CISMeF

Catalog & Index of Health Resources in French on the Internet

Prof. S.J. Darmoni, MD, PhD

TIBS, LITIS Lab
Rouen University Hospital & Rouen Medical School, France

Email: Stefan.Darmoni@chu-rouen.fr

Oslo August 2009
Knowledge Engineering & Health Information Systems (GCSIS)

Prof. S.J. Darmoni

N. Chaingaud
JP. Kotowicz
JP. Pécuchet

CISMeF terminology
- Encapsulated MeSH
- Metaterms
- Resource types
- Strategy searches
- Metadata

L. Soualmia
B. Thirion, C. Letord
G. Kerdelhué, J. Piot

Textual automatic indexing
- NLP, KNN
- Categorization

A. Rogozan
A. Neveol

MONOTERMINOLOGY
1995-2005
Knowledge Engineering & Health Information Systems (GCSIS)

**CISMeF terminology**
- Implicit Information Retrieval
- Encapsulated MeSH
- Metaterms
- Resource types
- Strategy searches
- Metadata

**Textual automatic indexing**
- NLP, KNN
- Categorization

**Multiterminology server**
- ICD 10
- CCAM
- SNOMED
- UMLS

**Other Medical Terminologies & Dictionaries**
- UMLF, VUMeF, VODeL, PIH

**Monotermology 1995-2005**
- A. Rogozan
- A. Neveol

**Multiterminology 2005-**
- M. Joubert / JF. Gehanno
- B. Thirion, C. Letord
- G. Kerdelhué, J. Piot
- L. Soualmia

**Health Information Systems**
- P. Massari

**French Infobutton**
- Contextual Knowledge

**Semantic Interoperability**
- Intra and Inter Terminologies in Health

**Multi terminology**
- B. Dahamna
- IM. Kergourlay

**Computer-assisted coding system**
- S. Pereira

**Semantic Interoperability**
- T. Merabti

**Contextual Knowledge**
- S. Pereira
Introduction

- Quality controlled subject gateways (or portals) were defined by Koch as Internet services which apply a comprehensive set of quality measures to support systematic resource discovery.

- CISMeF = quality controlled health gateway for French institutional health resources
  ✓ www.cismef.org
Introduction

The objective of CISMeF (Catalog and Index of French-speaking resources) is to assist the health professional & lay people during the search for electronic information available on the Internet. CISMeF covers healthcare disciplines and medical sciences.

CISMeF was a project originally initiated by Rouen University Hospital (RUH).


CISMeF began in February 1995

Doc’CISMeF in 2000: creation of a generic search tool using the CISMeF semi-informal ontology

URL: http://doccismef.chu-rouen.fr/
Introduction (cont.)

- Three priority axes:
  - evidence based medicine
  - teaching material
  - patient information
- >68,000 resources included
- 30,000 unique machines/working day
- CISMeF team in 2009: N= 12
  - 1.5 medical informaticians
  - 1 chief medical librarian + 2.5 medical librarians
  - 1 computer scientist (one junior lecturer)
  - 3 research engineers
  - 3 PhD students + 8 students from engineering school (1/4 time)
- Budget ≈ 300 K€/y; 30% RUH
- 20 grants in the last five years for CISMeF (GCSIS)
Main explanation of the CISMeF success

- Cooperation between
  - Medical librarians
  - Computer Scientists
  - Medical Informaticians
- Will to go ahead +++
- Minimum of complaints in front of adversity
Update of the CISMeF database

Methodology in 4 steps

- collection management
- filtering, based on Net Scoring (49 criteria) & MedCIRCLE project, and more recently Health on the Net code of conduct
- resource description (an extensive and documented metadata set)
- resource indexing (+ “super-indexing”) (using a controlled vocabulary system)
CISMeF terminology

- Two standard tools for organising information:
  - the MeSH (Medical Subject Headings) thesaurus from the US National Library of Medicine
  - Several metadata element sets
    - the Dublin Core metadata format + CISMeF specific fields
    - For teaching resources, IEEE 1484 LOM metadata format
      11 elements of the LOM Educational category => DC.Education
    - For evidence-based medicine resources, CISMeF specific fields: level of evidence + method to evaluate it
    - The HIDDEL metadata set is used to enhance transparency, trust and quality of health information on the Internet.

