# Can F-MTI semantic-mined drug codes be used for Adverse Drug Events detection when no CPOE is available?

**Béatrice Merlin [a], Emmanuel Chazard [a], Suzan Pereira [b], Elisabeth Serrot [c], Saoussen Sakji [b], Régis Beuscart [a], Stefan Darmoni [b]**

[a] *Medical Information and Records Department EA2694, University Hospital, Lille, France*
[b] *TIBS, LITIS EA 4108, Institute of biomedical research, University Hospital, Rouen, France*
[c] *Vidal SA, France*

**Abstract**

*Background: Adverse Drug Events (ADEs) endanger the patients. Their detection and prevention is essential to improve the patients' safety. In the absence of computerized physician order entry (CPOE), discharge summaries are the only source of information about the drugs prescribed during a hospitalization. The French Multi-Terminology Indexer (F-MTI) can help to extract drug-related information from those records. Methods: In first and second validation steps, the performance of the F-MTI tool is evaluated to extract ICD10 and ATC codes from free-text documents. In third step, ADE detection rules are used and the confidences of those rules are compared in several hospitals: using a CPOE vs. using semantic mining of free-text documents, diagnoses and lab results being available in both cases. Results: The F-MTI tool is able to extract ATC codes from documents. Moreover, the evaluation shows coherent and comparable results between the hospitals with CPOEs and the hospital with drugs information extracted from the reports for ADE detection. Conclusion: semantic mining using F-MTI can help to identify previous cases of ADEs in absence of CPOE.*

***Keywords:***

Adverse drug events, electronic health records, semantic mining, F-MTI, medical reports, discharge summaries, terminology as topic.

## Introduction

Adverse drug events (ADEs) are situations where some drugs, eventually combined with other drugs, lab abnormalities or clinical context lead to adverse events. Those events endanger the patients and induce extra costs. Many of those events can be prevented thanks to appropriate prescriptions or monitoring. ADE study often relies on electronic health records (EHRs). In those EHRs, computerized physician order entries (CPOEs) are the most important components because they provide reliable and available information about drug prescriptions. Unfortunately, many hospitals do not have any CPOE at disposal. But in that case, the free-text documents usually contain the main drug prescriptions: discharge summaries, discharge letters, exam reports, etc. A semantic-mining tool can extract that information from the text. The objective of this work is to check if a semantic mining tool can be used to extract drug codes in order to identify previous ADEs and to compute the confidences of ADE detection rules.

A tool has been developed since 2005 by the CISMEF team (Rouen University Hospital) and the Vidal company [1] and is called "F-MTI" French Multi-Terminology Indexer (a generic automatic indexing tool able to index documentation in several health terminologies). The aim of this study is to evaluate the F-MTI's performances for ADE detection.

## Rationale

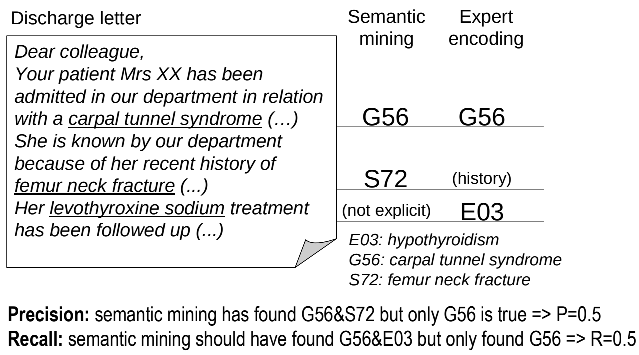


*Figure 1 - example of semantic mining applied on a discharge letter; precision and recall computation*

Semantic Mining is mainly oriented towards automatic indexing. For the evaluation of automatic indexing, different criteria can be measured, according to the literature [2-4]. The quality of the automatic indexing is evaluated by comparing the results of this automatic indexation (the candidate set) and the results of a gold standard (the gold standard set) on an evaluation dataset. The gold standard is the manual indexing performed by a human expert (Figure 1). For that purpose, different measures are commonly recognized as pertinent:

- **Precision (P)** is the number of indexing terms present in both candidate and gold standard sets divided by the total number of indexing terms in the candidate set. It measures the ratio of signal.
- **Recall (R)** is the number of indexing terms present in both candidate and gold standard sets divided by the total number of indexing terms in the gold standard set. It measures how well gold standard indexing terms are retrieved.
- **F-measure (F)** is the weighted harmonic mean of precision and recall. The traditional F-measure or balanced F-score is: $F = 2 * P * R / (P + R)$ where F is the F-measure, P is the precision and R is the recall.

