*MEDINFO 2010*
*C. Safran et al. (Eds.)*
*IOS Press, 2010*

*doi:10.3233/978-1-60750-588-4-252*

# Evaluation of a French Medical Multi-Terminology Indexer for the Manual Annotation of Natural Language Medical Reports of Healthcare-Associated Infections

**Saoussen Sakji[a], Quentin Gicquel[b], Suzanne Pereira[c], Ivan Kergourlay[a], Denys Proux[d], Stéfan Darmoni [a], Marie-Hélène Metzger[b]**

[a] *CISMeF, LITIS EA 4108 Institute of Biomedical Research, University of Rouen, Normandy, France*
[b] *University of Lyon 1 - CNRS UMR 5558, LBBE-UCBL, Lyon, France*
[c] *Vidal, Issy-les-Moulineaux, France*
[d] *XRCE, Meylan, France*

**Abstract**

*Background: Surveillance of healthcare-associated infections is essential to prevention. A new collaborative project, namely ALADIN, was launched in January 2009 and aims to develop an automated detection tool based on natural language processing of medical documents. Objective: The objective of this study was to evaluate the annotation of natural language medical reports of healthcare-associated infections. Methods: A software MS Access application (NosIndex) has been developed to interface ECMT XML answer and manual annotation work. ECMT performances were evaluated by an infection control practitioner (ICP). Precision was evaluated for the 2 modules and recall only for the default module. Exclusion rate was defined as ratio between medical terms not found by ECMT and total number of terms evaluated. Results: The medical discharge summaries were randomly selected in 4 medical wards. From the 247 medical terms evaluated, ECMT proposed 428 and 3,721 codes, respectively for the default and expansion modules. The precision was higher with the default module ($P_1$=0.62) than with the expansion ($P_2$=0.47). Conclusion: Performances of ECMT as support tool for the medical annotation were satisfactory.*

***Keywords***

Abstract and indexing, Cross-infection, Decision making, Computer-assisted, Semantic Mining

## Introduction

Surveillance of Healthcare-Associated Infections (HAI) is an important activity and a real burden in the context of control and prevention. The impact on patients' health and related healthcare cost are highly significant and a major concern even for the richest countries. Furthermore, some issues were addressed concerning the workload and costs generated by this surveillance.

Alternative methods based on automation of detection procedures were experimented in different facilities. In this context, the electronic health records (EHR) represent a unique opportunity for infection control practitioners (ICP) to automate manual processes. Few experiences of applying data and text mining techniques for monitoring adverse events are reported in literature [1-2]. Data mining offers methods that can recognize patterns in these large data sets and make them actionable: e.g. the Data Mining Surveillance System uses data from the clinical laboratory and hospital information systems to create association rules linking patients, sample types, locations, organisms, and antibiotic susceptibilities [1]. The Geneva team was testing a classifier to distinguish between HAI and non-HAI [3].

The objective of the overall ALADIN Project [5] is to develop an automated HAI detection tool based on screening French natural language documents and reports of the EHR, especially from discharge summaries. This project began in January 2009 and will last 3 years. For the evaluation of the tool, the gold standard used will be manual annotations of these medical documents by using a French medical multi-terminology indexing tool (French acronym: ECMT). The annotation will firstly provide the correspondence between terms used in current medical language and not directly available in the standardized terminologies and secondly it will provide standardised data for building algorithms of detection. The objective of this study was to evaluate the French ECMT tool, in terms of recall and precision for the annotation of natural language medical reports of healthcare-associated infections.

## Material and Methods

### Development of the HAI detection tool

The first step of the HAI detection tool was the development of a clinical questionnaire by the CNRS-UMR 5558 team (on a MS Access computer application named NosIndex). The purpose of this questionnaire is to enable ICPs from 4 French University hospitals (Lille, Lyon, Nice, Rouen) to manually collect all relevant medical terms (symptoms or diagnosis, medical intervention, medication, microorganisms, medical imaging…) from 2,000 medical reports. The ICPs will also enter their conclusions regarding the outcome: suspicion of HAI or not. The manual annotation of each medical report will be conducted independently by two different ICPs. In case of annotation discordance, the two ICPs will meet for a consensus procedure. The final decision regarding the outcome "suspicion of HAI or not" will serve as gold standard for the evaluation of the detection tool performances. From the 2,000 medical reports selected for this study, 400 HAI reports and 400 reports without HAI will be randomly selected for the evaluation of the detection tool. The gold standard used for this evaluation in terms of sensitivity and specificity will be the manual medical annotation.

