Semantic methods and tools of the CISMeF team

Prof. SJ. Darmoni, MD, PhD

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

Email: Stefan.Darmoni@chu-rouen.fr
Strategic revolution in 2005

- Between 1995 & 2005, CISMeF (Catalog and Index of Health Resources in French; www.cismef.org) = mono-terminological world around the MeSH

- Since 2005, shift to multi-terminological universe:
  - Multi-Terminological information system
  - Multi-Terminological extraction (several AMIA presentations) *
  - Multi-Terminological Information Retrieval (submitted to AMIA 2010)

- 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

Situation in January 2012

- Multi terminology information retrieval
- Doc’CISMeF search engine
- European Health multi-Termino-Ontology Cross-lingual Portal EHTOP
- CISMeF Information System
- 32 health terminologies

Multi terminology automatic indexing
ECMT InfoRoute
Introduction

- To index manually with several terminologies, necessity to have an health multi-terminology portal, with several functionalities
  - ✓ Description
  - ✓ Navigation
  - ✓ Relations (=> role of semantic harmonization)
  - ✓ Access to resources
- When? When the MeSH thesaurus is not sufficient
  - ✓ e.g. to index “procedures evaluation” => necessity to use the CCAM (French equivalent to US CPT)
  - ✓ e.g. for rare diseases, use of Orphanet, SNOMED or MedDRA
  - ✓ 12% of resources are manually indexed with several terminologies
Example with the « rokitansky syndrome »

- the « rokitansky syndrome » is not a MeSH term, when using the current CISMeF tool (MeSH 2010)

Terminologie CISMeF = MeSH Browser in French, which gives access to MEDLINE with a French query*
http://terminologiecismef.chu-rouen.fr/

*Poster AMIA 2007.

Thirion, B et coll.. French MeSH Browser: a cross-language tool to access MEDLINE/PubMed
No answer in MeSH 2010
During 2009, in collaboration with 8 students-engineers from the INSA de Rouen, and with LERTIM & MONDECA, the CISMeF team has developed the Health Multi-Terminology Health Portal (PTS as the French acronym; HMTP as the English acronym).

Since 2007, modelization of a generic model to integrate main terminologies and ontologies available in French

The current health terminologies included in PTS are:
- MeSH (+ MeSH SC + CISMeF extension), SNOMED Int, CCAM, ICD10 & ICPC2 (InterSTIS project)
- ATC, ICF, WHO-ART, WHO-ICPS (WHO)
- DRC, MedDRA, MEDLINEPlus
- CIS, CIP, CAS, EC, INN, Brand Names, PSIP taxonomy, IUPAC, NCC-MERP (PSIP Project)
- Orphanet (rare diseases), LPP & Cladimed (medical devices)
- TUV (Vidal), FMA, SNOMED CT, LOINC
- ADICAP, NCIT (to be included)
HMTP generic model
Health Multi-Terminology Portal (HMTP; PTS)

- URL: http://pts.chu-rouen.fr/
- Access for humans and computers (Web services)
  - Since September 2010, daily used by CISMeF team to index manually and automatically Web resources
  - Since January 2011, MeSH is freely available (600 unique users per working day)
- Restricted access to the other terminologies (760 registred)
- Current objective: to become the European Health Terminology & Ontology Portal (EHTOP)
  - Multi-linguality (June 2011) ; European T/O (e.g. ATC)
- Cooperation with BioPortal: Clement Jonquet & Mark Musen
Interest of PTS

- **Access to MEDLINE**
  - Main application
  - Already exists for French; potential extension to 15 languages (mainly European languages: e.g. German need to translate 83 subheadings)
  - Taught in half of medical schools

- **Teaching**
  - Information science (Montreal)
  - Medicine (anatomy; rare diseases) in second year med students
    - Satisfaction = 7.5/10; easy to use = 5.5/10
Interest of PTS (cont.)

