

Automatic computation of CHA2DS2-VASc score: Information extraction from clinical texts for thromboembolism risk assessment

Cyril Grouin, MSc¹, Louise Deléger, PhD¹, Arnaud Rosier, MD^{2,3,4},
Lynda Temal, PhD^{2,3}, Olivier Dameron, PhD^{2,3}, Pascal Van Hille, MSc^{2,3},
Anita Burgun, MD, PhD^{2,3}, Pierre Zweigenbaum, PhD¹
¹ LIMSI-CNRS, F-91043 Orsay, France ² INSERM U936, F-35000 Rennes, France
³ Université de Rennes 1, U936, F-35000 Rennes, France ⁴ Groupe Hospitalier de
l'Institut Catholique Lillois, Faculté Libre de Médecine, F-59000 Lille, France

Abstract

The CHA2DS2-VASc score is a 10-point scale which allows cardiologists to easily identify potential stroke risk for patients with non-valvular fibrillation. In this article, we present a system based on natural language processing (lexicon and linguistic modules), including negation and speculation handling, which extracts medical concepts from French clinical records and uses them as criteria to compute the CHA2DS2-VASc score. We evaluate this system by comparing its computed criteria with those obtained by human reading of the same clinical texts, and by assessing the impact of the observed differences on the resulting CHA2DS2-VASc scores. Given 21 patient records, 168 instances of criteria were computed, with an accuracy of 97.6%, and the accuracy of the 21 CHA2DS2-VASc scores was 85.7%. All differences in scores trigger the same alert, which means that system performance on this test set yields similar results to human reading of the texts.

Introduction

Unstructured clinical notes contain a wealth of information, some of which is absent from the structured part of the electronic patient record. This has motivated a long-running stream of work on natural language processing applied to clinical texts.^[1,2,3] The bottom line of that work is the detection of information elements in clinical texts, encompassing the main types of concepts involved in clinical practice: diseases and other problems,^[4] tests, treatments including medication,^[5] and their component concepts such as anatomy.

A second level of target information aggregates this elementary information to compute clinical indicators for a patient, for instance their smoking status,^[6] obesity status,^[7] or the presence of a congestive heart failure (CHF).^[8,9] Such information is useful for a variety of applications such as triggering alerts or recruiting patients for clinical trials.^[8]

Project Akenaton^[10] addresses the extraction of medical information from French free text patient reports in the domain of cardiology, focusing on patients who have a pacemaker. In this domain, a key information element is the CHA2DS2-VASc score,^[11] a new recommendation of the European Society of Cardiology. It has been proposed to determine the stroke risk for patients with non-valvular fibrillation.^[12] This score is computed from eight criteria. Each criterion counts for 1 or 2 points in the final score: (i) Congestive heart failure or left ventricular dysfunction, 1 pt, (ii) Hypertension, 1 pt, (iii) Age \geq 75, 2 pts, (iv) Diabetes mellitus, 1 pt, (v) Stroke, transient ischemic attack, or thromboembolism, 2 pts, (vi) Vascular disease (prior myocardial infarction, peripheral artery disease, aortic plaque), 1 pt, (vii) 65 \leq Age $<$ 75, 1 pt, (viii) Sex category, 1 pt if female gender. The final computed score varies from 0 to 9. In conclusion, the higher the score, the higher the risk of thromboembolism. This CHA2DS2-VASc score constitutes one of the elements taken into account when a clinician has to decide whether an anticoagulation therapy is required to prevent potential stroke. Some criteria such as age and sex can be extracted from the hospital information system, while others need a careful analysis of clinical documents. To the best of our knowledge, no previous work has addressed the automatic computation of this score based on clinical texts.

Furthermore, this task is harder than, e.g., computing a score such as the body mass index (BMI): whereas BMI only needs to find the two numeric measures of body weight and height, the CHA2DS2-VASc score needs to assess the presence or absence of concepts that have a more complex definition. For instance, the concept of peripheral artery disease (PAD) encompasses a range of diseases that are more specific than the generic description called PAD. Ontological knowledge is therefore useful to help link the specific disease names that can be found in patient reports to the generic detection of presence or absence of a peripheral arterial disease. We use a module^[13] which computes the CHA2DS2-VASc formula based on concepts in an ontology,^[14] which allows us to detect the presence of the relevant

concepts at the right level of abstraction. The present paper focuses on the natural language processing part of the CHA2DS2-VASc computation pipeline. More detail on the ontology-based module can be found in Dameron *et al.*^[13]

Since the score depends on the presence or absence of concepts for a given patient, a proper treatment of negation and other expressions of modality is important to avoid false detection of negated concepts. Negation and probability processing can be handled in English clinical texts by Chapman's NegEx system.^[15] A more complete detection of other modalities was addressed in the i2b2/VA 2010 challenge^[4] task on the detection of *assertions* on medical problems, which additionally included categories *conditional*, *hypothetical* and *not associated with the patient*. However, NegEx works on English text, and there is no equivalent system available for French clinical texts. We therefore designed a version of NegEx extended to the i2b2 assertion categories and adapted to French.