Supplementary parameters were introduced to add a supplementary weight to precision or recall depending on the task that are to be evaluated:

- **Silence** corresponds to the proportion of terms not extracted (silence=1-Recall; false negatives).
- **Noise** corresponds to the proportion of false terms extracted by the system (Noise=1-Precision; false positives).
- **Purity** evaluates the proportion of indexation mistakes (extraction of a false term) avoided by the system.

In this study, three main metrics are calculated to show the performance of F-MTI indexing compared to the gold standard manual indexing: Precision, Recall, F-measure. These metrics are often used to evaluate the performances of automatic indexing tools [5-7].

## Materials and Methods

In order to evaluate the equivalence of Semantic Mining and complete EHRs including CPOEs for ADE detection, three complementary validation methods are applied.

The 10th revision of the International Classification of Diseases (ICD10) classification is used for diagnoses [8]. The Anatomical Therapeutic Chemical classification is used for drugs [9].

**Step 1- extraction of ATC codes from free-text documents: agreement between F-MTI and experts**

The aim of this first phase is to measure the accuracy of the extraction of the drug names included in the various free-text documents by means of the F-MTI Semantic Mining Analyzer.

Several de-identified discharge letters are obtained:

- 4,000 from the Rouen University Hospital (F), from which 50 are used for the validation task
- 10,000 from the Denain General hospital (F) , from which 32 are used for the validation task

The drug names extracted by automatic semantic mining (F-MTI) are compared with the ones obtained from human medical expertise (Figure 2).

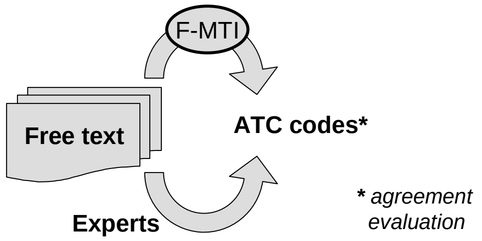


*Figure 2 - First validation step*

In the discharge letters, the drug names appear as brand or commercial names in 90% of cases, or as international names (INN). The list of brand names and INN names available in France are provided by the Vidal Company [1].

F-MTI indexing tool is used to extract the drug names and index them into ATC Codes: the results are gathered in the candidate set. The gold standard set is the result of the manual indexing performed by a human expert: the gold standard set. Human experts are a pharmacist and a medical archivist in Rouen; and two physicians in Denain.

In each free-text document, the Experts list

- the drug names recorded in the document (this is the "gold standard"),
- the drug names extracted by the F-MTI semantic tool.

Those lists are used to compute the precision and the recall.

**Step 2- extraction of ATC & ICD10 codes from free text: agreements between F-MTI and EHR**

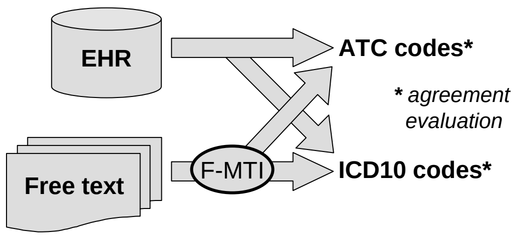


*Figure 3 - Second validation step*

In the Denain General Hospital, both the CPOE and the free-text documents are available. In this phase, the results of the semantic mining of the free-text documents (for the identification of the drugs prescribed or administered to the patient) are compared with the ones registered in the CPOE (Figure 3). This phase allows for computing the concordance between semantic mining analysis results and CPOE extraction for the identification of the drugs potentially linked with ADEs. This phase is only feasible in a hospital equipped both with a Hospital Information System containing the free-text documents and a CPOE System.

37 anonymized patients' complete electronic health records (EHRs) from the Denain General Hospital are used. Those records include:

- data from the EHR and the CPOE: ICD10 codes for diagnoses, ATC codes for drugs,
- the free-text documents and the results of the automatic indexing of these letters by Semantic Mining (F-MTI): ICD10 codes and ATC codes too.

