Evaluation of Multi-Terminology Super-Concepts for Information Retrieval

Griffon N\textsuperscript{a}, Soualmia LF\textsuperscript{b,c}, Névéo\textsuperscript{l} A\textsuperscript{d}, Massari P\textsuperscript{a}, Thirion B\textsuperscript{a,b}, Dahamna B\textsuperscript{a,b} & Stéfan J Darmoni\textsuperscript{a,b}

\textsuperscript{a} CISMeF, Rouen University Hospital, France
\textsuperscript{b} TIBS & LITIS EA4108, Rouen University, France
\textsuperscript{c} LIM&Bio, University of Paris 13, Sorbonne Paris Cité, France
\textsuperscript{d} National Center for Biotechnology Information, NLM, Bethesda, MD 20894, USA

Email: Stefan.Darmoni@chu-rouen.fr

MIE August 2011
Introduction

- Quality-controlled subject gateways were defined by Koch as Internet services which apply a comprehensive set of quality measures to support systematic resource discovery.
- CISMeF ([French] acronym for Catalog and Index of French Language Health Resources on the Internet) was designed to catalog and index the most important and quality-controlled sources of institutional health information in French:
  - began in *February 1995*
  - [www.cismef.org](http://www.cismef.org)
CISMeF terminology

- Initially based on the **MeSH** (Medical Subject Headings) thesaurus from the US National Library of Medicine
  - Granularity
  - Well known

- MeSH terms were gathered under **Super-Concepts**
  - MeSH super-concepts\(^a\)
  - Correspond roughly to medical specialties (e.g. surgery), biological sciences (e.g. genetics) or health topics (e.g. diagnosis)
  - Semantic links manually created to MeSH terms, MeSH subheadings and CISMeF resource types
  - To maximize information retrieval in CISMeF\(^b\); allowing categorization\(^c\)

\(^a\)Thirion B, Darmoni SJ. Simplified access to MeSH tree structures on CISMeF. Bull Med Libr Assoc. 1999 Oct;87(4):480-1
\(^b\)Gehanno JF, Thirion B, Darmoni SJ. Evaluation of meta-concept for information retrieval in a quality controlled health gateway. AMIA Annu Sylo Proc. 2007;269-73

MIE August 2011
CISMeF super-concepts

- The use of super-concepts came up to cope with the relative restrictive nature of these MeSH terms
- To illustrate the difference between MeSH terms and super-concepts in terms of IR in CISMeF, two queries
  - 'guidelines in cardiology’
  - 'databases in virology’,
- The query 'guidelines in cardiology’ retrieves 11 resources when 'cardiology' is considered as a MeSH term vs. 143 resources when 'cardiology' is considered as a CISMeF SC
- The query 'databases in virology’ retrieves 0 resource when 'virology' is considered as a MeSH term vs. 4 resources when 'virology' is considered as a CISMeF SC
CISMeF terminology (cont.)

- Since 2009, CISMeF is fully « multi-terminological »
  - CISMeF backoffice contains the main health terminologies available in French (e.g. SNOMED Int, ICD-10, ATC, CCAM) (n=32)
  - Multi-terminological automatic indexing (better recall)
  - Multi-terminological information retrieval

- Enrichment of super-concepts:
  - Multi-terminology super-concepts for the following T/O:
    - ICD-10, ATC, CCAM, FMA, SNOMED Int (Disease axis)
  - Available at pts.chu-rouen.fr

---

Darmoni, SJ; Sakji, S; Pereira, S; Merabti T; Prieur E; Joubert M & Thirion B. Multiple terminologies in an health portal: automatic indexing and information retrieval. Artificial Intelligence in Medicine, Verona, Italy, July, Lecture Notes in Computer Science, Pages 255-259, Springer, 2009.

MIE August 2011
**CISMeF Super-Concepts**

- **Descriptors**
- **Qualifiers**
- **Ressource types**

---

**CISMeF based on MeSH only**

- **Term**
  - Is-a or Part-of relationship
  - Super-concept – term association
  - Synonym

---

**Surgery**

---

**Super-Concepts**
CISMeF Super-Concepts: an example

CISMeF SC = cardiology

- MeSH (n=39 + descendants)
  - Cardiopulmonary bypass …
- ICD10 (n=94)
  - Acute pericarditis
- CCAM (n=206)
  - DAQL012 - Scintigraphie des cavités cardiaques à visée rythmologique
- ATC (n=9 + descendants)
  - Antihypertensives
- Etc..
Objectives

To assess the effect of multi-terminology SC (MT-SC) definition compared to MeSH-only SC (MeSH-SC) definition on information retrieval performance in CISMeF.
Methods: defining queries