**Building a French Health Multi-Terminology automatic indexer (ECMT)**

All relevant medical terms selected by the ICPs and entered in the questionnaire will be coded by using the tools developed by the CISMeF team, which include several health terminologies.

In 2005, the CISMeF team decided a strategic shift: from a mono-terminology approach based on the MeSH thesaurus to a multi-terminology approach based on the main health terminologies available in French. Some of them are integrated in the UMLS metathesaurus; some are not because they are French terminologies (e.g. CCAM, DRC, Orphanet). These terminologies have their respective objectives and may explain why the health sector is so rich in terminologies: the MeSH is devoted to documentation, SNOMED to describe patient records, ICD10 to epidemiology, CCAM to procedures, ATC to drugs, ICPC2 and DRC to general (or family) medicine, ICF to disability. To adapt the CISMeF information system to the new paradigm "Multi-Terminology approach", the CISMeF team had restructured the CISMeF database models to integrate as generically as possible the various health terminologies. These terminologies are also integrated into a Health Multi-Terminology Server (HTMS) [6], which is based on the ITM platform of the Mondeca private company (URL: http://www.mondeca.com/).

The new ECMT tool presented here is largely inspired by the CISMeF algorithm for information retrieval for the Doc'CISMeF search engine [9] and F-MTI [7], which is a multi-terminology automatic indexer, developed in collaboration with the Vidal Company. As most of indexing tools, it is language-dependant and works for the French language.

The EMCT tool has two query modules: one default module based on bag of words algorithm [7] and one expanded module based on textual indexing, using Oracle text indexing®. The information retrieval allows retrieving all the terminologies descriptors that contain the words of the query, with a minimal score, which depends on the number of query words. By construction, the expanded module will provide more results than the default module.

The ECMT tool has four main steps for the default module:

- Step 1: Query normalization

The unimportant words are removed from the initial query, and then the phonetic spelling and the stems of the remained significant terms are extracted and alphabetically sorted out.

- Step 2: Identification of the descriptors

For each terminology, 0 to (n-1) words from the set of the significant terms are removed in order to identify the descriptors.

| |
|---|
| Query Example: lead intoxication in France (intoxication du plomb en France)<br>step1: {intoxication, intoxi, 1toksikasion, France, franc, fr4s, plomb, plomb, plan}<br>step2: example for the thesaurus MeSH<br>-remove 0 word: identify the descriptor composed of {intoxication, france, plomb}<br>-remove 1 word: identify descriptors composed of {intoxication, france}, {intoxication, plomb}, {france, plomb}.<br>=> Identification of the descriptor "intoxication plomb"<br>-remove 2 words: identify descriptors composed of {france} |

- Step 3: Determination of affiliated descriptors (MeSH qualifiers) for each identified descriptor.
- Step 4: Supplement the query indexation by the MeSH indexation rules and pharmacological action rules (MeSH is the main thesaurus for indexing in CISMeF). For example, if the query is indexed by the supplementary concept "racecadotril", it will be indexed also by the descriptor "antidiarrheals" describing the pharmacological action.

The ECMT tool can be queried by a human but much more interestingly by any software application using a dynamic URL as long as this software application is connected to the Internet. In both cases, the tool is provides an XML answer which can be displayed for the human or integrated in the software application, eventually after a filtering process among the 9 available terminologies.

The ECMT algorithm is now available freely for the research community. It will become a commercial product thanks to the collaboration with the Vidal Company in the coming months. The URL for human (POST method) is: http://doccismef.chu-rouen.fr/Interpreteur.html. The URL for computer (GET method) is: http://doccismef.chu-rouen.fr/servlets/Interpreteur?Mot=enfant+asthmatique[2].