- Translation +
  - Many subscribers are translaters

- Audit of T/O
  - Exemple of HPO & Orphanet
Terminologies/Ontologies (n=43)

- MeSH : Medical Subject Headings (2011)
- CISMef : Catalogue et Index des Sites Médicaux Francophones : extension du MeSH
- SNOMED int : Systematized Nomenclature of Medicine (version 3.5, internationale)
- CIM-10 : Classification statistique International des Maladies et des problèmes de santé connexes (version 10)
- CCAM : Classification Commune des Actes Médicaux (v.22)
- CISP-2 : Classification Internationale des Soins Primaires (v.2)
- ATC : classification Anatomique, Thérapeutique et Chimique (2010)
- CIF: Classification Internationale du Fonctionnement et du handicap (2001)
- Cladimed : Classification de Dispositifs Médicaux (v.6)
- IUPAC: International Union for Pure and Applied Chemistry (2009)
- LPP : Liste des Produits et des Prestations (2011)
- MEDLINEplus : Thésaurus patients
- Orphanet : Classification des maladies rares (2010)
- WHO-ICPS : WHO – International Classification of Patient Safety (1.1)
Terminologies/Ontologies (2)

- PSIP taxonomy : (2010)
- Médicaments : Codes CIS, CIP, UCD, DCI, …
- FMA : Foundational Model of Anatomy (v.3.0 : 2009)
- VCM : Visualisation de Connaissances Médicales (2010)
- NABM : Nomenclature des Actes de Biologie Médicale (2011)
- GO : Gene Ontology (2011)
- SNOMED CT : Clinical Terms (2010), en cours de validation
- LOINC : Logical Observation Identifiers Names and Codes (2010), en cours de validation
- Interface Terminologies of prescription: biologie, radiologie, soins infirmiers, endoscopie...
- BNCI
  => Multi-disciplinary (n=3): since the PlaIR project (FEDER)
  - UNIT: Université Numérique Ingénierie et Technologies (2010)
  - IDIT : Institut du Droit International des Transports (2010)
  - Sanofi-Aventis MG : Market Glossary (first industrial application)
## Main figures

### May 2010

<table>
<thead>
<tr>
<th>Terminologies</th>
<th>Concepts</th>
<th>Synonyms</th>
<th>Definitions</th>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>&gt; 580 000</td>
<td>&gt; 840 000</td>
<td>&gt; 220 000</td>
<td>&gt; 1 200 000</td>
</tr>
</tbody>
</table>

### May 2011

<table>
<thead>
<tr>
<th>Terminologies</th>
<th>Concepts</th>
<th>Synonyms</th>
<th>Definitions</th>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>&gt; 980 000</td>
<td>&gt; 2 300 000</td>
<td>&gt; 220 000</td>
<td>&gt; 4 000 000</td>
</tr>
</tbody>
</table>

### May 2012

<table>
<thead>
<tr>
<th>Terminologies</th>
<th>Concepts</th>
<th>Synonyms</th>
<th>Definitions</th>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>1 570 301</td>
<td>3 683 023</td>
<td>192 815*</td>
<td>&gt; 4 891 423</td>
</tr>
</tbody>
</table>

* Deleting automatic MeSH definitions
Le Portail Terminologique de Santé (PTS) vous est proposé par l’équipe CiSMéF du CHU de Rouen (Liens Rouen) et l’INSA de Rouen, en collaboration avec le LERTIM de l’Université d’Aix-Marseille et la société MON. Le PTS inclut les principales terminologies de santé disponibles en français. Il a pour objectif de permettre la recherche en santé dans des bases de données spécialisées.


Cette interface PTS propose :
1. des fonctionnalités de recherche au sein des terminologies
2. un accès aux attributs de chaque descriptor comme les définitions ou synonymes éventuels, mais aussi intra-terminologiques
3. une navigation hiérarchique au sein de chaque terminologie
4. un accès aux ressources Internet (CiSMéF, PubMed, InfoRoute).
Syndrome de mayer-rokitansky-kuster-hauser
CISMeF InfoRoute

- Service on demand (SOD)
- French InfoButton « à la française »
- URL: inforoute.chu-rouen.fr
- Searchable by humans and computers (via Web services); integrated into HMTP/EHTOP
- Strategy for licencing
  - Free for the scientific community
  - Fee for industrials and for industrial applications of the scientific/public community
- Not yet compliant HL7 InfoButton
- Access in French and English to 50 sites (e.g. PubMed)
  - Semantic expansion (syn UMLS) for several Web sites (e.g. Google)
French InfoButton (Poster AMIA 2007)
Multi-terminology information retrieval (MedDRA SNOMED)

Did you mean rokitansky sequence OR "lay-er-okitansky-cluster-cause" syndrome?

