Natural Language Processing to detect Risk Patterns related to Hospital Acquired Infections

Denys Proux[1], Pierre Marchal[1], Frédérique Segond[1], Ivan Kergourlay[2], Stéfan Darmoni[2]
Suzanne Pereira[3], Quentin Gicquel[4], Marie-Hélène Metzger[4]

[1]Xerox Research Center Europe, Meylan, France
Denys.Proux@xrce.xerox.com, Pierre.Marchal@xrce.xerox.com, Frederique.Segond@xrce.xerox.com

[2]CISMeF, Rouen, France
Ivan.Kergourlay@chu-rouen.fr, Stefan.Darmoni@chu-rouen.fr

[3]Vidal, Issy-les-Moulineaux, France
suzanne.bento-pereira@vidal.fr

[4]Service d'Hygiène, Epidémiologie et Prévention des Hospices Civils de Lyon,
Hôpital Henry Gabrielle, Saint-Genis-Laval, France
quentin.gicquel@chu-lyon.fr, marie-helene.metzger@chu-lyon.fr

Abstract

Hospital Acquired Infections (HAI) has a major impact on public health and on related healthcare cost. HAI experts are fighting against this issue but they are struggling to access data. Information systems in hospitals are complex, highly heterogeneous, and generally not convenient to perform a real time surveillance. Developing a tool able to parse patient records in order to automatically detect signs of a possible issue would be a tremendous help for these experts and could allow them to react more rapidly and as a consequence to reduce the impact of such infections. Recent advances in Computational Intelligence Techniques such as Information Extraction, Risk Patterns Detection in documents and Decision Support Systems now allow to develop such systems.

Keywords

Natural Language Processing; Anonymization; Terminologies Mapping; Risk Pattern Detection.

1. Introduction

Patients' security is a key issue in hospitals. Specific prevention programs were developed in most of the European countries, including involvement of Infection Control Teams promoting prevention guidelines, control practices and implementing surveillance systems based on national standards.

Surveillance systems of adverse events are key elements for prevention as it has been demonstrated by various studies ([1], [2], [3] and [4]). An efficient surveillance system should meet several criteria: it should encompass clear definition of targeted infections, be able to detect and react in a very timely effective manner, be sensitive enough to detect small variations in the occurrence rate and should not require too much effort and time investment from the medical staff which is already overworked. Such a system should also be able to take into account a collection of data such as patient's risk factors (morbidity, invasive devices, surgical procedure…). These data have to be gathered from patient records to be recorded on specific standardized forms for further analysis. However the organization of hospital information systems does not help collecting this information.

Expertise gained over the last years in Computational Intelligence and more specifically in Risk Patterns detection from the literature allows now to address this problem. The detection of specific combinations of events and underlining relations between symptoms, treatments, drugs, reactions, and biological parameters can allow automatic systems to identify potential adverse events. Alerts could then be sent to risk management teams to help them identifying events that require immediate action and correction measures.

The following paper describes a project aiming to detect HAI by using risk patterns identification methods in patient records. The goal is to apply appropriate state of the art technologies included in a global process involving synergies between medical and technical experts to reduce the number of unnoticed cases and time for reaction. To do so Natural Language Processing (NLP) techniques will be applied to identify specific terms and sequences of facts in Patient Discharge Summaries.

2. Hospital Acquired Infections

2.1 Current Status

A Hospital Acquired Infection can be defined as: An infection occurring in a patient in a hospital or other health care facility in whom the infection was not present or incubating at the time of admission. This includes infections acquired in the hospital but appearing after discharge, and also occupational infections among staff of the facility. If the exact status of the patient is not clearly known when he first came in a medical unit, a

Workshop Biomedical Information Extraction 2009 - Borovets, Bulgaria, pages 35–41

period of 48 hours (or superior to the incubation period if it is known) is considered to separate HAI from other kinds of infections coming from outside. As for infections related to surgery a period of 30 days is considered and extended to 12 months in case of implanted device [5].

2.2 A Document workflow issue

HAI in hospitals is identified as a major issue and many multidimensional efforts have been undertaken to provide solutions to this problem. These solutions cover staffing, organizational and methodological dimensions.