- Do not reinvent the wheel +++ but adapt it

DC-2004, International Conference on Dublin Core and Metadata Applications
The heterogeneity of Internet health resources led the CISMeF team to enhance the MeSH thesaurus with the introduction of two new concepts:

- resource types (N≈300),
- metaterms (N≈120),
- predefined queries (N≈200)
MeSH ‘enhancements’

- Improvement of the MeSH thesaurus itself
  - Add-on of 10,000 synonyms, including (ambiguous) acronyms
  - Manual translations of 6,000 definitions (semi-automatic translation for the rest of the MeSH soon)
  - Translation of 7,500 MeSH Supplementary Concepts (SC) & add-on of 6,000 synonyms
CISMeF resource types

- CISMeF resource types (RT) are a generalization of the publication types of Medline
- As defined by the DC Metadata Initiative, a RT is used to categorize the nature of the content of the resource, when MeSH (term/subheading) pairs describe the topic of the resource
- Specific of the health resources available on the Internet, such as association, patient information, community networks

CISMeF resource types

- CISMeF RTs are organized similarly to MeSH terms and subheadings, in a hierarchical structure with subsumption relationships (allowing the explode property) and a maximum of five-level depth vs. the Medline publication types mainly a flat list.

- This RT list was manually built and maintained by the CISMeF team since 1997.

- This list is largely inspired by the MeSH thesaurus as 187 RTs (76%) are deliberately ambiguous because they are also MeSH terms (e.g. magnetic resonance imaging).

- education [MH]
  - teaching material [MH]
    - audiovisual aids [MH]
    - instruction [PT]
  - critical appraisal
  - lectures [PT]
    - French pre-residency program examination
  - practicals
  - problems and exercises [PT]
    - case report
    - clinical reasoning learning
    - correct version of the examination
    - critical appraisal
    - examination questions
      - multiple choice quiz (MCQ)
      - open and closed questions
    - problem-based learning [MH]
  - textbooks [MH]
  - tutorials
- teaching structure
  - hospitals, teaching [MH]
  - schools, health occupations [MH]
    - schools, dental [MH]
    - schools, medicine [MH]
    - schools, nursing [MH]
    - schools, pharmacy [MH]
    - schools, public health [MH]
    - schools, veterinary [MH]
- training
- establishment, institution, organization
- Resource guides
  - software
  - text
    - abstract
    - bibliography [PT]
    - biobibliography [PT]
    - biography [PT]
    - comics
    - comparative study [MH]
    - congresses [PT]
    - dictionaries, medical [MH]
    - dissertations, academic [MH]
    - encyclopedias [PT]
    - evaluation studies [PT]
      - economic evaluation study
      - professional practice evaluation study
      - public health evaluation study
      - health technology assessment
    - forms [PT]
    - government publications [MH]
    - guide
    - guidelines [MH]
      - health policy guidelines
      - patients guideline
      - practice guidelines [MH]
        - clinical practice guidelines
        - consensus development conferences [MH]
        - formal expert consensus
      - public health guidelines
    - journal article [PT]
    - laboratory manuals [PT]
    - lectures [PT]
    - legislation [PT]
    - meta-analysis [PT]
    - monograph [PT]
CISMeF metaterms

- Semantic links manually created to MeSH terms, MeSH subheadings and CISMeF resource types
- Can be considered in the CISMeF terminology as super-concepts
- Semantic links were created on the basis of the CISMeF librarians' technical know-how and the expertise of medical specialists from the Rouen University Hospital
- Biomedical specialities (most cases also MeSH terms, in the G02.403 MeSH tree )
- used in IR & Categorization & visualization and navigation through the concept hierarchies

CISMeF terminology (cont.)

