



# The terminology needs for evaluation of care pathways

17-23 October 2015  
Manchester, United Kingdom

Poster Number

N Tvardik<sup>(1)</sup>, C Bouvry<sup>(1)</sup>, I Kergourlay<sup>(2)</sup>, SJ Darmoni<sup>(2)</sup>, MH Metzger<sup>(1)</sup> WHO/CTS to insert

(1) Université Lyon 1, LBBE, France; (2) CISMef, Rouen, France

**Abstract** Electronic medical records (EHR) are a very important source of information for secondary use as epidemiological analysis and health services evaluation. The normalisation of the medical terms extracted from medical documents is an essential step of the preprocessing. Its quality partly depends on the medical terminologies available in the corresponding language. The objective of this presentation is to highlight the terminology needs for optimizing the data analysis of medical documents in the context of care pathway evaluation.

## Introduction

Different challenges were identified for secondary use of EHR through different research projects aiming at developing a generic IT solution for exploiting data extracted from French medical documents (URL: [www.synodos.fr](http://www.synodos.fr)). One of them is the medical language normalisation.

The objective of this presentation is to highlight the terminology needs for optimizing the data analysis of medical documents in the context of care pathway evaluation.

## Methods & Materials

**Source of data:** There are two different sources of data. The first was part of the research project ALADIN-DTH in which 1600 medical documents (discharge summaries, operative reports, consultation reports, etc.) were extracted and de-identified in four French university hospitals participating. The second source of data was part of the SYNODOS project and consisted of 300 medical documents of patients diagnosed with colon cancer in a French referral center for oncology care.

**Multi-terminology server:** For the standardization of medical language, a health cross-lingual multi-terminology portal HeTOP developed by the CISMef team is being used in this project (URL: [www.hetop.eu](http://www.hetop.eu)). It contains 57 health terminologies and ontologies, including 1 million health concepts in French and 1.7 million in English of which 48 were available for the project partners. From this terminology portal, a health extractor (ECMT V2) was developed to automatically code the health concepts included in the medical reports. These health concepts should be included in at least one of the 18 terminologies selected in ECMT V2.

**Gold-standard annotation:** A subsample of the medical documents was randomly selected for building a gold standard annotated corpus for the evaluation of the SYNODOS solution. The terms not coded or miscoded by the ECMT were manually coded using the 48 terminologies of HeTOP portal, with the help of MedIndex. MedIndex is an application developed by LBBE on R software and MS Access.

This application was developed for the annotation of medical concepts by semi-automatic methods, completing manually the ECMT V2 automatic annotation. The "gold standard database of facts" was populated by semi-automatic annotation of 246 medical records (learning dataset : 96 medical records - test dataset: 150 medical records) using MedIndex. The gold standard database of facts obtained in MedIndex is based on the conceptual model developed by LBBE. The model is organized on the classical care pathway of a patient in hospital. Different categories corresponding to the chronology of events were built: medical history, current clinical history, reasons for health care access, evolution, the results of the episode of care. For each of these categories, subcategories were defined. For example, for the category "current clinical history", the subcategories "chief complaints", "clinical examination", "paraclinical texts", "therapeutic procedures", "therapeutic management (drug therapy, nursing care...) were defined. Each medical term is stored in one of this subcategory.

## Results

The number of annotated medical terms coded in this corpus was 42,109 corresponding to 11,129 unique terms by category (10,670 unique terms in the whole database).

Table 1: Distribution of the terminologies used for the normalisation of medical concepts in each category of the knowledge base conceptual model