Best practice guidelines have been designed and specialists assigned to provide guidance to the medical staff in case of infection surge. However many problems remain. They are related both to the difficulty to isolate HAI signs from other normal symptoms associated with what brings a patient to the hospital, and to the way information workflows are organized and accessible.

First of all detecting symptoms related to an HAI is inherently a difficult task as patients coming to a hospital are already sick. Furthermore some time they suffer from several diseases or infections at the same time that generate various different symptoms.

Time frame is also an issue because symptoms related to an HAI may take several days to appear. During that period a patient may have moved to different medical units and even may be back home. Information is therefore diluted in various documents covering several days or weeks.

The great heterogeneity of information systems adds to the task complexity. Each hospital can use its own tools to process and store data. Therefore it is extremely difficult for HAI experts to track down elements that could lead them to detect a problem, not to say to access documents in real time. This is why most of the time they react only when the issue is obvious and require urgent damage control actions.

2.3 Solutions to overcome this problem

Several directions to develop an efficient surveillance system are currently explored [6]. Among them we can identify several main categories:

- Passive systems that take into account what is declared by the medical staff or by patient themselves
- Systems based on a retrospective analysis made by HAI experts from patient records
- Predictive systems based on pre-identified risk factors
- Automated systems performing a systematic analysis on patient records

Most of them are decision support systems where rules have been designed thanks to human expertise or statistical data. The approach here is to compute the risk for a given patient to get a specific HAI according to various parameters such as age, gender, pathologies, medical unit where he is treated etc. But in this case it is only a prediction system ([7], [8], [9]).

Other techniques use microbiological analysis results to perform predictions. But here once again we are dealing with a statistical system.

However very few attempts have been made in the domain of text-mining to identify HAI risk factors ([10]). Melton *et al.* have for instance used the MedLEE semantic extraction tool to detect potential problem from patient records. The recall of this system has been evaluated to 28% and the precision to 98%. In this case the priority was to detect only very serious events which stressed the importance of precision. The same tool has also been applied to radiology reports to detect pulmonary infections [11]. In this case the recall was 71% and precision 95%. This evaluation stresses the important of tools customized for very specific targets in order to improve their efficiency. But more generally applying Natural Language processing technologies to detect from medical reports risk fact for HAI is a very promising trends where lot of work remains to be done.

3. Text Mining for Risk Patterns Detection

3.1 The ALADIN project

This project is developed in close collaboration between HAI surveillance experts and Linguistic and Knowledge Management experts in order to both characterize HAI risk factors and to design the necessary set of rules to identify such risk factors from patient records.

On a first hand only some specific medical units will be targeted, those where most deadly infections occur (Intensive Care Unit and Surgery).

The project agenda is divided into 4 major steps.

3.1.1 Selection of a corpus

1000 patient records reporting HAI and 1000 not dealing with HAI will be gathered from 4 French University hospitals. Patient records are written in French. These documents will deal with surgical activities (digestive, neurosurgery and orthopaedics) and Intensive Care Units. This step requires that all personal data should be removed (anonymized) from these documents. They will also be annotated before being moved outside these local hospitals.

3.1.2 Characterization of Risk Factors (adverse events and links between them)

This step will be done by HAI experts. They will work on patient records indexed by the Medical multi-Terminology Indexer server proposed by CISMeF. Links between these entities will be encoded as rules for the

Xerox Incremental Parser (XIP). The definition of risk factors, terminologies, and rules will be an interdisciplinary work between HAI experts and linguists.

3.1.3 Development of the detection tool
Detection rules applied during the parsing step will allow to find both specific Medical concepts and specific relations between them. This analysis will not be applied only at the sentence level but at the full patient record (which may contain several documents) level which implies to take into account complex combinations of events and a specific chronology.

3.1.4 Evaluation of the system performances in terms of precision and recall
For this step, new patient records will be analyzed (400 reports dealing with HAI and 400 reports without HAI). The gold standard will be the manual analysis of these patient records by two independent HAI experts.