In this paper, we present the system we designed and implemented to extract medical concepts from clinical records in cardiology, within the framework of thromboembolism risk. We then present its evaluation through a comparison of human reading *vs.* automatic extraction, focusing on the criteria that are used to compute the CHA2DS2-VASc score. We conclude with a discussion of current results and perspectives for further work.

Background

Previous work has addressed the classification of patients into predefined classes: such classes can be binary, e.g. CHF *vs.* non-CHF,^[8,9] or involve a choice among a few disjoint categories, e.g. five categories of smoking status (past smoker, current smoker, etc.).^[6] In such cases, the problem can be modeled as a supervised classification task where the system must predict the correct class based on features which represent the patient cases. These features are extracted from the texts by natural language processing methods with a varying degree of sophistication, including some handling of negation.^[9]

A characteristic of most work on the classification of patient information is the use of flat lists of terms, often obtained from existing controlled vocabularies (possibly through the UMLS Metathesaurus^[16]), aggregated in such a way as to detect each relevant feature. This was the case for instance in the i2b2 medication extraction challenge,^[5] where most systems used lists of drug names to detect medications and lists of findings and disease names to find the reason for a prescription. It would be possible to adopt the same approach to detect, e.g., peripheral arterial diseases. However, we opted for a more principled approach in which the natural language processing component deals with the recognition of specific concepts in the texts, and a separate ontological component decides whether a specific concept (e.g., "artérite des membres inférieurs" (*lower limb arteritis*)) is an instance of a generic concept (e.g., *peripheral arterial disease*) which instantiates a criterion or not. This avoids the overspecialization of concept detection in the texts, and allows the NLP component to output a more versatile representation where concepts can be used for a larger variety of tasks beyond the specific question of CHA2DS2-VASc score computation.

The supervised classification approach used by many recent systems also loses much relevance in the case of the computation of a formula. In the case of the CHA2DS2-VASc score, the number of possible values (10 values, from 0 to 9) would make it more difficult for supervised classification to take these values as discrete target classes. Besides, the mathematical relation of the score to the underlying features would be lost by a straightforward application of classical feature-based supervised learning methods. We found no compelling argument not to rely on the existing formula to compute the CHA2DS2-VASc score once the basic information elements (or features) are determined, that is, to apply a knowledge-based approach to that task. Our knowledge-based approach to the computation of the CHA2DS2-VASc score therefore departs from other works based on supervised classification of patient reports.

Negation processing has been extensively studied for English-language medical texts.^[15,4,17] Recent work transferred NegEx to Swedish by transposing its trigger phrases from English to Swedish.^[18,19] Although Swedish and English are both Germanic languages, a simple translation was not enough, because of differences in grammatical features (e.g., gender and number agreement in Swedish), constructs (e.g., *do* auxiliary in English negations), or word order. French is a Romance language, so transfer to French might raise other issues.

Methods

We need to determine the relevant information elements that will be used to compute the eight criteria on which the CHA2DS2-VASc score computation depends. We then need to design an information extraction system which will identify medical concepts in French clinical documents in this purpose. We consider as a relevant information element

a piece of information obtained from a text, more specifically a clinical record, that could be of interest within the framework of thromboembolism risk assessment. The final objective of this framework is to automatically identify patients with risk of thromboembolism attack in case of atrial fibrillation.

With the help of a cardiologist, we defined a list of nine topics the system must focus on to deal with this framework. These are topics around which information elements needed to compute the eight CHA2DS2-VASc score criteria can be found, although there is not a one-to-one correspondence between these topics and the criteria: one topic may contribute to several criteria, and vice-versa. These topics are the following: (i) age of patient, (ii) atrial arrhythmia episode, (iii) blood clot or thrombus formation, (iv) arterial embolism, (v) cardiovascular risk factors (tobacco addiction, diabetes, etc.), (vi) heart disease (aortic valve regurgitation, left ventricular ejection fraction (LVEF), left ventricular end diastolic diameter (LVEDD), mitral failure, etc.), (vii) atrial fibrillation duration and characteristics, (viii) rate-lowering drug/treatment and anticoagulation treatment, and (ix) pacemaker and defibrillator information. These topics are general categories which we have to expand to be able to extract the specific types of medical information that are relevant here. Example instances of information elements to extract are given in Table 1.