The Method consists in the careful comparison of the codes obtained from semantic mining of the free-text documents with the codes contained in the EHR and CPOE. The comparison of drug codes (through ATC Classification) and the comparison of diagnosis codes (through ICD10 classification) are performed separately.

The so-obtained codes are compared. The recall R and the precision P are computed in each case.

**Step 3- validation of the use of the semantic mining results for data-mining-based ADE detection**

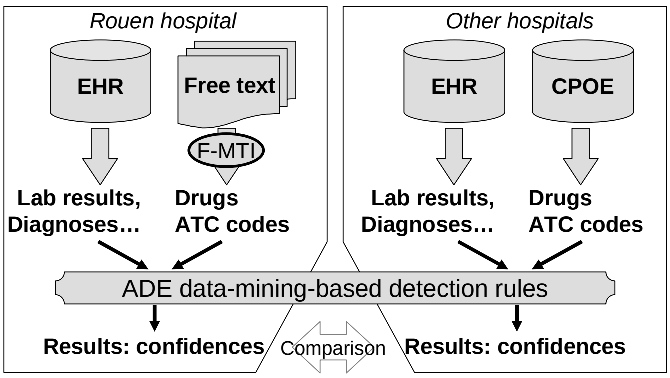


*Figure 4 - Third validation step*

This third validation phase consists in exploring the results of data-mining-based ADE detection rules when drugs are obtained from Semantic Mining of the various free-text documents, in case of absence of CPOE. This is done by studying the frequency of potential ADEs in the Rouen university hospital and comparing this frequency with the ones observed in hospitals where a CPOE is implemented (Copenhagen and Denain) (Figure 4).

The Material is represented by 245 data-mining-based detection rules obtained from various departments [10]. Those rules are a set of conditions that can lead to a traceable ADE. For each rule, the confidence is computed in Denain, Copenhagen and in the Rouen University Hospital where Drugs are obtained from Semantic Mining Analysis. Each rule is characterized by its confidence (1: proportion of outcome knowing that all the conditions are met) and its support (2: proportion of records matching both conditions and outcome).

$$\text{Confidence} = P(\ E \mid C_1 \cap \ldots \cap C_k\ ) \qquad (1)$$

$$\text{Support} = P(\ E \cap C_1 \cap \ldots \cap C_k\ ) \qquad (2)$$

The Method consists in the comparison of the confidences (positive predictive values) of the rules in the different places:

- the Rouen hospital where ATC codes are extracted from summaries,
- the other hospitals where ATC codes are extracted from CPOEs. The datasets from Denain and Copenhagen are pooled together to have only 2 datasets to compare. Moreover, pooling all the other datasets allows to get a better estimate of the confidence of the rules.

For each rule, all the stays that match the conditions of the rule are considered. The aim is then to test the independency between two binary variables using a Fisher's exact test:

- the occurrence of the effect (0 = "No" / 1 = "Yes"),
- the drug extraction method (CPOE/semantic mining)

For a given rule two results can be obtained:

- if p value $< 0.05$ then there is a significant difference between the confidence of the rule in Rouen and in other hospitals (the variables are not independent)
- if p value $> 0.05$ then no significant difference is observed between the confidence of the rule in Rouen and in other hospitals.

None of those results is interesting rule by rule. If significant p value is obtained for one rule, it is not surprising because the PSIP project showed that the confidences of the rules depend on the context in which they are used in (the patients, the practices and the knowledge are different) [11]. But if most of the rules look like having similar confidences in Rouen than in other places, it is an argument to say that the results of rules evaluation are consistent in Rouen compared with other hospitals.

## Results

**Step 1- extraction of ATC codes from free-text documents: agreement between F-MTI and experts**

The main results in the Rouen university hospital are:

- the overall Precision is **P = 0.84**
- the overall Recall is **R = 0.93**
- the F-measure is **F = 0.88**.

The main results in the Denain General Hospital are:

- the overall Recall is **R = 0.88**
- the overall Precision is **P = 0.88**
- the F-measure is **F = 0.88**

These results are coherent although the hospitals use different Hospital Information Systems, employ different physicians and take in care different populations of patients.