3.2 Current work and experiments
In what follows we describe the work currently performed on the first step of the ALADIN project. This step implies the annotation of the corpus with semantic tags provided by a multi-terminology server and the development of an anonymization tool dedicated to patient records. We also provide in the remaining sections an overview of the work that will be performed to setup the risk pattern detection mechanisms.

In the context of the ALADIN project we use XIP to perform all Natural Language Processing tasks. This parser is robust that is to say it has already been used in various projects to process large collections of unrestricted documents (web pages, news, encyclopedias, etc.) This engine has been developed by a research team in computational linguistics. It has been designed to follow strict incremental strategies when applying parsing rules. The system never backtracks on rules to avoid falling into combinational explosion traps which makes it very appropriate to parse real long sentences from scientific texts for example [12]. The analysis is relying on three processing layers which are: Part of Speech Disambiguation, Dependency Extractions between words on the basis of sub-tree patterns over chunk sequences, and a combination of those dependencies with boolean operators to generate new dependencies or to modify or delete existing dependencies.

Figure 1 presents the results of a sentence parsed by this tool. It shows only syntactical dependencies, however it is also possible to apply more semantic rules based on classes of terms and dependency types to identify more complex information such as risk patterns.

"*Patient reaction to its treatment suggests a Clostridium Infection.*"

MOD_PRE(reaction,Patient)
MOD_POST(reaction,treatment)
VDOMAIN(suggests,suggests)
SUBJ-N_PRE(suggests,reaction)
PREPD(treatment,to)
SREL(Infection,infect)
SREL(reaction,react)
SREL(treatment,treat)
SREL(suggestion,suggests)
SREL(Infection,infect)
OBJ-N(suggests,Clostridium)
OBJ-N(suggests,Infection)

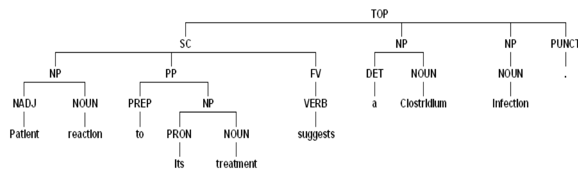

Figure 1. Identifying relations between key factors

4. Annotation and Terminology fusion

4.1 Objectives
In order to prepare the corpus and identify documents that deal with Hospital Acquired Infections, it is important to pinpoint inside texts valuable information that characterize HAI. This will be done manually by doctors. In order to standardize the type of information to be detected a guideline has been defined by HAI experts coordinating the medical part of the project. Information such as patient details, surgery type and symptoms is captured using standardized forms and terminologies.

This information is extracted manually (copy-pasted from original texts) but in order to prepare the design of the automatic indexing system (and also the automatic validation of the risk factor detection systems) we have to assign to each extracted information the related identification code coming from standard terminologies.

This step is made available thanks to the multi-terminology server developed in collaboration between CISMeF and Vidal. A specific annotation tool allows doctors to review document contents, to select useful information and to request a standard terminology tag for selected texts.

4.2 Fusion of Terminologies
The Multi Terminologies Indexer is a generic automatic indexing tool able to tag [13] an entire document with all terminologies necessary for the ALADIN project: SNOMED 3.5 (International Systemized Nomenclature of human and veterinary MEDicine), MeSH (Medical Subject Heading) , ICD10 (Classification of Diseases) and CCAM (French CPT), TUV (Unified Thesaurus of Vidal), ATC (Anatomical Therapeutic and Chemical

Classification), drug names with international non-proprietary names (INN) and brand names, Orphanet terms (rare diseases), CIF (International Functional Terminology), CISP2 (International Classification for Primary care), DRC (Consultation results), MedlinePlus.

This server provides a Web Service interface that allows terminology queries from a remote application through the Internet. These queries (sequence of words extracted from a text) are processed in order to remove all empty words, then normalized (stemmed), sorted then matched with available terminologies and filtered. The system proposes then all exact matching results and also expanded matching to help the user choose the most corresponding terms. If too many answers are possible than a ranking algorithm depending on the number of keywords searched is applied to filter out these results.

Tags selected by the user are then encoded along with the original word sequence extracted from the text, inside an XML data structure related to the processed document. This structured information will be used later to automate the comparison between information extracted by the risk assessment system and what should be detected.