1. **Mayer-Rokitansky-Kuster-Hauser Syndrome** (MAGIC Foundation)
   What is Mayer-Rokitansky-Kuster-Hauser Syndrome? Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome is an uncommon variation in the prenatal development of the female genital...
   www.magicfoundation.org/...ksy-kuster-hauser-syndrome.html - External Health Links

2. Genetics Home Reference: WNT4 Mullerian aplasia and ovarian dysfunction **NIH** (National Library of Medicine)
   WNT4 Mullerian aplasia and ovarian dysfunction? Biason-Lauber syndrome Mayer-Rokitansky-Kuster-Hauser-like syndrome Mullerian aplasia and hyperandrogenism Mullerian duct...
   ghr.nlm.nih.gov/...nt4mullerianaplasiaandovarianfunction - External Health Links

3. **Birth Defects** (National Library of Medicine)
   Birth defects are problems that happen while a baby is developing in the mother's body. Most birth defects...
Pour vous connecter, cliquez sur le bouton ci-dessous

Connexion

résultat généré en 1.116 s pour l'expression : asthme
Union of MeSH syn. and UMLS syn.
Semantic harmonization: mapping, alignment

Three methods employed

- **Conceptual**
  - ✓ Same CUI
  - ✓ Other relations: close match, BT-NT, NT-BT (SKOS)
  - ✓ On UMLS (n=11 included in HMTOP)

- **NLP**
  - ✓ More or less same algorithm of automatic indexing
  - ✓ Bag of words
  - ✓ on n*n-1/2 T/O (included in the HMTP), n=number of T/O

- **Statistical**
  - ✓ Co-occurrence matrix
  - ✓ CCAM-ICD10; CCAM-LPP (potential licencing and €€€)
Conceptual Same CUI

NLP
Bag of words
Automatic indexing & information retrieval

- Bag of words (since Aurelie Neveol)
- New concept: automatic affiliation of a subheading to a MeSH term
- Manual affiliation of a subheading to a MeSH Supplementary Concept (Evaluation to perform)
- Stoilo & Lewenstein distance (PhD Z. Moalla Y2)
- MeSH Indexing at the concept level and not anymore at the descriptor level => improve precision
  - Interesting for rare diseases and drug
  - JMLA 2012 (in press)
  - Affiliation of a subheading to a MeSH Concept
Results in the Doc’CISMeF search engine

- Use of multi-terminology indexing with SNOMED & MedDRA + MeSH indexing

Multi-terminology information retrieval

Multi-terminology manual indexing using PTS
Future work

- EHTOP
  - ICD10 & MeSH in European languages
  - Procedures & medical devices T/O

- Information retrieval on EHR (RAVEL project ANR TecSan)
  - Extension to *omics data: modelization to begin in Sept. 2012 (MSc)

- Interface Terminologies (TerSan project ANR TecSan)
  - Done for radiology (validation to begin with radiologists first)
  - Biology: Manual alignment IT Prestation (LIS) to LOINC; modelization of LOINC
  (new hierarchies to automatically generate mappings)
  - Endoscopy: in process
  - Nursing: in process
  - Inclusion of Billing reference terminologies (e.g. NABM in Biology); outside TerSan project

- Multi-lingual search engine (already multi-T/O)
Situation in few months

- **Generic search engine**
  - Web resources, terminologies, EHR

- **Multi terminology information retrieval EHR**
  - New operators (e.g. time, numerical)
  - Ravel ANR TecSan?

- **Multi-Termonology automatic indexing & semantic harmonization**
  - ECMT InfoRoute

- **Multi-lingual Health T/O Portal**
  - EHTOP

- **Interface terminologies**
  - TerSan ANR TecSan?

