBL)

- Large scale (+1,100 participants)
- 4.5+ year prospective, longitudinal study of ageing
- 4 collections of data in 18 month intervals since 2006
- Research streams:
  - Cognitive
  - Imaging
  - Biomarkers
  - Lifestyle

- “Medication” definition
  - Pharmaceuticals & Nutraceuticals
- Medication records include:
  - Name
  - Dose
  - Frequency
  - Duration of use
Problem

• AIBL’s medication use data quality was poor
  • Participant self- and/or carer-assisted reporting to paper records
  • Manual entry from paper to an electronic data capture system without medication support (free text fields)

• Types of issues
  • Misspelling of medication names
  • Incomplete records
  • Mix of brand/product names & generic names (e.g. “Cartia” vs “Aspirin”)

• Consequence
  • Difficult to analyse medication use within the cohort due to paper records

• How do we
  • Improve the quality of legacy data?
  • Ensure this problem doesn’t occur at later time points?
Solution Candidates

• Candidate evaluation
  • Support for use cases:
    – “How many Participants are taking <medication>?”
    – “Is there a correlation between <medication> and <observation>?”
  • Match to user community expectations
  • Long term sustainability
  • “Secret sauce”?  
    – Is there value in being able to exploit semantics in the data during analysis?

• Candidates
  • Commercial: MIMS Integrated
  • Standards: Australian Medicines Terminology (AMT), SNOMED CT-AU
  • Public domain: DrugBank
Method

• AMT Mapping
  • Concepts below Trade Product
• SNOMED CT-AU Mapping
  • Concepts below Substance
• “Direct” algorithm
  • Exact string match on Preferred Term, Fully Specified Name & Description (via Lucene)
• “Least common ancestor” (LCA) algorithm
  • LCA: The common ancestors that are not ancestors of any other common ancestor
  • Navigate hierarchy of candidate concepts’ ancestors to find least common of those ancestors
Ontoserver API

• RESTful API providing a number of useful operations including

  XMLResponse `findConcepts`(term, context, ...)

  XMLResponse `findConceptsByTerm`(term, max)

  XMLResponse `concept`(id)

  XMLResponse `subsumedBy`(term, predicate)

  XMLResponse `parents`(term, max)

  XMLResponse `children`(term, max)
# Results

<table>
<thead>
<tr>
<th>Mapped To</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>523</td>
<td>43.2</td>
</tr>
<tr>
<td>LCA</td>
<td>687</td>
<td>56.8</td>
</tr>
<tr>
<td></td>
<td>1210</td>
<td>56.1</td>
</tr>
<tr>
<td><strong>SNOMED CT-AU</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td>147</td>
<td>42.4</td>
</tr>
<tr>
<td>LCA</td>
<td>200</td>
<td>57.6</td>
</tr>
<tr>
<td></td>
<td>347</td>
<td>16.1</td>
</tr>
<tr>
<td><strong>Unknown</strong></td>
<td>601</td>
<td>27.8</td>
</tr>
<tr>
<td><strong>Total Mapped</strong></td>
<td>1557</td>
<td>72.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2158</td>
<td>100</td>
</tr>
</tbody>
</table>
Limitations & future work

- Relatively naive direct mapping algorithms
  - Do different string matching algorithms improve Direct algorithm performance?
- More formal evaluation
  - No “ground truth” when the work was done.
  - Recently acquired later time point data that includes a manual mapping to MIMS Australia terms.
  - Planning an evaluation and improvement of our mapping algorithm using the manual mapping “ground truth”.
- Implementation of AMT/SNOMED CT-AU concept lookup during data entry to avoid the need to complete this task in future.
- Implementation of tools to exploit the relationships between AMT/SNOMED CT-AU concepts in query and visualisation for research purposes – Wordle example
Fun with Wordle

With thanks to Wordle (see http://wordle.net/)
Snapper & Ontoserver for fun and profit

• AEHRC has tools that you can play with
  • Ts & Cs say “fun” only in the first instance, talk to us if you’re after profit
• Snapper
  • Terminology mapping tool
  • Free download at http://aehrc.com/snapper
• Ontoserver
  • Free access to an Ontoserver instance:
    http://ec2-23-20-239-33.compute-1.amazonaws.com:8080/ontoserver/
  • Ontology server providing a useful API (SNOMED variants & AMT)
  • RESTful API – WADL at http://ec2-23-20-239-33.compute-1.amazonaws.com:8080/ontoserver/resources/application.wadl
  • Don’t try writing it down, contact me after the session
Questions, thank you & more information

Simon McBride
Research Project Leader
Australian E-Health Research Centre

+61 7 3253 3631
simon.mcbride@csiro.au
http://aehrc.com/

Many thanks to co-authors
• Dr Michael Lawley
• Dr Hugo Leroux
• Mr Simon Gibson

AIBL:
• http://aibl.csiro.au/

Australian E-Health Research Centre (AEHRC)
• http://aehrc.com/
• Come visit us at Booth 26