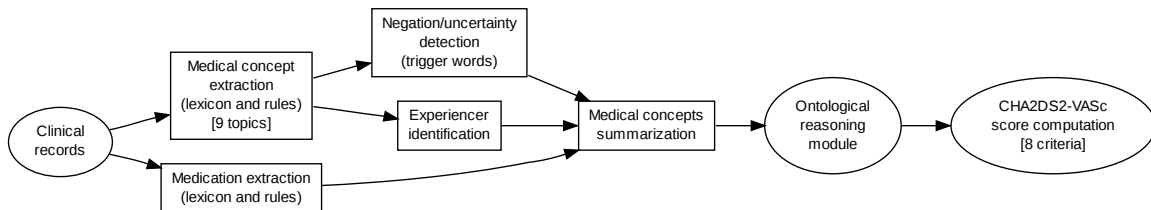
Table 1: Relevant information elements for thromboembolism risk identification in French clinical documents.

French examples	English translation
un <i>tabagisme majeur</i> , de l'ordre de 40 <i>cigarettes/jour</i> poursuivi depuis l'âge de 12 ans...	a <i>major tobacco addiction</i> , about 40 <i>cigarettes/day</i> continued since the age of 12...
un <i>cholestérol total</i> à 2,9 g/l et des <i>triglycérides</i> à 1,82 g/l avec un <i>LDL</i> à 2,12 g/l.	a <i>total cholesterol</i> of 2.9g/l and <i>triglycerides</i> of 1.82g/l with an <i>LDL cholesterol</i> of 2.12g/l.
la <i>fraction d'éjection</i> est retrouvée à 44% avec un <i>diamètre télédiastolique du ventricule gauche</i> à 63 mm.	the <i>left ventricle ejection fraction</i> is 44% with a <i>left ventricular end diastolic diameter</i> of 63mm.
il existe une <i>insuffisance mitrale modérée</i> 1,5/4	there exists a <i>moderate mitral valve regurgitation</i> 1.5/4.

Additionally, some drug prescriptions are also a precious clue of patient condition, such as hypertension or other coronary diseases. Extraction of drug prescriptions thus brings additional help to compute the CHA2DS2-VASc score. For instance, if hypertension is not explicitly mentioned but some antihypertensive medication is found in a patient report, the hypertension point can be added to the score.

We created a system to extract these information elements using several NLP modules (see Figure 1). Each module of this system is based on two main characteristics: first, the use of a domain-restricted lexicon to identify important terms as well as trigger words, and second, extraction rules to refine concept identification.

Figure 1: Global architecture of the system used to compute the CHA2DS2-VASc score. NLP modules are inside rectangle shape boxes.



The system first performs a basic sentence segmentation before applying the lexicon and extraction rules. This helps to process the documents at a linguistically sound level of granularity so that syntactic and semantic processes are applied to a controlled input. This system can be categorized as human-knowledge-based, in contrast to machine-learning-based systems, since it relies on human-defined lexicons and extraction rules.

Medical concept extraction

Lexicon We created a global lexicon composed of 106,639 entries we gathered from three distinct lexicons. Not all entries focus on thromboembolism risk assessment (e.g., we gathered all existing drug names, not only those used in cardiology, assuming that ignoring information is easier than looking for missing information). The three lexicons are the following:

1. a drug name lexicon we gathered from both professional and general public sources (Vidal,^{*} Doctissimo,[†] etc.);
2. a list of medical problems extracted from the Unified Medical Language System (UMLS) Metathesaurus;^[16]
3. and a list of specific cardiological terms provided by a cardiologist.

Each entry in our lexicon contains a medical term, a general category the term belongs to (anatomy, disease, drug, family, laboratory results, procedure), the corresponding concept in a home-made ontology,^[14] and the parent of the concept (as found in the ontology). Table 2 presents example medical concepts from each of these categories. The first step of the system uses this global lexicon to identify these medical terms.

Table 2: Example entries of the global lexicon.