Keywords Hierarchy  Sub headings Hierarchy  Resource Types Hierarchy

Term
Is-A relation
Metaterm-term association
Synonyms
CISMeF terminology (cont.)

```html
<html>
<meta name="DC.Description" ... />
<meta name="DC.Format" ... />
<meta name="DC.Subject" scheme="MeSH"
  content="epidemiology; public health"/>
<meta name="DC.Subject" scheme="F-MeSH"
  content="épidémiologie; santé publique"/>
...
<meta name="CISMeF.Date.Consultation"
  scheme="ISO8601"/>
...
<meta name="CISMeF.Type"
  scheme="CISMeF"
  content="Réseaux coordonnés"/>
...
<meta name="Keywords"
  content="*épidémiologie, epidemiology; *santé publique, public health"/>
</html>
```
CISMeF metaterms

- The use of metaterms came up to cope with the relative restrictive nature of these MeSH terms.
- To illustrate the difference between MeSH terms and metaterms 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 MT.
- 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 MT.
Méta-terme "toxicologie"

**Toxicologie** est un "métatérme". Voir ci-dessous les arborescences, mots clés et qualificatifs du thesaurus MeSH, utilisé notamment pour la base de données bibliographiques Medline, le concernant. Les termes utilisés sont ceux du MeSH en français réalisé par l'INSERM. Les informations à destination des patients sont accessibles par CISMeF-p.
Le lien proposé sur Doc'CISMeF exécute une requête sur l'union de tous ces termes.

- antidotes [MeSH]
- antitoxine [MeSH]
- centre anti-poison [MeSH ; type]
- centre traitement toxicomanie [MeSH ; type]
- détoxication médicaments [MeSH]
- effets indésirables [qualificatif]
- induit chimiquement [qualificatif]
- intoxication [MeSH]
- intoxication [qualificatif]
- pesticides [MeSH]
- polluants environnement [MeSH]
- pollution environnement [MeSH]
- produits illicites [MeSH]
- tabagisme [MeSH]
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>
</tbody>
</table>
Query Translation:


Result:

234114

Database:

PubMed

User query:

Predefined queries

- Some important medical terms are not existing in the MeSH thesaurus
  - e.g. insulin therapy, colonic surgery
- Query including MeSH descriptors, MeSH qualifiers and boolean operators
  - Terme en français : insulinothérapie
  - Terme en anglais : insulin therapy
  - Requête : insuline/usage thérapeutique.mc
Predefined queries

- chirurgie colorectale
- Terme en anglais : colorectal surgery
- Query
  

Metaterms vs. PubMed Subsets

- PubMed Subsets: prerequests including MeSH terms, MeSH subheadings, textword and journals
- Very interesting for brand-new publications (PreMEDLINE)
- Metaterms: better precision but far worse recall… nonetheless
Metaterms vs. Journal Descriptors

- JDs medical specialties
- Manual link between JDs and Journals indexed in the Medline database
  - e.g. Journal of Cardiac Surgery => two JDs: Cardiology & Surgery
- Statistical score between JDs and MeSH terms
- Study design MTs vs. JDs
- Step 0: manual mapping between MTs and JDs
- Step 1: improving MTs using JDs

Feasibility study: heterogeneous results according to JDs

Very interesting for JD Cardiology, less interesting for Gastroenterology

- Step 2: comparison of MTs vs. JDs on 100 manual categorization MEDLINE articles (consensus of 2 experts)
  collaboration with Patrick Ruch & Julien Gobeill (HUG)
- JD > MT for categorization
- NLM JDs to improve CISMeF MTs by 3-5%

CISMeF-patient

- Very ambitious objective: to become the equivalent of US NLM MEDLINEplus for France
- Same model than CISMeF
  - Instead of MEDLINEplus Topics, CHI synonyms of MeSH terms
    - ‘mal’ for ‘douleur’
    - ‘chaude pisse’ (vulgar or ) for ‘gonorrhea’
  - Possibility to modify the query to search the same CHI synonym on teaching document or guideline (EBM)
  - Possibility to search the same topic on MEDLINEplus CHI (Fr) => MeSH term (Fr) => MeSH term (Eng)
CISMeF-patient (cont.)