Category	ADI (%)	ATC (%)	BNC (%)	CCA (%)	CIS (%)	CLA (%)	DRC (%)	FMA (%)	ICD (%)	ICF (%)	ICN (%)	LNC (%)	MED (%)	MSH (%)	NCIT (%)	SNO (%)	OTH (%)	NOT (%)	Total (n)
Symptoms	2.0	0.0	0.8	0.1	0.3	0.0	1.4	2.3	1.2	0.3	1.8	0.9	4.6	28.5	2.0	28.1	1.9	23.7	2,342
Paraclinical exams and results																			
Physical examination	1.6	0.0	0.3	0.1	0.1	0.3	2.2	2.5	0.4	0.0	2.8	1.5	4.5	19.8	1.5	30.4	1.2	30.4	667
Bacteriological exam	0.0	0.0	0.0	1.1	5.6	0.0	3.3	0.0	0.0	0.0	0.0	5.6	4.4	27.8	3.3	25.6	0.0	23.3	90
Bacteriological results	0.0	0.0	2.6	0.0	0.9	0.0	0.9	1.7	0.0	0.0	0.0	2.6	0.9	49.1	0.9	20.7	1.7	18.1	116
Other biological exam	0.0	0.0	0.7	0.0	3.5	0.0	0.0	0.0	0.0	0.0	2.8	2.8	2.8	43.1	4.2	22.9	3.5	13.9	144
Other biological results	0.9	0.0	2.8	0.0	1.6	0.0	0.6	0.9	0.0	0.0	2.8	2.2	6.9	31.1	4.7	23.0	2.5	19.8	318
Histological exam	2.4	0.0	0.0	5.5	0.0	0.0	0.8	2.4	0.0	0.0	0.8	1.6	2.4	25.2	6.3	22.8	1.6	28.3	127
Histological results	9.7	0.0	0.8	0.0	0.3	0.3	0.3	1.0	0.0	0.0	2.0	0.5	0.8	18.4	4.9	29.2	1.3	30.7	391
Imaging test	1.6	0.0	1.0	1.0	1.6	0.0	1.0	1.9	0.0	0.0	3.5	0.6	3.8	34.3	1.6	13.8	12.2	22.1	312
Imaging results	4.1	0.0	0.1	0.1	0.3	0.2	0.9	5.5	0.7	0.1	2.0	0.8	2.6	15.4	4.4	26.7	1.4	34.9	1,333
Other paraclinical test	1.6	0.0	3.1	3.1	2.3	0.0	1.6	0.8	0.8	0.0	5.5	0.8	4.7	37.5	1.6	21.1	1.6	14.1	128
Other paraclinical results	1.2	0.0	0.6	0.3	0.3	0.0	0.0	3.8	0.6	0.9	1.7	1.2	7.0	19.8	3.8	26.5	0.9	32.0	344
Therapeutic management																			
Surgical procedures	1.4	0.0	0.2	5.9	0.6	1.1	0.4	3.3	0.2	0.0	1.9	0.4	3.3	19.9	2.6	26.7	0.9	31.1	935
Invasive non-surgical procedures	1.1	0.0	0.0	3.4	0.0	1.7	0.6	1.7	0.6	0.0	3.4	0.6	3.4	24.7	5.1	24.7	0.0	29.2	178
Other immediate therapeutic management	0.7	0.0	2.1	2.1	0.0	2.1	0.7	1.4	0.0	0.0	1.4	1.4	2.9	21.4	0.7	27.9	1.4	33.6	140
Medication	0.2	8.3	0.4	0.3	1.0	0.6	0.0	0.2	0.1	0.1	1.8	0.6	1.1	41.0	2.4	10.4	6.7	25.0	1,045
Nursing care	0.8	0.8	0.0	1.6	0.0	0.8	3.3	0.8	0.0	0.0	7.4	0.0	4.1	22.1	2.5	22.1	0.8	32.8	122
Physiotherapy management	1.4	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	7.5	0.7	0.7	19.2	1.4	20.5	0.0	45.9	146	
Therapeutic or preventive advices	0.0	0.0	0.0	3.1	0.0	0.0	0.0	0.0	1.5	0.0	9.2	3.1	0.0	36.9	3.1	10.8	0.0	32.3	65
Other therapeutic management	0.5	0.5	0.0	1.1	1.6	5.4	0.0	0.5	0.0	0.0	2.7	2.2	5.4	33.3	1.6	18.8	0.5	25.8	186
Expert medical advice	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0	33.8	5.4	21.6	0.0	31.1	74
Seeking health care	0.0	0.4	0.0	0.0	5.0	0.8	0.4	0.0	0.0	0.4	7.3	1.2	0.0	35.9	2.7	15.8	0.4	29.7	259
Socio-professional category	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0	0.0	24.3	16.2	13.5	2.7	40.5	37	
Other categories	1.9	0.1	0.3	0.1	0.9	0.1	1.5	1.7	1.5	0.1	2.6	0.8	4.6	34.0	2.8	24.6	2.0	20.3	1,630
Total (n)	219	92	60	93	91	45	108	245	72	19	272	108	395	3,086	319	2,679	252	2,974	11,129
Total (%)	2.0	0.8	0.5	0.8	0.8	0.4	1.0	2.2	0.6	0.2	2.4	1.0	3.5	27.7	2.9	24.1	2.3	26.7	100.0

Note: ADI: ADICAP (French histological terminology); BNC: BNCI; CCA: CCAM (French terminology for therapeutic procedures); ICP: ICPC; CIS: CISMef (French MESH extension); CLA: CLADIMED; DRC: French primary care terminology; ICN: ICNP; LNC: LOINC; Med: MedDRA; MSH: MESH; SNO: SNOMED; OTH: other terminologies; NOT: no code available