5. Anonymization

5.1 Objectives

The anonymization step aims to detect and replace by appropriate values all data that make people identification possible. However, while there are English official list of types identifying relevant information, such list do not exist for French (as mentioned in Grouin *et al.* [18]).

Information to be anonymized is agreed upon with domain's specialist and CNIL (Commission Nationale de l'Informatique et des Libertés) the organism in charge of protecting personal data and private life.

These data are: people names (not only patient but also the medical staff), locations (hospital names, address), dates (birth, death, entry in a specific medical unit, ...) and all identification data such as phone number, room number, email address, ... This process is required by the fact that patient records have to be extracted from local hospitals to be centralized by the Service d'Hygiène, Epidémiologie et Prévention des Hospices Civils de Lyon which is the medical coordinator of the ALADIN project to work on the definition of risk factors. This is required by national regulation protecting anonymity. However beyond the scope of the ALADIN project, medical information sharing between hospitals or other healthcare organisms, or research projects also require to have a preliminary anonymization step. According to the amount of data to be processed, automated systems would be a tremendous help.

We have therefore reused in combination with XIP a set of rules ([16], [17] and [19]) already designed for named entity detection.

We have started with the named entity module developed at XRCE which recognizes the following types of entities: person names, dates, organizations names, places, events, email addresses as well as phone and fax numbers. This system has been evaluated in the context of the French ESTER2 campaign where it was ranked first in one of the evaluation exercise.

However, while it gave good general results, because of the idiosyncrasies of the patient records we had to do some customization work.

We have made a first customization of this set of rules to cope with the specificity of medical reports then performed a first evaluation on a short corpus to evaluate the system performance. These results gave us clues to improve our tools and to design its 2nd generation which is currently under development.

5.2 Targeted Named Entities

Anonymization faces a double challenge which is first to be able to detect specific Named Entities ([14], [15]) then to generate appropriate encoded terms that both remove any direct link to a specific person but also to keep the distinction between different persons.

In patient records, named entities often appear by themselves (in particular in the header). As a consequence, some names of person and places where not properly identified. For person name we have written new rules using list of first names with document case.

For place names, we have modelized the postal addresses of hospital relying on a lexical base of terms that could appear in hospital names (e.g. *CHLS*)

Because of the specificity of the application we also had to change distribution of semantic tags. For instance, *SAMU 38* in our application is important not because it is an organization but because it provides information about the place (first aid in Isère as 38 is the number of the French administrative division called Isère).

We also had to deal with the fact that patient records are often quickly written and contain typos. The most common ones are person names that are all written in lowercase. To solve this issue we have tagged as person name any unknown unit following a unit with the feature Title. However such a solution does not solve ambiguity cases such as *Monsieur gros (Mister big)*. More work on typos correction will be part of the next version of the system. The last issue concerns medical terms containing strings corresponding to named entities (e.g. *maladie de Parkinson*, *Glasgow 15*, where *Parkinson* and *Glasgow* should not be anonymized). At the moment we have written ad hoc rules but in the future we will have access to a specialized lexical database to deal with these cases.

The management of time stamps inside medical reports represents an important challenge. It is very important not only to detect all dates but to keep in the anonymization process the same chronology and time lap between each event. This chronology will be used latter both by experts to analyze the problem and by the ALADIN system to identify complex risk factors that involve a specific sequence of events. Our anonymizer takes as a starting point for the chronology the oldest date (which should not be a birth date) indicated inside the patient record and compute a new time stamp that embed chronological information referring to the starting date (e.g. T+14 means 14 days after Time 0). Birth date are not taken into account by this chronological recoding, however, in order to remove information that may help to identify the patient, any explicit birth date is replaced by the age of the patient computed with respect to the date when the report has been written.

5.3 1st experiment and results analysis

Based on this requirement, a first evaluation of our customized tool has been performed on a corpus of 5 patient records. The standard length of these documents is 4 pages and 1500 words.

The relatively small number of documents for this evaluation is due to the difficulty to access at this step of the project to patient records (as these documents should be anonymized before living hospital databases).