- Need for the MEDLINEplus Topic-MeSH table to finish the loop
- Generic approach
- MIE 2006 (in press)
CISMef-patient (cont.)

- Two main uses of CISMef patient
  - MDs (GPs) = infomediaition; to give to the patient the relevant URLs => patient education
  - Patient or consumer to search by himself
  - If the MDs are not going towards infomediaition, patients and consumers will do it… without their will and consent

- URL: www.chu-rouen.fr/cismefp
Metaterms in Categorization

- major score and a minor score computed
- $\text{major}(T_i) = \text{Card}\{M^*, Q^*, R^*, \text{inducing } T_i\}$, the number of *major* indexing terms and resource types which induce $T_i$
- $\text{minor}(T_j) = \text{Card}\{M, Q, R, \text{inducing } T_j\}$, the number of *minor* indexing terms which induce $T_j$
- Precision 81%; Recall 93; 62% of the resources were assigned a “fully relevant” or “fairly relevant” categorization according to strict standards

International Journal of Medical Informatics (2004) **73**, 57-64
<table>
<thead>
<tr>
<th>Contenu</th>
<th>Spécialités</th>
<th>Mots-clés</th>
</tr>
</thead>
<tbody>
<tr>
<td>présentation des services proposés, annuaire des structures d'aide</td>
<td>*virologie</td>
<td>hépacivirus</td>
</tr>
<tr>
<td>psychologique et/ou psychiatrique pour les personnes vivant avec le</td>
<td>*infectiologie</td>
<td>hépatites virales humaines</td>
</tr>
<tr>
<td>VIH ou l'hépatite en Ile-de-France, lexique, sélection de sites, liste</td>
<td>*allergologie et immunologie</td>
<td>infections à VIH</td>
</tr>
<tr>
<td>de ressources en cas d'urgence</td>
<td>hépatologie</td>
<td>soutien social</td>
</tr>
<tr>
<td></td>
<td>gastroentérologie</td>
<td>*syndrome d'immunodéficience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>acquise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VIH</td>
</tr>
<tr>
<td></td>
<td>*réseaux coordonnés</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cible(s)</th>
<th>patient / grand public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Langue(s)</td>
<td>français</td>
</tr>
<tr>
<td>Pays/ville</td>
<td>France - Paris (Paris)</td>
</tr>
<tr>
<td>Type(s)</td>
<td>*réseaux coordonnés</td>
</tr>
</tbody>
</table>
Medline Categorization

- URL: http://documvf.crihan.fr/servlets/Categoriseur
- Categorization of any MEDLINE file by CISMeF Metaterms

### Exemple for « Darmoni »

<table>
<thead>
<tr>
<th>Specialties</th>
<th>Score</th>
<th>Majors</th>
<th>Minors</th>
<th>PubMed link</th>
<th>CISMeF link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keywords Qualifiers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information science</td>
<td>2027</td>
<td>161</td>
<td>139</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medical informatics</td>
<td>1391</td>
<td>116</td>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>education</td>
<td>366</td>
<td>28</td>
<td>4</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>environment and public health</td>
<td>182</td>
<td>8</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>epidemiology</td>
<td>147</td>
<td>2</td>
<td>42</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>organization and administration</td>
<td>101</td>
<td>39</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physiology</td>
<td>85</td>
<td>4</td>
<td>12</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>statistics</td>
<td>74</td>
<td>8</td>
<td>19</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>gastroenterology</td>
<td>70</td>
<td>4</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Information Retrieval in CISMeF: history