**Development of a software application (NosIndex) for the manual annotation of medical reports**

In the context of this project, a MS Access application (NosIndex) was developed by the CNRS-UMR 5558 team in order to collect the manual annotation of the medical reports. Figure 1 summarizes the process and exchanges between NosIndex and ECMT. When an ICP enters an original term

[1] The query in French means: asthmatic child

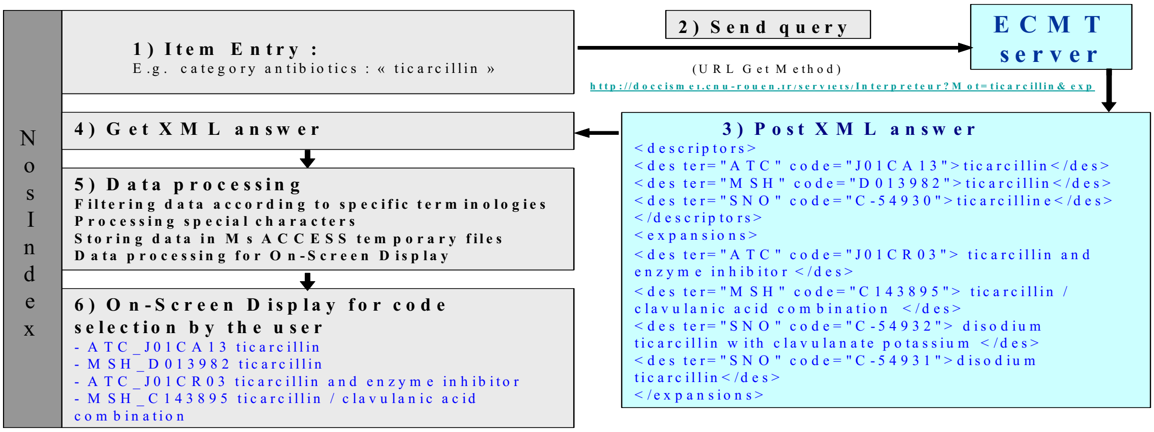


*Figure 1 - Process and exchanges between NosIndex and ECMT*

from the medical report into NosIndex by using a standardized clinical questionnaire, the application exports the request to the ECMT tool and imports the XML file containing all the corresponding codes. The process lasts 2 to 3 seconds before the ICP can visualize all the proposed codes. Then, the ICP can select the most appropriate one. If it does not find any satisfying code, the term will not be coded by the ICP and the term will be in a second step analyzed by semantic experts for decision. The structured questionnaire was developed on the application in order to classify all the medical terms relevant for implementing the detection algorithms of HAI.

The following categories (symptoms/diagnoses, bacteriological exams, type of microorganism, biological exams, radiological exams, antibiotics, and type of surgical intervention) were selected by HAI experts and have been coded specifically in order to limit the noise generated by the use of the 9 terminologies. (a) Symptoms/diagnosis: ICD10, SNOMED 3.5, MeSH; (b) Bacteriological exams: SNOMED3.5, MeSH; (c) Type of microorganisms: SNOMED 3.5; (d) Biological exams: SNOMED3.5, MeSH; (e) Radiological exams: SNOMED 3.5, MeSH, CCAM; (f) Antibiotics: ATC, MeSH; (g) Type of surgical intervention: CCAM, MeSH. The NosIndex application filters the results provided by ECMT, depending on the category of the medical terms. These filters were chosen manually by the expertise of the ALADIN project members.

### Evaluation methods of the ECMT tool

#### *Data sources*

Medical discharge summaries from patients with hospital acquired infection managed in intensive care units (n=42), orthopaedic surgery (n=12) and digestive surgery (n=8) of the Lyon University Hospital were randomly selected. An ICP evaluated the terms proposed by the ECMT when a medical term was entered in NosIndex. From these medical reports, the first 50 terms (not redundant) for symptoms/diagnosis and the first 30 terms for other categories were manually selected by an ICP. He entered the terms in NosIndex and evaluated each term proposed by the ECMT. The level of satisfaction of the term proposed was evaluated by using the following score: (a) 0 if no match; (b) 0.5 if approximate match (too precise or on the contrary too little accuracy); (c) 1 if perfect match.

#### *Evaluation of the ECMT default module*

We defined Recall $R_1$ as the number of relevant terms found by the ECMT default module divided by the total number of relevant terms provided by the two ECMT modules.

We defined precision $P_1$ as the number of relevant terms found by the ECMT default module divided by the total number of terms provided by the ECMT default module. $F_1$-measure was calculated as follows:

$$F_1 = \frac{2P_1R_1}{P_1 + R_1} \quad (1)$$

An average precision and average recall were calculated by category of medical terms. It is then possible to calculate an average precision by category, as the average of precision values for each term of this category (e.g. the average precision = $(P_1 + P_2 + P_i \ldots + P_n)/n$, with $P_i$ = precision for the term i and n = total number of terms evaluated for this category). It is also possible to calculate an average recall by category.
Then, an overall mean average precision and mean average recall were calculated, as followed:

- Mean average precision = average of the average precision for each category;
- Mean average recall = average of the average recall for each category.