Table 1. Anonymisation results

	Nb	Recall	Precision
PEOPLE	108	96.7%	99.1%
LOCATION	52	85.9%	97.8%
DATE	123	95.2%	98.9%
PHONE/FAX/E-MAIL	65	100%	100%
TOTAL	**348**	**95.6%**	**99.2%**

First results show that among recognition errors some of them are related to spelling errors that corrupt the proper formatting of a name or number (e.g. *011 novembre* instead of *11 novembre*, *MonsieurDupont* instead of *Monsieur Dupont*).

It will be difficult to overcome this problem. Some names or location also appear in the text without a proper introducing or contextual disambiguation sequence (e.g. *à la Salpêtrière* instead of *à l'hôpital de la Pitié-Salpêtrière*). The improvement of propagation rules combined with more exhaustive location lexicons should cover partly this issue.

5.4 Improvements for the 2nd generation Anonymiser

A new experiment on a new set of documents has shown that most of remaining detection errors come from either complex location or people names appearing in the text or named entity without significant lexical context to allow disambiguation. In these cases several solutions are possible. They are currently being implemented in the new version of the anonymization tool.

In the future we plan to improve the system as follows:

- Take into account the most frequent causes of typos (e.g. missing character space between words or inverted characters into words).
- Add the event type to the list of possible named entity types. It is important as this type of entity carries information about place and/or date and can show up in patient records (e.g. *lors de la course à vélo L'Ardéchoise*).
- Fine grained entity types: at this stage the system just replaces person names and location names by their corresponding type. However, we have no finer grained indication (if the anonymized person name is a doctor or a patient name for instance).
- Take co-reference into account. We do not keep track of the different occurrences of the same person name in the text. In other words all occurrences of person names receive the same annotation independently of the fact that they have been already mentioned in other part of the document. This step is however important if we want to perform information extraction tasks in anonymized documents.
- Use an encrypted (for confidentiality reasons) local dictionary. It could allow the user to improve the efficiency of the system by adding new names that are not part of the original detection rules. This would definitely improve the detection rate on new documents.
- Development of a two layers anonymizer. The first layer replaces every named entity detected with a high level of confidence, the second layer applies more flexible rules on remaining untagged expressions to suggest possible replacements.

However, once again if the purpose is to have a 100% accuracy on this anonymization step, it is important to have a human validation at the end and therefore to display in a user friendly interface what has been anonymized and what remains to be done manually

6. Next steps

6.1 Characterization of Risk Factors

Once the corpus is anonymized and annotated the second step deals with the definition of what is a risk factor. This means defining what type of Named Entities should be retrieved, what types of link should relates these entities and what chronology should be respected to validate the risk. This work will be done in close collaboration between HAI and linguistic experts in order to encode syntactico-semantic related rules.

Figure 2 show the result of a POS tagging and Named Entity detection based on a sentence extracted from a patient record.

"*The postoperative consequences were marked by abdominal pain and fever due to multiple intra-peritoneal abscesses and peritonitis without anastomotic dehiscence that required a peritoneal toilet on September 29th of this year.*"

Part of Speech detected for each sentence tokens

The+DET postoperative+ADJ consequence+NOUN be+VBPAST mark+VPAP by+PREP abdominal+ADJ pain+NOUN and+COORD fever+NOUN due+ADJ to+PREP multiple+ADJ intra-peritoneal+guessed+ADJ abscess+NOUN and+COORD peritonitis+NOUN without+PREP anastomotic+guessed+ADJ dehiscence+guessed+NOUN that+PRONRE require+VPAST a+DET peritoneal+guessed+ADJ toilet+NOUN on+PREP September+PROP 29+ORD of+PREP this+DET year+NOUN .+SENT

Figure 2. POS tagging

Applying now XIP to the same text enables the system to detect chunks of related words. In particular the parser extracts the following chunks from the above sentences.

MOD_PRE_[1593]_[2108](consequence,postoperative)
MOD_PRE_[1593]_[2108](pain,abdominal)
MOD_PRE_[1598]_[2108](abscess,multiple)
MOD_PRE_[1598]_[2108](abscess,intra - peritoneal)
MOD_PRE_[1593]_[2108](toilet,peritoneal)

Figure 3: Chunks Detection

Now the combination of extracted syntactic dependencies with specific terminology tags assigned by the multi-terminology server allows the parser to compute pertinent semantic dependencies.