- Small size (N=14,500 resources in 2002)
- Recall >> Precision
  - e.g. metaterms
  - True from 1995 to 2005
- Query segmentation
- Character normalisations
- Stop words
- Exact expression
- « Phonemization »
- Spelling correction (optional)
- Stemming
- Bag of words
Lina Soulmia’s PhD

✓ In the Simple Search, display of
  • See also
  • Pharmacological actions
  • Do not confuse with
  • Derivations
  • Rules (text mining)

✓ Indexing rules
  • A => B ; IR: A => A OR B ; B => B or A
  • Spine/injuries => spinal injuries demo

✓ Expanding or limiting the original query
IR in CISMeF: history

- Five steps
  - Reserved terms (∈ CISMeF terminology)
  - Title
  - Mixing the reserved terms and the titles
  - Mixing the reserved terms, all fields and adjacency in the titles (word adjacency: (n-1)*5)
  - Mixing the reserved terms, all fields and adjacency in the plain texts (word adjacency: (n-1)*10)


- Since 2006, only three steps (1 U 2 U 3 => 1’)

Since 2006, only three steps (1 U 2 U 3 => 1’)
### Repartition of the queries matched with documents in the CISMef database

<table>
<thead>
<tr>
<th>Operation</th>
<th>Number of queries with documents in return</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>57</td>
<td>27%</td>
</tr>
<tr>
<td>Step 2</td>
<td>14</td>
<td>7%</td>
</tr>
<tr>
<td>Step 3</td>
<td>8</td>
<td>4%</td>
</tr>
<tr>
<td>Step 4</td>
<td>22</td>
<td>10%</td>
</tr>
<tr>
<td>Step 5</td>
<td>36</td>
<td>17%</td>
</tr>
<tr>
<td>Total</td>
<td>176/250</td>
<td>65%</td>
</tr>
</tbody>
</table>
IR in CISMeF: currently

- Only three steps
  
  Step 1: Reserved terms (∈ CISMeF terminology) OR document's title
  
  Step 2: The CISMeF metadata
    Mixing the reserved terms, all fields and adjacency in the titles (word adjacency: (n-1)*5)
  
  Step 3: Adjacency in the plain texts
    Mixing the reserved terms, all fields and adjacency in the plain texts (word adjacency: (n-1)*10)
Since 2005, four levels of indexing in CISMeF

- Level 1: manuel indexing (e.g. guidelines)
- Level 2: supervised indexing (e.g. technical report or teaching document from national medical societies)
- Level 3: automatic indexing (e.g. SCPs, teaching document from one medical school)
- Level 4: extending the CISMeF corpus => Google CISMeF (restricted to publishers included in CISMeF)
CISMeF Information Retrieval

- Some differences with PubMed
  - Resources automatically indexed included
- CISMeF resource ranking
  - Analysis of the query
  - MeSH Major (or Title) first (display of score)
    - Then, date (as PubMed)
  - Automatic (Title or SubTitle)
  - Minor MeSH
**CISMeF Information Retrieval**

- F-UMLS + terminologies in French (not yet) included in the UMLS
  - ✓ Orphanet, CCAM, DRC
- Mapping these terminologies to UMLS via F-MTI
  - PhD Tayeb Merabti 2Y (Algeria)
  - Bag of words: perfectly matching 18% of the Orphanet thesaurus

---

**Granted by the French NSF**
Current CISMeF search engine

1 ressource(s) trouvée(s) en 0,3 secondes, pour : mason (titre) et wyburn (titre) - Interprétation de la requête :

1. Wyburn-Mason, syndrome de - Bonnet-Dechaume-Blanc, syndrome de - [2007]

[ Site éditeur : Orphanet base de données sur les maladies rares et les médicaments orphelins. Direction Générale de la Santé / INSERM]

mots-clés:

types:

accès:

http://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=FR&Expert=53719
CISMeF search engine in 2010?