Category	Examples	Number
Anatomy	oreillette gauche (<i>left atrium</i>), valve aortique (<i>aortic valve</i>), ventricule gauche (<i>left ventricle</i>), etc.	5,792
Disease	insuffisance de la valve mitrale (<i>mitral valve regurgitation</i>), lésion aortique (<i>aortic lesion</i>), myxome de l'oreillette gauche (<i>left atrial myxoma</i>), thrombose aortique (<i>aortic thrombosis</i>), etc.	61,103
Drug	atenolol (<i>atenolol</i>), avk (<i>anti-vitamin k</i>), coumadine (<i>coumadin</i>), héparine (<i>heparin</i>), etc.	33,639
Family	beau-père (<i>father in law</i>), jumeaux monozygotes (<i>monozygotic twins</i>), mère (<i>mother</i>), etc.	147
Laboratory results	débit cardiaque (<i>cardiac output</i>), ventilation pulmonaire (<i>pulmonary ventilation</i>), etc.	602
Procedure	pontage aortique (<i>aortic bypass</i>), valvulotomie mitrale (<i>mitral valvulotomy</i>), etc.	5,357

Extraction rules To refine medical concepts located in the documents, we defined specific extraction rules, which we implemented using regular expressions. We based these rules upon empirical observation of the clinical documents in our corpus. As we are focusing on thromboembolism risk, we defined a set of rules to deal with 25 cardiological cases.[‡] These rules allow us to take into account variant expressions (full word, abbreviation, etc.) and/or different cases of precision (with adjectives or different formulations, etc.).

Negation and modality As medical information is written in natural language in clinical documents, a basic identification of clinical concepts is not sufficient. Indeed, a system can detect a medical concept in a negated expression (the patient does not exhibit the mentioned problem); natural language can also express information with varying degrees of uncertainty; finally, within specific sections of the clinical documents, such as history of present illness or family antecedents, medical information can involve someone else than the patient (in the case of a hereditary disease). We created three modules to deal with these linguistic cases, as shown on Table 3.

^{*}Vidal is a medical reference in France for health care professionals, <http://www.vidal.fr/> visited on 03/15/2011.

[†]Doctissimo is a general public health and wellness portal with articles written by clinicians, <http://www.doctissimo.fr/> visited on 03/15/2011

[‡]Aortic leak, aortic stenosis, atrial failure, atrium, hypertension, body mass index, bypass, crackles, left ventricle ejection fraction, fibrillation, flutter, hereditary antecedent, high blood pressure, internationalized normalized ratio, LDL cholesterol, left ventricular end diastolic diameter, mitral valve regurgitation, mitral leak, mitral stenosis, overweight, pacemaker, prothrombin rate, ventricular septum, shortening fraction, tobacco addiction

Table 3: Modules to handle negation and other modalities of the extracted information.

Module	French examples	English translation
Negation	<i>ne retrouve pas</i> le moindre œdème des membres inférieurs	<i>does not find</i> the slightest lower limb edema
Uncertainty	afin de rechercher une <i>éventuelle</i> ischémie myocardique	in order to find a <i>potential</i> myocardial ischemia
Experiencer	un infarctus du myocarde chez son <i>père</i>	a myocardial infarction in her <i>father</i>

In order to deal with negation, we used the NegEx algorithm.^[15] To adapt this algorithm to French, we created a list of 318 negation triggers for French. This trigger list is based on a translation of the existing English triggers and empirical observation of the French cardiology corpus. We also used the 9 major categories of NegEx to categorize negation triggers. Table 4 displays examples of the triggers used in our adapted NegEx system to identify negation in French clinical documents.

Table 4: French negation triggers used with the NegEx algorithm.

Category	Trigger
Pre negation	absence de (<i>lack of</i>), jamais eu (<i>never had</i>), aucun (<i>no</i>), pas de signes de (<i>no sign of</i>), etc.
Pre possible negation	pour écarter (<i>to rule out</i>), etc.
Post negation	est écarté (<i>is ruled out</i>), ont été éliminés (<i>have been eliminated</i>), etc.
Post possible negation	peut être écarté (<i>can be ruled out</i>), sera écarté (<i>will be ruled out</i>), etc.
Conjunction	cependant (<i>nevertheless</i>), sauf (<i>except</i>), etc.
Pseudo negation	pas de changement significatif (<i>no significant change</i>), pas sûr de (<i>not certain whether</i>), ne cause pas (<i>does not cause</i>), etc.

A second module allows us to determine the uncertainty of the expressed information within a sentence. We created a module based upon trigger words of two kinds:

- Pre uncertainty trigger words: *éventuel* (*eventual*), *hypothèse* (*hypothesis*), *possible* (*possible*), *probable* (*probable*), *risque de...* (*risk of...*), *dépistage du...* (*screening of...*), *prévention du...* (*prevention of...*), *recherche du...* (*search for...*), etc.;
- Post uncertainty trigger words: *suspecté* (*suspected*), *comme hypothèse* (*as hypothesis*), etc.

In case of uncertainty, we decided not to extract the medical concept, considering that a human interpretation of such a phrase is needed to decide whether this concept must be considered or not.