SYMPTOM(postoperative consequences)
SYMPTOM(abdominal pain)
SYMPTOM(fever)
DIAGNOSIS (multiple intra-peritoneal abscesses)
DIAGNOSIS(peritonitis)
DIAGNOSIS(infection)
PROCEDURE(peritoneal toilet)
TREATMENT(Tienam)
BACTERIA(Klebsiella)

Figure 4. Named Entity Detection

This extracted information can now be used to find possible matches with HAI scenarios.

6.2 HAI scenarios

Once texts have been parsed, pertinent named entities detected and semantic dependencies computed between these entities (at the sentence level), the next step is to find possible match between these dependencies and HAI scenario defined by experts.

What needs to be identified at this step is high level information such as: who is the patient, what are the treatments involved, what symptoms are detected, are characteristic adverse events terms appearing inside the text (e.g. name of a virulent bacteria).

And more than just a detection of isolated pieces of information it is important to be able to recognize specific sequences of events such as: what was the situation at the beginning, what analysis have been made to the patient, what treatments have been provided (e.g. which specific combination of drugs), what are the reactions to these treatments

The connection between these elements is important because according to their order it may characterize an HAI or just a normal case.

Finally one last thing that should be considered for scenario matching is flexibility. This should be taken into account because most of the time HAI are not clearly indicated inside texts. Some pieces of information are more important than others and are significant enough by themselves to characterize an HAI such as the name of a given bacteria (e.g. *infection with Klebsiella*). In other cases it is a combination of less significant pieces of information that all together allows to characterize an HAI, such as the use of a specific type of antibiotic drug in a specific department (e.g. *tienam* and *Intensive Care Unit*). It is therefore necessary to define scenarios with associated level of importance for information branches in the graph for appropriate decision making (figure 5).

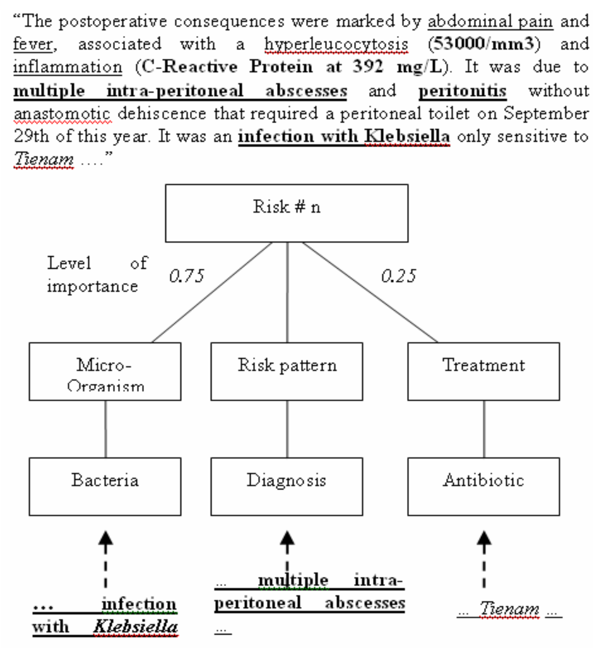

Figure 5. HAI scenario

These scenarios will be defined in the last part of the ALADIN project.

7. Conclusion

In this paper we presented a project that develop an Information Extraction system based on Natural Language Processing techniques to mine patient records to detect HAI risks. This project benefits from a strong collaboration between HAI surveillance experts to formalize HAI scenarios and linguists to convert this knowledge into detection rules for a semantic parser. This collaboration is achieving already the first step of the project which is the corpus preparation thanks to the development of appropriate anonymization and annotation tools.

8. Acknowledgement

ALADIN is a 3 year project funded by the French *Agence Nationale de la Recherche* (National Research Agency - ANR) in the context of the TecSan (*Technologies pour la Santé et l'Autonomie*) program. We also want to thank Caroline Hagège for her contribution to this paper.

9. References

[1] R.W. Haley, J.W. White *et al.* The Efficacy of Infection Surveillance and Control Programs in Preventing Nosocomial Infections in US Hospitals. Am J Epidemiol, 1985; 121:182-205.