- Wyburn-Mason, Syndrome
- Orphanet, Syndrome de Wyburn-Mason
- MedDRA
Inclusion of definitions

Définition [MeSH Scope Note, traduction CISMef]: Présence de bactéries viables circulant dans le sang. La fièvre, les frissons, la tachycardie, et la tachypnée sont les manifestations aiguës communes de la bactériémie. La majorité des cas sont observés chez des patients déjà hospitalisés, dont la plupart ont des maladies ou des modalités de soins pré-existants qui rendent leur...
Affiliation of a resource type

- Affiliation of a RT to a MeSH term or to a MeSH (term/subheading) pair
- To obtain a new affiliation concept [MeSH (term/subheading)\CISMeF RT] constituting a triplet, to be more precise during the manual indexing process =>
  to be more precise during the IR
- Affiliation of a RT is similar conceptually to the affiliation of SH, taking into account the respective definitions of a RT and a SH

Submitted to BDI
Any search could be limited to Major Topics for each level of the CISMeF terminology

Major/Minor for resource types
- Manually indexed

Major/Minor for metaterms
- Automatically calculated
- A metaterm is “major” for a CISMeF resource if and only if at least one keyword, qualifier or resource type semantically linked to this metaterm is major for the same CISMeF resource (otherwise, the metaterm is minor).
Aurélie Névéol’s PhD work

- Automatic Indexing (MeSH)
- Categorization (Metaterms)
- Extraction of French synonyms (from a FR/EN parallel corpus) (may be presented by Aurélie at LH NLM next October around AMIA meeting)
- Since Feb. 2006, postdoc NLM Aronson + NCBI
Aurélie Névéol’s PhD work

- MAIF is similar to MTI in that it is a combination of two MeSH indexing approaches:
  - a NLP approach, and
  - a statistical, knowledge-based approach.

- One originality of her work
  - Both approaches support the extraction of keyword/qualifier pairs from the resources
MeSH Automatic Indexing for French (MAIF)

- Developed by CISMeF (Rouen, France)

- Combines Natural Language Processing techniques and Statistical approaches

- Features:
  - MeSH-FR
  - Uses MeSH resources
  - Keyword or keyword/qualifier pair retrieval
  - Breakage function: selection of the most likely candidates
## Results

<table>
<thead>
<tr>
<th>Rk</th>
<th>MAIF-NLP</th>
<th>MAIF</th>
<th>MAIF 10-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54.90 - 9.71 - 16.50</td>
<td>37.25 - 6.14 - 10.54</td>
<td>21.95 - 3.49 - 6.02</td>
</tr>
<tr>
<td>3</td>
<td>40.47 - 28.67 - 33.56</td>
<td>34.61 - 21.71 - 26.68</td>
<td>14.56 - 8.05 - 10.37</td>
</tr>
<tr>
<td>T</td>
<td>38.56 - 35.00 – 36.69 (T=5.24)</td>
<td>27.20 - 36.06 – 31.01 (T=7.46)</td>
<td>12.35 - 23.18 – 16.11 (T=11.55)</td>
</tr>
</tbody>
</table>

Precision, recall and F-measure for MAIF at fixed ranks, and adaptive threshold $T$. 
Knowledge coupling

- Between knowledge bases
  - CISMeF to PubMed
  - CISMeF BP to NGC
  - CISMeF Patient to MEDLINEplus
  - And many more…

- From electronic patient to knowledge bases
  - Interoperability between medical terminologies
  - Business €€€
Strategic revolution in 2005

- Between 1995 & 2005, mono-terminological world around the MeSH
- Since 2005, shift to multi-terminological universe:
  - CCAM, CIM10, SNOMED Int., CIF/CIH, CISP2, DRC
  - Creation of a French Health Multi-Terminological Server (HMTS): ANR, InterSTIS
  - Multi-Terminological extraction (7th FP EU, PSIP)
  - Multi-Terminological Information Retrieval (JFIM 2009)
- Several health terminologies for the **automatic indexing** and the **information retrieval** in the CISMeF quality-controlled health portal... and beyond
- Can be reused in any European language if health terminologies are available in your language!!! In particular in Norway
Multi-Terminological extraction