- **CISMeF Information System**
  - Multi-lingual
  - Multi-discipline
  - Multi-T/O
Teaching anatomy using HMTP/EHTOP

Use of the FMA ontology
French translation

Contextual access to Bioportal

Contextual access to CISMeF Information System (maintenance)
Rich relation

Question: what the other muscles of the forearm?
List of muscles of the forearm

- Abductor pollicis longus
- Extensor indicis
- Flexor carpi radialis breve
- Extensor carpi radialis brevis
- Pronator quadratus
- Pronator teres
- Supinator
- Flexor carpi ulnaris
- Flexor digitorum profundus
Palais osseux
Vue inférieure
Teaching rare diseases using HMTP/EHTOP

Use of the Orphanet thesaurus and its rich relations
Orphanet Disease: Marfan syndrome

French term: Syndrome de Marfan

English term: Marfan syndrome

Original code: 568

Specific attributes:
- Transmission
- Autosomique dominant
- Autosomal dominant
- Référence MIM
- QOMIM
- Prevalence
- 1-5 / 10 000
- Apparition
- French
- Etiology
- French

Contextual Link to Orphanet Web site

Contextual link to OMIM

Contextual query to PubMed
Health Multi-Terminology Portal

Orphanet Disease - Marfan syndrome

- Full tree
  - Orphanet Diseases tree
    - Rare developmental defect during embryogenesis
      - Marfan syndrome
        - MASS syndrome
    - Rare eye disease
      - Rare genetic eye disease
        - Genetic lens and zonula anomaly
          - Lens position anomaly
          - Marfan syndrome
            - MASS syndrome
    - Genodermatosis with an eye disease
      - Connective tissue disease with eye involvement
        - Marfan syndrome
          - MASS syndrome
    - Keratoconus
      - Syndromic keratoconus
        - Marfan syndrome
          - MASS syndrome
        - Lens and zonula anomaly
          - Lens position anomaly
          - Marfan syndrome
            - MASS syndrome
        - Rare genetic glaucoma
Name of the genes related to the Marfan syndrome
Question: what are the other rare diseases related to this gene?
In two clicks and two seconds, the answer to the question FIVE other rare diseases are related to this gene FBN1
Signs of the Marfan syndrome

What are the other rare diseases related to Glaucoma?
Multi lingual version of EHTOP

URL: www.ehtop.eu or www.hetop.eu

Click on Log in; id/pwd=fmauser/fmapass

Focus on European languages and Latin alphabet but not only… Japanese, Mandarin, Arabic, Hebrew

In Japanese, two terminologies available:
• MeSH
• ICD10
### Connexion

**Connecté**

Vous êtes connecté(e) avec le profil "museif". Vous avez désormais accès aux thésaurus liés à celui-ci.

**Mon PTS**

Vous pouvez modifier vos préférences PTS.

Vous pouvez aussi modifier vos données personnelles.

Vous pouvez aussi modifier votre mot de passe.

Vous pouvez consulter vos terminologies.
オオカモノマツ属 (MeSH Descriptor)

Description
Hierarchies
Relations
Resources

Preferred term(s)

- オオカモノマツ属

Original code:
D031110

Definitions:

MeSH definition

A plant genus of the family ASCLEPIADACEAE. Members contain phenanthro-indolizidine alkaloids.

Synonyms:

T_UP_MESH_SYNONYM

Specific attributes:

T_ATT_CISMЕF_SUBSTANCE
T_ATT_CISMЕF_URL
Terminologies choice
- Select all
- MeSH FTP (default)

Results
255 term(s) found in 975 ms.

ICD-10 (12)
ICPC-2 (1)
MeSH (19)

Original code:
D03110

Definitions:
MeSH definition
A plant genus of the family ASGLEPIADACEAE. Members contain phenoanthro-indolizidine alkaloids.

Synonyms: Suggest another synonym

Specific attributes:
T_ATT_CISMEF_SUBSTANCE
T_ATT_CISMEF_URL
Query in Japanese

Preferred term(s)
- 呼吸困難-発作性
  - Dyspnea, paroxysmal

Definitions:
A disorder characterized by sudden attacks of respiratory distress in at rest patients with HEART FAILURE and PULMONARY EDEMA. It usually occurs at night after several hours of sleep in a reclining position. When there is significant WHEEZING, it is called cardiac asthma.