Finally, we also designed a module which tries to identify who is the experiencer of a medical problem within a window of 9 words before or after the studied medical problem. This module uses a list of 147 entries from our global lexicon to identify the subject of the disease. It detects whether the mentioned medical problem affects the patient or someone else from their family. If a problem affects someone else than the patient, we do not take this problem into account.

Medical prescription

Our medication extraction module is an adaptation to French of a system we designed for the 2009 i2b2 natural language processing challenge.^[5] This challenge was dedicated to medication prescription extraction from clinical documents in English. It aimed at extracting drug names and all related information (dosage, mode of administration, frequency, duration, and reason for prescription). We took advantage of this challenge to develop a medication extraction system.^[20] Using a rule-based system, we ranked 8th out of 22 participants with a 0.773 F-measure.

Our system relies on the use of lexicon and extraction rules based on trigger words (abbreviations and expressions for all related information classes). We built three lexicons: (i) a drug lexicon to detect drug names based upon drug names from the UMLS Metathesaurus and therapeutic classes, (ii) signs and symptoms lexicons to identify the reason why a given medication was prescribed, based upon the UMLS Metathesaurus using entries with the “Signs and Symptoms” semantic type and the “MetaMap NLP View” flagged terms, and (iii) lists of abbreviations and expressions to extract drug-related information, where each entry has been associated with the type of information it denotes.

We performed an extraction in several steps, as follows: (i) we split the document into sentences, (ii) we then applied the lexicon to identify drug names within each sentence as an exact match, (iii) we split the sentences into parts, where one part begins with a drug name, and (iv) we searched related information inside each part, considering that related information often follows a drug name, but we also extended the search to the sequence closely preceding the drug name.

We then adapted our system to the French language.^[21] We kept the general architecture of the English system (sentence splitting, identification of drug names, and detection of associated information) using a lexicon and rules. We modified the lexicon (gathering drug names, pharmacological substances and abbreviations or spelling variants in French) and the rules (by adapting English rules and adding new rules designed by observation of the development corpus). We also kept the same classes of target information (medication, dosage, mode of administration, frequency, duration, and reason for prescription). We evaluated our French medication extraction system over a test corpus composed of 50 French patient records that we manually annotated as a reference (257 drug names to identify with their related information): it obtained a 0.867 F-measure.

Concept compilation

Having identified medical concepts and prescriptions within clinical documents, the system builds a global XML file that sums up all information extracted from the document; it also completes each kind of information with generic attributes that allow subsequent processing to easily access information items independently of the way information was given in the source clinical document.

For each medical concept, we add specific values denoting the type of concept and the corresponding concept in the Akenaton ontology. Table 5 lists examples of extracted concepts from a clinical document with the additional information we add for each medical concept.

Table 5: Examples of concepts extracted from a clinical document and additional information for this concept.

Concept in the document	Type	Concept in the ontology
Cardensiel	Bradycardisant	Not found
Mode AAI safe R	Pacemaker	AAI Mode
Pace maker double chambre	Pacemaker	Dual chamber pacemaker

For each prescription, we also linked each drug name to the ATC (Anatomical Therapeutic Chemical) classification system, indicating both ATC code and ATC general class. Table 6 provides examples of medication prescriptions we extracted from a clinical document. The resulting XML file summarizes the results of the natural language processing modules of our text-based CHA2DS2-VASc computation pipeline.

Table 6: Examples of prescriptions extracted from a clinical document and additional information for their concepts.

Medication	ATC Code	ATC Class	Dosage	Frequency	Reason for prescription
Triatec (ramipril)	C09AA05	Inhibiteurs de l’enzyme de conversion non associés (<i>ACE inhibitors</i>)	5 mg	/jour (<i>per day</i>)	—
Zanidip (lercanidipine)	C08CA13	Antagonistes calciques non associés (<i>calcium channel blockers</i>)	10 mg	/jour (<i>per day</i>)	Hypertension (<i>high blood pressure</i>)

Computation of CHA2DS2-VASc score

This computation is performed by an ontological reasoning module based on OWL and SWRL.^[13] This score computation module completes our full CHA2DS2-VASc computation pipeline. In order to process the output of the NLP modules, the ontological reasoning module focuses on the concept in the ontology given for each extracted concept (see Table 5), and on the ATC code given for each drug name (see Table 6).

We give in Table 7 an example of medical information extracted by our system, and the computed CHA2DS2-VASc point obtained for each extracted concept.