- Collaboration with Vidal company
- F-MTI & RIMT tools
  - PhDs of Suzanne Pereira & Saoussen Sakji: PSIP Project
- Bag of words algorithm, stemming (or lemmatization)
- Inclusion of health terminologies available in French
  - SNOMED Int, ICD 10, MeSH, MeSH SC, ICDC (included in UMLS)
  - ATC, CIF (WHO)
  - CCAM, DRC, Orphanet, TUV, CIS, CIP, INN, Brand Names
  - MedDRA, WHO-ART, LOINC (to be included)
- Recent study on CISMeF corpus de CISMeF: MonoT vs. MultiT (AMIA 2009; in press): +7% recall; -12% precision
Multi-Terminological Information Retrieval

- RIMT using the same health terminologies, integrated to the CISMeF backoffice
  - Operational in Doc’CISMeF since April 2009 (test)
  - Bi-terminological in the PSIP DIP since September 2008
- Bag of words algorithm, stemming
- Double context
  - Knowledge (CISMeF) + contextual knowledge
    - PhD Saoussen Sakji 2Y (Tunisia)
  - Electronic Health Record (EHR)
    - PhD AD Dirieh-Dibaba 1Y (Djibouti)
- Evaluation fin 2009
  - E.g. : « laserbrasion » CCAM term
# Context of the MTIR

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Document Knowledge</th>
<th>Care (EHR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeSH</td>
<td>Manual XOR Auto</td>
<td>Auto</td>
</tr>
<tr>
<td>CIM-10</td>
<td>Auto</td>
<td>Manual OR Auto</td>
</tr>
<tr>
<td>CCAM</td>
<td>Auto</td>
<td>Manuel OR Auto</td>
</tr>
<tr>
<td>SNOMED IT</td>
<td>Auto</td>
<td>Auto</td>
</tr>
</tbody>
</table>
Information Retrieval in EHR

- PhD Dirieh-Dibaba 1Y
- Differences with IR in knowledge bases
  - Importance of the dates +++
  - Explicit query manual or automatic indexing
  - Mixture of unstructured data (e.g. information from discharge summaries) and structured data (e.g. biology)
    - Query: myocardial infarction [ICD 10] [manual] AND bilirubin > 20 AND same hospitalization (same hospitalization ID) XOR Delta dates < 5 days
- Use of Oracle semantic tools (SQL + SPARQL)
- Extension of the categorization to SNOMED
Creation of MT CISMeF backoffice
Indexation Multi-Terminologique

Financements

Projet PSIP Grant agreement n°216130

Projet ALADIN ANR-08-TECS-001

bronchite asthmatique chez l’enfant

Cet indexeur multi-terminologique est accessible gratuitement pour la communauté scientifique. Si vous êtes une entreprise privée, merci de contacter par courriel le Pr. Stéfan Darmoni de l’équipe CISMeF ou le Dr. Jean-François Forget, directeur scientifique du Vidal, avant toute utilisation.

Si vous l’utilisez, merci de citer un des papiers suivants :

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <interpretation>
  - <descripteurs>
    <des ter="CIM" id="CIM_LIB_4747" code="J45">asthme</des>
    <des ter="DRC" id="DRC_D_717" code="D717">asthme</des>
    <des ter="DRC" id="DRC_RCE_27" code="RCE_27">asthme</des>
    <des ter="MSH" id="MSH_D_001249" code="D001249">asthme</des>
    <des ter="MSH" id="MSH_D_001991" code="D001991">bronchite</des>
    <des ter="MSH" id="MSH_D_002648" code="D002648">enfant</des>
    <des ter="SNO" id="SNO_D_251000" code="D2-51000">bronchite asthmatique</des>
    <des ter="SNO" id="SNO_S_10170" code="S-10170">enfant</des>
  </descripteurs>
  - <expansions>
    <des ter="CCA" id="CCA/GLRP_001" code="GLRP001">GLRP001 - Séance de ré entraînement à l'exercice d'un enfant asthmatique, sur machine</des>
    <des ter="CIM" id="CIM_LIB_16709" code="J45.9">J45.9 - bronchite asthmatique SAI</des>
    <des ter="CIM" id="CIM_LIB_16695" code="J44.8">J44.8 - bronchite asthmatique (obstructive) SAI chronique</des>
    <des ter="CIM" id="CIM_LIB_27093" code="J44">J44 - bronchite asthmatique (obstructive) chronique</des>
    <des ter="SNO" id="SNO_D_230500" code="D2-30500">bronchite chronique obstructive</des>
  </expansions>
</interpretation>
94 ressource(s) trouvée(s) en 0,1 secondes, pour : Healthcare Common Procedure Coding System (mot réservé) ou Healthcare Common Procedure Coding System (titre) - Interprétation de la requête : ★★★