Synonyms: 
- 呼吸困難-発作性
- Dyspnea, paroxysmal
- Paroxysmal dyspnea
- Cardiac asthma
- Paroxysmal nocturnal dyspnea
- Asthma paroxysmal nocturnal
### Terminologies choice
- Select all
- My HMTQ (default)

### Results
1 term(s) found in 514 ms.

**MeSH (1)**
- 呼吸困難-発作性

**MeSH Descriptor (1)**

### Relations (full)
#### Allowable qualifiers (36)
- History
- MeSH Qualifier
- Diagnosis
- MeSH Qualifier
- Ethnology
- MeSH Qualifier
- Nursing
- MeSH Qualifier
- Radiologic imaging
- MeSH Qualifier
- Veterinary
- MeSH Qualifier
- Blood
- MeSH Qualifier
- Diet therapy
- MeSH Qualifier
- Endocrinology
- MeSH Qualifier
- Pathology
- MeSH Qualifier
- Radiation therapy
- MeSH Qualifier
- Carbohydrate metabolism
- MeSH Qualifier
- Drug therapy
- MeSH Qualifier
- Genetics
- MeSH Qualifier
- Pharmacology
- MeSH Qualifier
- Rehabilitation
- MeSH Qualifier
- Chemicals
- MeSH Qualifier
- Economics
- MeSH Qualifier
- Immunology
- MeSH Qualifier
- Physiology
- MeSH Qualifier
- Surgery
- MeSH Qualifier
- Classification
- MeSH Qualifier
- Embryology
- MeSH Qualifier
- Metabolism
- MeSH Qualifier
- Preventive medicine
- MeSH Qualifier
- Therapy
- MeSH Qualifier
- Complications
- MeSH Qualifier
- Enzymology
- MeSH Qualifier
- Microbiology
- MeSH Qualifier
- Psychology
- MeSH Qualifier
- Ultrasonography
- MeSH Qualifier

#### T_REL_CIS_MT_MSH_T (2)
- Cardiology
  - CIOMP MeSH term
- Pulmonary disease (esp.)
  - CIOMP MeSH term

#### Automatic exact mapping(s) (from CIOMF team) (12)
- Cardiaca asthma (disorder)
  - SNOMED CT Concept
- Paroxysmal nocturnal dyspnea
  - SNOMED CT Concept
- Paroxysmal nocturnal dyspnea (finding)
  - SNOMED CT Concept

#### T_REL_UMLS_BTNT_LERTIM (5)
- Heart failure
  - ICD-10 Sub-category
- Paroxysmal nocturnal dyspnea
  - SNOMED CT Concept
- Shortness of breath
  - ICD-10 Sub-category

#### UMLS correspondence (same concept) (11)
- Paroxysmal nocturnal dyspnea
  - ICD-10 Sub-category
- Paroxysmal dyspnea
  - SNOMED CT Concept
- Paroxysmal dyspnea (finding)
  - SNOMED CT Concept

#### Exact mapping(s) from UMLS - LERTIM (1)
- Paroxysmal dyspnea
  - SNOMED CT Concept

#### Terminology (1)
- Terminology
Pubmed search results for "dyspnea, paroxysmal/therapy[MH] OR "paroxysmal nocturnal dyspnea"".

1. Catheter ablation for atrial fibrillation.
   Related citations

   PMID: 22166809 [PubMed - indexed for MEDLINE]
   Related citations

3. [Correct information about fetal movements saves life].
   PMID: 22165537 [PubMed - indexed for MEDLINE]
   Related citations
Marfan-syndrom (MeSH Descriptor)

Bevorzugte Begriff(e)
- Marfan-syndrom
- Marfan Syndrome
- Syndrome de Marfan
- Marfan-syndrom
- Sindrome de Marfan

Original code:
D008382

Begriffsbestimmungen:
MeSH definition
An autosomal dominant disorder of CONNECTIVE TISSUE with abnormal features in the heart, the eye, and the skeleton. Cardiovascular manifestations include MITRAL VALVE PROLAPSE, elongated DURA MATER, and dislocated LENS. Marfan syndrome is often associated with mutations in the gene encoding fibrillin, a major component of extracellular matrix.

Synonyme:
- Marfan's syndrome
- Syndrome, marfan
- Syndrome, marfan's
- Marfan
- Marfan's disease

Spezifische Eigenschaften:

CiSMeF URL
Many thanks

Email: stefan.darmoni@chu-rouen.fr
HMTP generic model