Clinical record	Antécédents médicaux : HTA. [...] Traitement à l'entrée : PREVISCAN (0-0-1/2). [...] A l'examen clinique de ce jour, il n'y a aucun signe d'insuffisance cardiaque.	
	<i>Medical antecedents: high blood pressure. [...] Treatment on admission: PREVISCAN (0-0-1/2). [...] On today's physical examination, there is no sign of cardiac failure.</i>	
Medical concept	Extracted concept:	hta (<i>high blood pressure</i>)
	Concept in the ontology:	hypertension (<i>high blood pressure</i>)
	CHA2DS2-VASc point:	1
Medical concept	Extracted concept:	insuffisance cardiaque (<i>cardiac failure</i>)
	Concept in the ontology:	heart-failure
	Assertion:	negated
	CHA2DS2-VASc point:	0

Evaluation corpus and setting

Our global corpus is composed of 62 files for patients that attended a cardiology hospital department. Each patient file includes clinical reports (neuro facial radiology and surgical reports), diagnoses (with the corresponding code), and a list of all medical procedures with the corresponding codes. We created a reference corpus composed of 21 patient files. First, the CHA2DS2-VASc score was automatically computed through our complete pipeline for each of the 62 patients. The computed scores ranged from 0 to 7. We then selected at least 2 patients for each value of this score for this reference corpus. This reference corpus was then read by a cardiologist who studied the source patient files and manually recorded all the CHA2DS2-VASc criteria necessary to compute the CHA2DS2-VASc score. These criteria were fed to the CHA2DS2-VASc computation module which produced the CHA2DS2-VASc scores for these patients. These scores constitute the reference we aim to reproduce.

We evaluated the natural language processing pipeline at two levels. First, by comparing its computed criteria with those obtained by human reading of the same clinical texts. Second, by assessing the impact of the observed differences on the resulting CHA2DS2-VASc scores.

Results

For each patient file of the reference corpus, we evaluated the performance of the natural language processing subset of the pipeline by comparing its results with human-based results.

Table 8 displays the differences between the human-based values and the automatically computed values for the criteria: 0 means identity while -1 or 2 is the difference between the two values. It lists, for the three patient files for which a difference was found, its identifier, the number of documents it contains, and the two information elements obtained from the structured part of the record: Age and Sex (although both are generally listed in the texts, their systematic presence in the structured record makes it less useful to use the values found in the texts). It also shows the two CHA2DS2-VASc scores obtained through human (Hum) and fully automatic processing (Auto).

Table 8: CHA2DS2-VASc computed measures for both human-reading (Hum) and automatic method (Auto). CHF stands for “Congestive Heart Failure”, HTA for “Hypertension”, A2 for “Age \geq 75”, DIA for “Diabetes mellitus”, S2 for “Stroke”, PAD for “Peripheral Artery Disease”, A for “65 \leq Age $<$ 75, and Sc for “Sex category”.

Patient				Score		Delta in CHA2DS2-VASc criteria							
Id	# doc	Age	Sex	Hum	Auto	CHF	HTA	A2	DIA	S2	PAD	A	Sc
57	29	79	M	7	6	-1	0	0	0	0	0	0	0
59	16	75	F	6	5	0	0	0	0	0	-1	0	0
72	34	82	F	6	7	-1	0	0	0	2	0	0	0

Discussion

These results show that the CHA2DS2-VASc score computed using the automatically extracted medical concepts is very close to the scores computed through human reading of each document. At the level of criteria, $21 \times 8 = 168$ criteria had to be computed, among which 164 were exact. This corresponds to an accuracy of $164/168 = 0.976$.

Over a total of 21 patients, 18 obtained the same CHA2DS2-VASc score through the fully-automatic and human-based methods, an accuracy of 0.857. Besides, all three different assessments of the score were distant of only one point. Moreover, all involved scores greater than two. All computed CHA2DS2-VASc scores vary from 0 to 7 in our corpus. Traditionally, cardiologists rely on three categories defined over the CHA2DS2-VASc score: a score greater than 2 (where anticoagulation therapy is permanent), a score of one (anticoagulation therapy is recommended), and a null score (anticoagulation therapy is not indicated). The three score errors computed a final score of 6 instead of 7 (patient #57), a score of 5 instead of 6 (patient #59), and a score of 7 instead of 6 (patient #72). Indeed, computational errors are always serious in medicine, but in these cases, the alert raised for these patients is similar; they must take a treatment and the pacemaker alert must be considered as serious. Therefore, system response for all 21 patients would be adequate in this test set: this means that system performance on this test set yields results identical to human reading of the texts.

Closer analysis shows that for patient #59, the automatic approach missed a medical concept which counts for one point (peripheral artery disease): “athérome calcifié non sténosant de la bifurcation” (*non stenosing calcified atheroma in the bifurcation*) that does not exist in our lexicon; this problem can be easily solved by adding this concept to the lexicon. For patient #72, the automatic approach leads to one excess point in the score. Looking in more detail, we find two errors: one error (linked to the criterion “congestive heart failure” that counts for one point in the score) concerns a medical concept that is not present in the lexicon: we have to add: “décompensation cardiaque globale” (*global cardiac decompensation*); the second error concerns a concept that has been extracted as being present in the history of the patient “une AVC qui n’a pas été confirmée” (*a stroke that has not been confirmed*) whereas this concept (*stroke*) is not present; this counts for 2 points in the computed CHA2DS2-VASc score. In consequence, we have to improve the NegEx trigger list for French that we used to deal with this case of negation.

Looking at table columns, we can see that we missed one point twice in the criterion “Congestive Heart Failure” and one point in the criterion “Peripheral Artery Disease” due to medical concepts absent from our lexicon. We also added two points through the criterion “Stroke” because of a problem in our negation processing module.

We performed an evaluation of the negation detection module on the corpus of 21 patients which represents 424 clinical records to process. Over a total of 914 concepts, 59 are negated while 855 are not negated. Our system annotated 79 concepts as negated (among which 53 are correct) and 835 as not negated (among which 809 are correct). We obtained a global F-measure of 0.863.

Missing negations are due to non-existing trigger words in our lexicon (“non accompagnée de” *not accompanied by*) while false negations are due to an excessive factoring of the negation information from a concept to the following one: “sans gradient intraventriculaire, insuffisance mitrale toujours minime” (*without any intraventricular gradient, mitral insufficiency still negligible*); in this case, the negation “sans” (*without*) only focuses on the first concept “gradient intraventriculaire” (*intraventricular gradient*) while our adaptation also tagged the second concept “insuffisance mitrale” (*mitral insufficiency*) as negated without taking into account the comma as a clue separating two phrases in the sentence. Preventing the propagation of negation to several concepts is difficult because of the way some clinical

records are written, especially when several successive concepts are listed without any punctuation mark to split these entries.

Conclusion

Within the framework of the thromboembolism risk, the CHA2DS2-VASc score allows cardiologists to identify easily potential stroke risk for patients with non-valvular fibrillation. We presented in this paper a system based on natural language processing which automatically extracts medical concepts and prescriptions from French clinical records, taking into account the expressed negation for each concept. These are then used as a way to identify relevant criteria that are part of the CHA2DS2-VASc score. When evaluating the NLP modules, the prescription extraction module obtained a global F-measure of 0.867 while the negation module obtained a global F-measure of 0.863.

We performed an overall evaluation on 21 patients files, based on a comparison of the computed CHA2DS2-VASc score, one score being computed based on criteria extracted by a cardiologist reading the documents, the other score being computed using the automatically extracted medical concepts. This evaluation showed similar results, with an accuracy of 0.976 at the level of the 186 individual criteria and of 0.857 at the level of the CHA2DS2-VASc scores. The observed differences are mainly due to two problems: first, a medical concept that is not present in a lexicon, and second, a lack of precision in negation handling. However, further consideration of practical implications of the obtained scores shows that the three differing scores remain in the same categories and that the same patients would raise alerts.

There is still room for improvement, especially for the adaptation of the NegEx algorithm and related linguistic resources to French language. This part of the work is crucial since a medical concept mentioned in the clinical document can be presented in a negative way (the problem is not present or the problem could occur under certain conditions), and could eventually change the meaning of the patient results in a totally different way. We noticed that these assertion problems are a key asset to access the meaning of clinical records; in 2010, one task from the i2b2/VA challenge focused on the assertion annotation in order to detect if a medical concept was present, absent, possible, hypothetical, conditional, or associated with someone else.^[4] While participating in this challenge,^[22] we obtained high results using machine-learning approaches; further work is needed to better adapt the method we used in this challenge from English to French.

Acknowledgement

This work has been funded by the Akenaton project under grant number ANR-07-TecSan-001.

References

1. Naomi Sager, Margaret Lyman, Ng T. Nhn, and Leo J. Tick. Medical language processing: Applications to patient data representation and automatic encoding. *Methods Inf Med*, 34(1-2):140-6, 1995.
2. Carol Friedman, Philip O. Alderson, John H.M. Austin, James J. Cimino, and Stephen B. Johnson. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*, 1(2):161-74, 1994.
3. Stefan M. Meystre, Guergana K. Savova, KC Kipper-Schuler, and JF Hurdle. Extracting information from textual documents in the electronic health record: a review of recent research. In *Yearb Med Inform*, pages 128-44. Shattauer, Stuttgart, 2008.
4. Özlem Uzuner, Brett R. South, S. Shen, and Scott L Duvall. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc*, Jun 16 2011. [Epub ahead of print].
5. Özlem Uzuner, Imre Solti, and Eton Cadag. Extracting medication information from clinical text. *J Am Med Inform Assoc*, 17(5):514-8, 2010.
6. Özlem Uzuner, Ira Goldstein, Yuan Luo, and Isaac Kohane. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc*, 15(1):14-24, 2008.
7. Özlem Uzuner. Recognizing obesity and comorbidities in sparse data. *J Am Med Inform Assoc*, 16(4):561-70, 2009.