They appear as so successful as compare to the literature [12-14] particularly in the context of the French language where some particular difficulties have to be overcome (particularly negations, or some verbal passive forms).

**Step 2- extraction of ATC & ICD10 codes from free text: agreements between F-MTI and EHR**

***ATC codes extraction:***

The ATC codes from the semantic mining are considered as “candidates” while the ATC codes from the CPOE are given as “the “gold standard”. The results are (Table 2):

- the overall Recall is **R = 0.37**,
- the overall Precision is **P = 0.73**,
- the F-measure is **F = 0.49**

***ICD10 codes extraction:***

When the F-MTI tool is compared with the spontaneous encoding process, which is essentially based on economic considerations, the results are not as good as when compared with an expert encoding based on the free-text documents:

- the overall Recall is **R = 0.27**,
- the overall Precision is **P = 0.17**,
- the F-measure is **F = 0.21**

*Table 2 - Ability of F-MTI to replace EHR or CPOE codes*

| | Drugs: SM vs DB | Drugs: SM vs Experts | Diagnoses: SM vs DB |
|---|---|---|---|
| **Recall** | 37.4% | 88.4% | 26.7% |
| **Precision** | 72.6% | 88.4% | 17.3% |
| **F-measure** | 49.4% | 88.4% | 21.0% |

SM=semantic mining, DB=database

**Step 3- validation of the use of the Semantic Mining results for data-mining-based ADE detection**

The comparison between Rouen and other hospitals datasets is performed on each rule separately. Rule n°53 is provided as a detailed example.

<u>Rule N° 11:</u> Vitamin K antagonist (VKA) & Antiepileptic → Appearance of too low an International Normalized Ratio (INR; risk of thrombosis)

In that example, no significant difference is observed between Rouen and other hospitals pooled together.

The same method is applied on the 245 validated rules. A significant difference between the pooled confidence and the Rouen confidence can be observed in 50 rules (20.4% of the rules).

*Table 1 - comparison between Rouen and other hospitals*

| Object | Rouen | Other hospitals |
|---|---|---|
| Stays matching both the conditions and the effect: VKA & antiepileptic & too low INR | 2 | 43 |
| Stays matching the conditions: VKA & antiepileptic & no too low INR | 6 | 206 |
| Confidence | 33% | 21% |

*Fisher’s exact test: p=0.61 (independence)*

## Discussion

Predicting ICD10 codes is not an easy task when native data from the EHRs are used as the “gold standard” instead of an expert summary-based encoding. The ICD10 codes that are in the EHRs were most often encoded for economic objectives and include other information sources than the text documents. Moreover, the agreement between to experts is not so high [15].

Predicting ATC codes looks more successful although that task was performed on unstructured free text. Though mining the summaries poses problems. In discharge summaries, most often, only drugs previously taken by the patient and drugs prescribed at discharge are mentioned. In particular, some treatments only administered during the hospitalization (oxygen, pain killers, rehydration solutions, etc.) are never mentioned, which decreases the recall. Moreover the therapeutic information seems to be very rare when the patient has died during the hospitalization: there is no discharge treatment, and most often only clinical information is provided.

The F-MTI tool has to be improved. It encounters difficulties to recognize brand names in the discharge summaries due to identified problems that are currently being corrected.

Some additional problems are linked with incorrect spelling of the names in the discharge summaries. Some brand names are written improperly with dash ("-") or underscore ("_") or with an incorrect space " " (e.g. *di-antalvic*, *diffu k*, *di hydan*, *cacit D*, *calcidose vit D*, *co renitec*). On the contrary, some brand names are written without dash ("-") or underscore ("_") or space (" "), as normally they should have to (e.g. *chibroproscar* instead of *chibro-proscar*; *bi preterax* instead of *bipreterax*). Some other misspellings or mistyping are quite frequent (e.g. *triapridal* instead of *tiapridal*, *genopevaryl* instead of *gynopevaril*, *dextropropoxifene* instead of *dextropropoxyfene*, *piperacetam* instead of *piracetam*, *ketoderme* instead of *ketoderm*).