[2] R. Condon, W. Schulte *et al.* Effectiveness of a Surgical Wound Surveillance Program. Arch Surg, 1983; 118:303-7.

[3] S.D. Bärwolff, C. Geffers, C. Brandt, R.P. Vonberg *et al.* Reduction of Surgical Site Infections after Caesarean Delivery Using Surveillance. J Hosp Infect, 2006; 64:156-161.

[4] P. Gastmeier, C. Brandt, I. Zuschneid, D. Sohr *et al.* Effectiveness of a Nationwide Nosocomial Infection Surveillance System for Reducing Nosocomial Infections. J Hosp Infect, 2006; 64:16-22.

[5] J.S. Garner, W.R. Jarvis, T.G. Emori *et al.* CDC Definitions for Nosocomial Infections. Am J Infect Control, 1988; 16:128-40.

[6] R. Amalberti, Y. Auroy, P. Michel, R. Salmi, P. Parneix, J.L. Quenon, B. Hubert. Typologie et Méthode d'Evaluation des Systèmes de Signalement des Accidents Médicaux et des Evénements Indésirables. Revue sur les Systèmes de Signalement, Rapport d'Etape du Contrat Mire-DRESS, 2006.

[7] V. Sintchenko, E. Coiera. Decision Complexity Affects the Extent and Type of Decision Support Use. AMIA Symposium 2006, 2006; ():724-8.

[8] C.A. Schurink, P.J. Lucas, I.M. Hoepelman, M.J. Bonten. Computer-Assisted Decision Support for the Diagnosis and Treatment of Infectious Diseases in Intensive Care Units. Lancet Infectious Diseases, 2005; 5:305-12.

[9] C. Chizzali-Bonfadin, K.P. Adlassnig, W. Koller. MONI: an Intelligent Database and Monitoring System for Surveillance of Nosocomial Infections. Medinfo, 1995; 8(2):1684.

[10] H.J. Murff, A.J. Forster, J.F. Peterson, J.M. Fiskio, H.L. Heiman, D.W. Bates. Electronically Screening Discharge Summaries for Adverse Medical Events. J Am Med Inform Assoc, 2003; 10(4):339-50.

[11] J.P. Haas, E.A. Mendonca, B. Ross, C. Friedman, E. Larson. Use of Computerized Surveillance to Detect Nosocomial Pneumonia in Neonatal Intensive Care Unit Patients. Am J Infect Control, 2005; 33(8):439-43.

[12] S. Aït-Mokhtar, J.P. Chanod. Incremental Finite-State Parsing. Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP'97), 1997; ():72-9.

[13] S. Pereira, A. Névéol, G. Kerdelhué, E. Serrot, M. Joubert, S. Darmoni. Using Multi-Terminology Indexing for the Assignment of MeSH Descriptors to Health Resources in a French Online Catalogue. AMIA Symposium, 2008; ():586–90.

[14] T. Poibeau. Sur le Statut Référentiel des Entités Nommées. Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2005), 2005.

[15] L. Plamondon, G. Lapalme, F. Pelletier. Anonymisation de Décisions de justice. Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2004), 2004; ():367-76.

[16] C. Brun, C. Hagege. Intertwining Deep Syntactic Processing and Named Entity Detection. In Proceedings of the 4th International Conference, EsTAL 2004, Alicante, Spain, October 20-22, 2004.

[17] C. Brun, M. Ehrmann, G. Jacquet. A Hybrid System for Named Entity Metonymy Resolution. In proceedings of the 4th International Workshop on Semantic Evaluations (ACL-SemEva). Prague, June 23-24, 2007.

[18] C. Grouin, A. Rosier, O. Dameron, P. Zweigenbaum. Une Procédure d'Anonymisation à Deux Niveaux pour Créer un Corpus de Comptes Rendus Hospitaliers. In Risques, Technologies de l'Information pour les Pratiques Médicales, 2009.

[19] C. Brun, M. Ehrmann. Adaptation of a Named Entity Recognition System for the ESTER 2 Evaluation Campaign. In 2009 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE'09) Proceedings, 2009.