1. **Test Doppler pour la prédiction de la pré-éclampsie - CCAM : JQQM003 - JQQM002 - JQQM007 [ 2009 ]**

   ETSAD - L’évaluation des Technologies de Santé pour l’Aide à la Décision France
   "présentation, données d’évaluation, technologies et applications, état des lieux, réglementations et bonnes pratiques, éléments financiers et fournisseurs"

   - **MeSH**
     - JQQM003 - Échographie de surveillance de la croissance foetale avec échographie-doppler des artères utérines de la mère et des vaisseaux du foetus;
     - Échographie Doppler, *échographie prénatale;*
   - **accès :** http://www.etsad.fr/etsad/index.php?module=dmi&action=recap&p1=353
   - pertinence : 100%

2. **Endoprothèses dans le traitement des anévrismes et des dissections de l’aorte thoracique - CCAM : DGLF003 [ 2009 ]**

   ETSAD - L’évaluation des Technologies de Santé pour l’Aide à la Décision France
   "présentation, données d’évaluation, technologies et applications, état des lieux, réglementations et bonnes pratiques, éléments financiers et fournisseurs"

   - **MeSH**
   - **accès :**
1. **Cas de leucoencéphalopathies multifocales progressive observées chez des patients traités par Cellcept - [28/05/2008]**

   [Site éditeur: AFSSAPS Agence Française de Sécurité Sanitaire des Produits de Santé]

   "En accord avec l'agence Européenne du Médicament (EMEA) et l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS), Roche souhaite vous informer de nouvelles données de pharmacovigilance concernant CellSante (mycophénoléate mofétïl)...

   MeSH:
   - (L) antineoplasiqes et immunomodulateurs;
   - (L04) immunosupresseurs;
   - (L04A) immunosupresseurs (L04AA) immunosupresseurs selectifs;
   - (L04AA06) mycophenolique acide;

   *mots-clés:*
   - acide mycophénolique/analogiques et dérivés;
   - leucoencéphalopathie multifocale progressive/induit chimiquement;

   *substances:*
   - acide mycophénolique [mc]; immunosupresseurs [mc]; mycophénolate mofétïl [mc];

   *types:*
   - avis de vigilance sanitaire; résumé des caractéristiques du produit;

   *accès:*
   - http://afssaps.sante.fr/htm/t0/filltpsc/lp08503.htm

2. **CELLCEPT 250mg GELULE B/100 - [22/09/2008]**

   [Site éditeur: MEDISSIPAR - MEDicaments à DIspensation PARTiculières]

   prescription, première délivrance, renouvellement [France]

   MeSH:
   - (L) antineoplasiqes et immunomodulateurs;
   - (L04) immunosupresseurs;
   - (L04A) immunosupresseurs (L04AA) immunosupresseurs selectifs;
   - (L04AA06) mycophenolique acide;

   *mots-clés:*
   - acide mycophénolique/analogiques et dérivés;

   *substances:*
   - acide mycophénolique inclt. mycophénolate mofétïl inclt.
Many thanks

- Email: Stefan.Darmoni@chu-rouen.fr