The exclusion rate is defined as the ratio between the number of terms not found by the ECMT default module and the total number of terms evaluated. This exclusion rate was calculated by category and overall.

#### *Evaluation of the expansion performances*

We defined the precision $P_2$ as the number of relevant terms found by the ECMT expansion module divided by the total number of terms provided by the ECMT default module. Recall and F-measures were not calculated because no gold standard is available for determining the denominator. For example, the number of relevant terms existing in the multi-terminology indexing tool. An average precision was calculated by category. The same terminologies as for the ECMT default module were selected by category. The exclusion rate is defined as the ratio between the number of terms not found by the ECMT expansion module and the total number of terms evaluated. This exclusion rate was calculated by category and overall.

## Results

The results of the ECMT tool are summarized in Table 1. For the evaluation of the 237 medical terms of medical reports related to HAI, the number of terms proposed by the ECMT default module was 9 times less than the number proposed by the ECMT expanded module (428 vs. 3,721). Then, the overall precision $P_1$ (ECMT default module) was higher than the overall precision $P_2$ (ECMT expanded module) (0.62 vs. 0.47).

The overall F-measure was 0.59 with a minimum of 0.46 for the category "bacteriological exams" and a maximum of 0.71 for the category "antibiotics". The overall recall R1 was 0.58 but important variations were observed between categories (min: 0.45; max: 0.70). The only use of the default module could be satisfactory ($R_1$= 0.70) but for other categories and particularly for the type of surgical intervention ($R_1$ = 0.45), the expansion module improved consistently the number of relevant terms proposed for annotation.

The overall exclusion rate for the default module was 0.15. The expansion lowered significantly the exclusion rate of 0.15 to 0.11. By category, the expansion lowered significantly the exclusion rate for "radiological exams" (from 0.13 to 0.06) and for "type of surgical intervention" (from 0.22 to 0.09). Exclusion rate was rather small for four categories with both modules. The exclusion rate remained important for two categories: 0.26 for bacteriological exams and 0.23 for antibiotics.

*Table 1- Summarized results of the ECMT evaluation*

| | **Number of terms evaluated** | **Number of terms (ECMT default module)** | **Number of terms (ECMT expansion)** | **Evaluation "default module"** | | | **Evaluation "expansion"** | **Exclusion rate** | |
|---|---|---|---|---|---|---|---|---|---|
| | N | N | N | P1[3] | R1[4] | F-measure | P2[5] | Default module[6] | ECMT expansion[7] |
| Symptoms/ diagnosis | 50 | 141 | 1424 | 0.60 | 0.62 | 0.61 | 0.42 | 0,06 | 0.04 |
| bacteriological exams | 31 | 63 | 122 | 0.41 | 0.51 | 0.46 | 0.38 | 0,32 | 0.26 |
| type of microorganisms | 31 | 38 | 147 | 0.77 | 0.63 | 0.69 | 0.64 | 0,10 | 0.06 |
| biological exams | 30 | 43 | 639 | 0.77 | 0.60 | 0.67 | 0.43 | 0,07 | 0.03 |
| radiological exams | 32 | 55 | 989 | 0.53 | 0.53 | 0.53 | 0.38 | 0,13 | 0.06 |
| Antibiotics | 31 | 23 | 38 | 0.74 | 0.7 | 0.72 | 0.71 | 0,23 | 0.23 |
| Type of surgical intervention | 32 | 65 | 362 | 0.54 | 0.45 | 0.50 | 0.40 | 0,22 | 0.09 |
| Overall | 237 | 428 | 3721 | 0.62 | 0.58 | 0.59 | 0.47 | 0,15 | 0.11 |

[3] e divided by the total number of terms provided by the ECMT default module

[4] R1 = number of relevant terms found by the ECMT default module divided by the total number of relevant terms provided by the two ECMT modules

[5] P2 = number of relevant terms found by the ECMT expansion divided by the number of terms provided by the ECMT expansion

[6] Exclusion rate (default module) = = ratio between the number of terms not found by the ECMT default module and the total number of terms evaluated.