Some mistakes are redundant, e.g. the brand name is *cacit D3*. It is not automatically indexed and *cacit* is indexed instead of it. The same is occurring with *di-antalvic* & *antalvic* and *calcidose Vit D* & *calcidose*.

Some mistakes are more difficult to correct, as they refer to ambiguous terms. For instance in the lab results section of a discharge summary, *Albumin* refers to a lab result, while *Albumin* is also the brand name of a drug. This ambiguity will have to be handled.

## Conclusion

This validation task demonstrates that the F-MTI tool is able to identifying commercial and brand names of drugs in the free-text documents. The study allowed identifying courses of action to improve the tool.

A semantic mining tool is probably not able to automatically discover ADE prevention rules from previous hospital stays. It is not able to prevent ADEs as the discharge summaries and letters are always written after the end of the stay. Nevertheless, semantic mining of those documents can help to retrieve administered drugs in absence of CPOE in order to compute the confidence of ADE detection rules. Doing that, semantic mining of the free-text documents allows for ADE detection in former hospital stays.

### Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under Grant Agreement n° 216130 -the PSIP project [16-18].

## References

[1] Vidal S.A. [cited 2009 october 15]; Available from: http://www.vidal.fr/societe/vidal.

[2] Makhoul, John; Francis Kubala; Richard Schwartz; Ralph Weischedel: Performance measures for information extraction. In: Proceedings of DARPA Broadcast News Workshop, Herndon, VA, February 1999.

[3] Van Rijsbergen, C.V.: Information Retrieval. London; Boston. Butterworth, 2nd Edition 1979.

[4] Nakache D., Metais E., Timsit J. Evaluation and NLP. proceedings of DEXA Database and Expert System Application, 626–632, 2005.

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

[6] Névéol, A.; Pereira, S.; Kerdelhué, G.; Dahamna, B.; Joubert, M. & Darmoni, SJ. Evaluation of a Simple Method for the Automatic Assignment of MeSH Descriptors to Health Resources in a French Online Catalogue. Stud Health Technol Inform , U.S. National Library of Medicine, National Institutes of Health, Bethesda, USA., Volume 129, 407-411, 2007.

[7] Névéol, A; Rogozan, Aé & Darmoni, SJ. Automatic indexing of online health resources for a French quality controlled gateway. Information Processing & Management , May, Volume 42, Number 3, 695-709, 2006.

[8] International Classification of Diseases. [cited 2009 october 15]; Available from: http://www.who.int/classifications/icd/en.

[9] Anatomical Therapeutical Chemical classification. [cited 2009 october 15]; Available from: http://www.whocc.no/atcddd.

[10]Chazard E, Preda C, Merlin B, Ficheur G; PSIP consortium, Beuscart R. Data-mining-based detection of adverse drug events. Stud Health Technol Inform. 2009;150:552-6.

[11]Chazard E, Ficheur G, Merlin B, Serrot E, Beuscart R. Adverse Drug Events Prevention Rules: Multi-site Evaluation of Rules from Various Sources. Stud Health Technol Inform. 2009;148:102-11.

[12]Evans DA, Brownlow ND, Hersh WR, CampbellEM. Automating Concept Identification in the Electronic Medical Record: An Experiment in Extracting Dosage Information. AMIA 1996 Symposium Proceedings. 1996:388-92.

[13]Sirohi E, Peissig P. Study of Effect of Drug Lexicons on Medication Extraction from Electronic Medical Records. Pacific Symposium on Biocomputing; 10:308-18. 2005

[14]Gold S, Elhadad N, Zhu X, Cimino J J., George Hripcsak, Extracting Structured Medication Event Information from Discharge Summaries AMIA 2008 Symposium Proceedings 237-241

[15]Henderson T, Shepheard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. Med Care. 2006 Nov;44(11):1011-9.

[16]Patient Safety by Intelligent Procedures in medication. [cited 2009 october 15]; Available from: http://www.psip-project.eu.

[17]European Research Council. [cited 2009 october 15]; Available from: http://erc.europa.eu.

[18]Seventh Framework programme. [cited 2009 october 15]; Available from: http://cordis.europa.eu/fp7/home_en.html.

### Address for correspondence

Béatrice Merlin, service d'information et des archives médicales, CHRU Lille, 2 avenue Oscar Lambret, 59037 Lille Cedex, France