8. Serguei V. Pakhomov, James Buntrock, and Christopher G. Chute. Prospective recruitment of patients with congestive heart failure using an ad-hoc binary classifier. *J Biomed Inform*, 38(2):145–53, 2005.
9. Jeff Friedlin and Clement J. McDonald. A Natural Language Processing System to Extract and Code Concepts Relating to Congestive Heart Failure from Chest Radiology Reports. In *AMIA Annu Symp Proc*, pages 269–73, 2006.
10. Anita Burgun, Lynda Temal, Arnaud Rosier, Olivier Dameron, Philippe Mabo, Pierre Zweigenbaum, Régis Beuscart, David Delerue, and Christine Henry. Integrating clinical data with information transmitted by implantable cardiac defibrillators to support medical decision in telecardiology: the application ontology of the Akenaton project. In *AMIA Annu Symp Proc*, page 992, 2010. (Poster).
11. European Heart Rhythm Association, European Association for Cardio-Thoracic Surgery, A John Camm, Paulus Kirchhof, Gregory YH Lip, Ulrich Schotten, Irene Savelieva, Sabine Ernst, Isabelle C Van Gelder, Nawwar Al-Attar, Gerhard Hindricks, Bernard Prendergast, Hein Heidbuchel, Ottavio Alfieri, Annalisa Angelini, Dan Atar, Paolo Colonna, Raffaele De Caterina, Johan De Sutter, Andreas Goette, Bulent Gorenek, Magnus Heldal, Stefan H Hohloser, Philippe Kolh, Jean-Yves Le Heuzey, Piotr Ponikowski, and Frans H Rutten. Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (ESC). *Eur Heart J*, 31(19):2369–429, Oct 2010. PMID: 20802247.
12. Gregory YH Lip and Jonathan L Halperin. Improving stroke risk stratification in atrial fibrillation. *Am J Med*, 123(6):484–8, 2010.
13. Olivier Dameron, Pascal Van Hille, Lynda Temal, Arnaud Rosier, Louise Deléger, Cyril Grouin, Pierre Zweigenbaum, and Anita Burgun. Comparison of OWL and SWRL-based ontology modeling strategies for the determination of pacemaker alerts severity. In *AMIA Annu Symp Proc*, 2011.
14. Lynda Temal, Arnaud Rosier, Olivier Dameron, and Anita Burgun. Modeling cardiac rhythm and heart rate using BFO and DOLCE. In *International Conference on Biomedical Ontology*, 2009.
15. Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce G Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*, 34(5):301–10, 2001.
16. Donald A Lindberg, Betsy L Humphreys, and Alexa T McRay. The Unified Medical Language System. *Methods Inf Med*, 32(4):281–91, 1993.
17. Delphine Bernhard and Anne-Laure Ligozat. Analyse automatique de la modalité et du niveau de certitude : application au domaine médical. In *Proceedings of TALN 2011*, Montpellier, 2011.
18. Maria Skeppstedt. Negation Detection in Swedish Clinical Text. In *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, pages 15–21, Los Angeles, California, USA, June 2010. Association for Computational Linguistics.
19. Hercules Dalianis and Maria Skeppstedt. Creating and evaluating a consensus for negated and speculative words in a Swedish clinical corpus. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 5–13, Uppsala, Sweden, July 2010. University of Antwerp.
20. Louise Deléger, Cyril Grouin, and Pierre Zweigenbaum. Extracting medical information from narrative patient records: the case of medication-related information. *J Am Med Inform Assoc*, 17(5):555–8, 2010.
21. Louise Deléger, Cyril Grouin, and Pierre Zweigenbaum. Extracting Medication Information from French Clinical Texts. In *Stud Health Technol Inform*, volume 160(Pt 2), pages 949–53, 2010.
22. Anne-Lyse Minard, Anne-Laure Ligozat, Asma Ben Abacha, Delphine Bernhard, Bruno Cartoni, Louise Delger, Brigitte Grau, Sophie Rosset, Pierre Zweigenbaum, and Cyril Grouin. Hybrid methods for improving information access in clinical documents: Concept, assertion, and relation identification. *J Am Med Inform Assoc*, 18(5):588–593, 2011.