[7] Exclusion rate (expansion) = ratio between the number of terms not found by the ECMT expansion module and the total number of terms evaluated.

## Discussion

The main objective of this study was to evaluate the ECMT tool, in terms of F-measure, recall, precision and exclusion for the annotation of natural language medical reports of healthcare-associated infections in French. If the overall F-measure could be considered as satisfying (0.59), some efforts need to be made to improve the F-measure for certain categories (bacteriological exams, radiological exams and type of surgical interventions).

An important result of this study is the relative good overall precision $P_2$ (ECMT expanded module) compared to the overall precision $P_1$ (ECMT default module) (0.47 vs. 0.62), although the ECMT expanded module provides 9 times more terms than the ECMT default module (3721 vs. 428). Furthermore, the overall exclusion rate of the ECMT expanded module is rather small (0.11).

We need to improve the results for two categories: bacteriological exams and antibiotics. For the category "bacteriological exams", we plan to add a partial translation of LOINC.

For antibiotics, we have already translated in French 8,200 out more than 180,000 MeSH Supplementary Concepts [8]. We plan to go further. We have also planned to increase our cooperation with the Vidal company, leader of French drug database market.

This evaluation of the ECMT tool was done in a specific context, namely healthcare-associated infections, with the purpose to help the professionals who are not experts in standardized terminologies annotate medical reports. A software application can be easily developed as an interface for annotation of various medical topics (e.g. epidemiological studies, clinical research, adverse event surveillance…). By filtering relevant terminologies and by selecting default or expansion module depending on the medical category of terms, the investigators can find a satisfactory balance between inevitable noise and lack of precision.

## Conclusion

The aim of this study was to evaluate the performance of the ECMT as a support tool for the manual medical annotation. This study showed that the performance of the tool was good enough for helping ICPs to annotate medical reports with the different standardized terminologies. In parallel and in the context of the ALADIN project, the ECMT tool will also be used for the development of an automated HAI detection tool.

**Acknowledgments**

This work has been partly granted by the French National Research Agency (ANR) through the TecSan program (project ALADIN-DTH n° ANR-08-TECS-001)

## References

[1] Brossette SE, Hymel PA Jr. Data mining and infection control. Clin Lab Med. 2008 Mar;28(1):119-26, vii.

[2] Ma L, Tsui FC, Hogan WR, Wagner MM, Ma H. A framework for infection control surveillance using association rules. AMIA Annu Symp Proc. 2003:410-4.

[3] Cohen G, Sax H, Geissbuhler A. Novelty detection using one-class Parzen density estimator. An application to surveillance of nosocomial infections. Stud Health Technol Inform. 2008;136:21-6.

[4] Metzger MH, Gicquel Q, Proux D, Pereira S, Kergoulay I, Serrot E, Segond F, Darmoni SJ. Development of an Automated Detection Tool for Health Care-Associated Infections Based on Screening Natural Language Medical Reports. AMIA symp. 2009 (in press).

[5] Proux D, Segond F, Gerbier S, Metzger MH. Addressing risk assessment for patient safety in hospitals through information extraction in medical reports. In: Springer B, ed. Intelligent Information Processing IV. IFIP International Federation for Information Processing Vol. 288, 2009; 230-9.

[6] Joubert M, Dahamna J, Delahousse J, Fieschi M, Darmoni S. SMTS®: Un Serveur Multi-Terminologies de Santé. In: Risques, technologies de l'information pour les pratiques médicales, Informatique et santé, 17, Pages 47-56, 2009.

[7] Pereira S, Névéol A, Kerdelhué G, Serrot E, Joubert M, Darmoni S. Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. AMIA symp. 2008:586-590.

[8] Chazard E, Letord C, Beuscart MC, Kergourlay I, Darmoni SJ, Beuscart R. Data mining & semantic mining validation. PSIP deliverable D3.1. (internal report).

[9] Soualmia, L, Dahamna, B, Thirion, B, Darmoni SJ. Strategies for health information retrieval. Stud Health Technol Inform, 2006; 124: 595-600.

**Address for correspondence**

SJ. Darmoni, CISMeF, Rouen University Hospital, 1 rue de Germont, 76031 Rouen Cedex, France & TIBS, LITIS EA 4108, Institute of Biomedical Research, Normandy, University of Rouen, France