- MT-SC are based on MeSH-SC plus semantic links to some terms in other terminologies:
  - **MeSH-SC query**: query retrieving resources indexed by a term linked to MeSH-SC
    - “Surgery[MeSH-SC]”
  - MT-SC query: query retrieving resources indexed by a term linked to MT-SC
    - “Surgery[MT-SC]”
  - **Delta query**: query retrieving resources indexed by a term linked to MT-SC and not to MeSH-SC
Methods: evaluation

- Top 20 answers of MeSH-SC and Delta queries were evaluated by one resident (NG).
- Qualitative assessment of relevance using a 3-point Likert scale (fully, partly and not relevant).
- **Precision**: mean weighted precisions were computed for two levels of relevance. Comparison between indexing methods (automatic vs manual) were performed.
- **Recall**: relative recall of MeSH-SC queries was computed assuming MT-SC queries’ recall was 1.
## Results

### Mean weighted precision:

<table>
<thead>
<tr>
<th></th>
<th>MeSH-SC queries</th>
<th>Delta queries</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial relevance</td>
<td>80%</td>
<td>76%</td>
<td>0.3</td>
</tr>
<tr>
<td>Full relevance</td>
<td>66%</td>
<td>33%</td>
<td>&lt;10^{-4}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Automatic indexing</th>
<th>Manual indexing</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial relevance</td>
<td>48%</td>
<td>81%</td>
<td>&lt;10^{-4}</td>
</tr>
<tr>
<td>Full relevance</td>
<td>38%</td>
<td>50%</td>
<td>0.004</td>
</tr>
</tbody>
</table>

### Relative recall:

<table>
<thead>
<tr>
<th></th>
<th>MeSH-SC queries</th>
<th>MT-SC queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial relevance</td>
<td>86%</td>
<td>100%</td>
</tr>
<tr>
<td>Full relevance</td>
<td>92%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Discussion

- Shift from MeSH to multi-terminology:
  - Higher recall
  - Same or lower precision according to relevance level
- Lack in performance of automatic indexing
Discussion

- This study has two biases against MT-SC:
  - Links from MeSH to SC were made and improved for 15 years whereas MT to SC links were barely made. More erroneous hand-crafted links were found for MT-SC than MeSH-SC. However this is not really frequent.
  - Multi-terminology indexing concerns only new resources that are different from old MeSH-only indexed resources. These 2 sets of resources are not comparable (e.g. some of these new resources, providing very standardized and precise information, need new indexing strategy to avoid them inducing noise).
Perspectives: new applications of MT-SC

- information retrieval in EHR
  - e.g. select all patients with mt-sc='cardiology’ and elevated troponine
  - RAVEL project (TecSan program, French Research Agency); 2012-4

- Categorization: concept oriented views
  - Active on the patient summary (ICD10, CCAM)a; SNOMED to be used

---


MIE August 2011
Conclusion

- MT-SC queries will be best used when the MeSH-SC result set is small.
- Automated tools for indexing needs to be improved significantly
Thank you for your attention

Questions?
Additional Subject Subsets for PubMed

PubMed citations can be retrieved on the following specialized topics in subject subsets, namely by launching an in-built complex search strategy.

<table>
<thead>
<tr>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>addiction</td>
</tr>
<tr>
<td>allergology_and_immunology</td>
</tr>
<tr>
<td>anatomy</td>
</tr>
<tr>
<td>anesthesiology</td>
</tr>
<tr>
<td>bacteriology</td>
</tr>
<tr>
<td>cardiology</td>
</tr>
<tr>
<td>cytology</td>
</tr>
<tr>
<td>dentistry</td>
</tr>
<tr>
<td>dermatology</td>
</tr>
<tr>
<td>developmental_biology_and_medicine</td>
</tr>
<tr>
<td>diagnosis</td>
</tr>
<tr>
<td>disability</td>
</tr>
<tr>
<td>disease_transmission</td>
</tr>
<tr>
<td>economics</td>
</tr>
<tr>
<td>education</td>
</tr>
<tr>
<td>embryology</td>
</tr>
<tr>
<td>emergency_medicine</td>
</tr>
<tr>
<td>endocrinology</td>
</tr>
<tr>
<td>epidemiology</td>
</tr>
<tr>
<td>ethics</td>
</tr>
<tr>
<td>family_medicine</td>
</tr>
<tr>
<td>gastroenterology</td>
</tr>
<tr>
<td>genetics</td>
</tr>
<tr>
<td>geriatrics</td>
</tr>
<tr>
<td>gynecology</td>
</tr>
<tr>
<td>hematology</td>
</tr>
</tbody>
</table>
Query Translation:


Result:

234114

Database:
PubMed

User query: