Text Mining of Medical Documents in Spanish: Semantic Annotation and Detection of Recommendations

Carlos Tellería[1][a], Sergio Ilarri[2][b] and Carlos Sánchez[3]
[1]*Instituto Aragonés de Ciencias de la Salud, Zaragoza, Spain*
[2]*I3A, University of Zaragoza, Zaragoza, Spain*
[3]*University of Zaragoza, Zaragoza, Spain*

Keywords: Medical Documents, Information Extraction, Text Mining, Classification, Semantic Annotation, Detection of Recommendations in Texts, Spanish Texts.

Abstract: In medical practice, identifying relevant facts and therapeutic recommendations from health-related documents is a key issue to ensure an efficient and effective service to patients. However, the automatic analysis of text documents to extract relevant data is a challenging task. This is the case particularly when we deal with documents written in languages other than English, for which the availability of lexical resources and tools is much more limited and less experiences have been reported. In this paper, we present our experience dealing with texts written in Spanish in a medical context. By applying text mining techniques and exploiting semantic resources, we present an approach to automatically label documents using appropriate medical terms. Besides, we also describe a technique that attempts to detect practice recommendations for doctors automatically in clinical guides. An experimental evaluation shows the benefits of applying text mining techniques as a support system for doctors as well as its feasibility. The scarcity of experimental evaluations with medical documents in Spanish motivated our work.

1 INTRODUCTION

The amount of text documents containing relevant medical information is continuously growing. Whereas this is a positive trend that proves a significant dissemination of research results in the health area, it is also very challenging for doctors to identify the most relevant data and keep up with the latest research and medical recommendations. Even within a single document dealing with a specific health topic, it can be difficult to quickly find the key points and distinguish recent research results from well-established practice recommendations and guidelines, especially if this has to be done in a short time while examining a patient during a consultation. In this context, the development of software support tools that can help health professionals to filter and identify relevant information quickly would be very profitable. Thus, a tool assisting in the task of identifying relevant facts and therapeutic recommendations from health-related documents could improve the efficiency and effectiveness of health providers.

For this purpose, the application of text mining techniques could be very helpful. However, analyzing text written in natural language is challenging. Moreover, the difficulties increase significantly when we have to manage documents written in non-English languages, as appropriate lexical resources and tools are scarce in that case and the number of experiences reported is significantly much smaller.

As the benefits for both citizens and health professionals could be huge and the amount of research performed in this context is still quite limited, we are researching techniques to deal with unstructured medical documents written in Spanish. More specifically, in this paper, we present a practical experience developed to tackle two problems: the automatic labelling of medical documents using suitable medical concepts and the identification of recommendations and guidelines (practice recommendations for doctors) in health-related texts. Furthermore, an experimental evaluation using anonymized clinical histories (for the labelling task) and clinical guides (for the detection of recommendations) shows the benefits and the feasibility of applying text mining techniques as a

[a] https://orcid.org/0000-0002-6394-3212
[b] https://orcid.org/0000-0002-7073-219X

Tellería, C., Ilarri, S. and Sánchez, C.
Text Mining of Medical Documents in Spanish: Semantic Annotation and Detection of Recommendations.
DOI: 10.5220/0010059101970208
In *Proceedings of the 16th International Conference on Web Information Systems and Technologies (WEBIST 2020)*, pages 197-208
ISBN: 978-989-758-478-7

support system for doctors. The structure of the rest of this paper is as follows. In Section 2, we describe the state of the art. In Section 3, we present our approach for the automatic labelling of medical documents. In Section 4, we describe the technique used to detect practice recommendations in texts. Finally, in Section 5, we show our conclusions and outline some prospective lines of future work.

2 RELATED WORK

Exploiting data available in a non-structured format, such as text documents, is a difficult task for which a myriad of text mining techniques have been developed (Aggarwal and Zhai, 2012). Typical operations that can be performed with texts include: information retrieval (Manning et al., 2008) (obtention of relevant documents satisfying a given query, usually a keyword-based query), text classification (Sebastiani, 2002) (automatic allocation of documents to an appropriate category from a set of possible predefined classes), information extraction (Jiang, 2012) (retrieval of specific data from the text), textual annotation (Liao and Zhao, 2019) (automatic assignment of suitable labels to texts) and named entity recognition (Marrero et al., 2013) (detection of named entities such as references to people or company names, geographic places, etc.), and document summarization (Gholamrezazadeh et al., 2009).

Concerning specifically health documents, there is also a growing interest in applying text mining techniques to automatically process text data in order to maximize the probability of finding the relevant data and minimize the cost, which would lead to an overall improvement of health services. A typical example is the use of text mining in biomedicine (Simpson and Demner-Fushman, 2012; Spasic et al., 2005). Most works that apply text mining on medical documents focus on a specific area, such as oncology (Yim et al., 2016), radiology (Pons et al., 2016), geriatrics (Chen et al., 2019), or suicide prevention (Coppersmith et al., 2018). According to (Marrero et al., 2010), two relevant peculiarities that imply additional difficulties for the biomedical domain are the difficulties regarding terminological consensus and the lack of terminological patterns in practice.

Most existing text mining approaches over medical documents focus on texts written in English, where a number of tools and linguistic resources are available. According to (Névéol et al., 2018), where the challenges and opportunities of clinical natural language processing in languages other than English are studied, "Chinese and Spanish have recently attracted sustained efforts", but studies for Spanish are for the moment quite behind other non-English languages such as French, German and even Chinese. As examples of some efforts performed for the Spanish language, we can cite (Castaño et al., 2016; Costumero et al., 2014; Marimon et al., 2019). Thus, in (Castaño et al., 2016) an unsupervised machine learning approach to discover the equivalence between terms (considering synonyms, abbreviations, acronyms, and frequent typos) is presented; although this work focuses on documents in Spanish, no specific resource for Spanish was used. The work presented in (Costumero et al., 2014) tackles the problem of detecting negation regarding clinical conditions in Spanish medical documents. Techniques to process Spanish medical texts to remove sensitive patient information have also been proposed (Marimon et al., 2019). Finally, it is also interesting to mention the possibility of applying automatic machine translation to medical documents, which would potentially enable the application of resources and tools available for the chosen target language (Wu et al., 2011). As opposed to these works, in this paper we present our experience concerning the use of text mining for the semantic annotation of medical documents and for the detection of recommendations in clinical guides. This contributes to the state of the art by reporting how different techniques can be exploited to provide suitable results, thus increasing the scarce amount of experiences with medical texts in Spanish.

3 AUTOMATIC LABELLING OF HEALTH DOCUMENTS IN SPANISH

We have decided to use two lexical resources as a basis for automatic labelling of medical documents in Spanish: SNOMED CT (Spanish edition) and DeCS.

3.1 SNOMED CT

SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) (Cornet and de Keizer, 2008; SNOMED International, 2020) is a clinical terminology that has been translated to several languages, including Spanish. It is delivered through two CSV files, one containing the terms (more than 1 million terms) and their classes, and another one with the relations (more than 5 million relations) between the terms (e.g., subclass relationships and synonymy). Based on the Spanish edition of SNOMED CT, we have built a dictionary of terms containing 951213 terms (including synonyms) divided into 20 classes. Although the initial number of classes in SNOMED CT was 98, we have combined some classes into a single one when, based on the available information, they were considered too similar (e.g., "medicamento clínico", which is"clinical drug" in English, and "fármaco de uso clínico", which is "drug of clinical use" in English). Table 1 shows some examples of terms correctly detected in clinical histories of our dataset thanks to the use of SNOMED CT.

Table 1: SNOMED-CT: examples of terms detected in our dataset of clinical histories.

Term	**Associated class (or associated classes)**
cambio degenerativo	anomalía morfológica
canal	estructura corporal
cuerpo vertebral	estructura corporal
deshidratación	trastorno

3.2 DeCS

DeCS (Descriptores en Ciencias de la Salud / Health Sciences Descriptors) (BIREME, 2020a) is a multilingual dictionary aimed at facilitating the indexation and retrieval of scientific medical documents (stored in specialized repositories such as LILACS and MEDLINE). It was developed based on *MeSH (Medical Subject Headings)* (Trieschnigg et al., 2009), provided by the U.S. National Library of Medicine. Thanks to a hierarchy relating the different terms, it is possible to make searches more specific or more general, by moving down or up through the hierarchy, respectively. It contains 33966 descriptors and qualifiers: 29431 of them come from MeSH and 4535 are exclusive of DeCS. DeCS is delivered through several files. The one we have used is an XML file containing all the terms in Spanish; through the DeCS Web Services (BIREME, 2020b), we have retrieved information related to the field *treeId*, such as the upper class of the hierarchy for a given term, by using a URL with the structure http://decs.bvsalud.org/cgi-bin/mx/cgi=@vmx/decs/?tree_id=⟨id⟩. Overall, we have obtained 91823 medical terms (less than a 10% of the number of terms obtained with SNOMED CT). Table 2 shows some examples of terms correctly detected in clinical histories of our dataset thanks to the use of DeCS.

Table 2: DeCS: examples of terms detected in our dataset of clinical histories.

Term	**Associated class (or associated classes)**
alergia	ENFERMEDADES - SALUD PÚBLICA
anamnesis	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
bocio	ENFERMEDADES - SALUD PÚBLICA
CEC	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS

3.3 Annotation Methods Considered

Our main goal is to evaluate whether using the given semantic resources (SNOMED CT and DeCS) can help to achieve satisfactory annotations. Therefore, with the two resources described above, we have compared several annotation methods. Given the difficulty to find annotated medical datasets in Spanish, it is challenging to have a large gold-standard corpus available for machine learning modeling and evaluation. Building such a corpus through manual annotation would be time consuming and would require the participation of health care professionals. Therefore, rather than applying machine learning techniques to try to learn appropriate annotation models, we rely on other types of methods (not based on machine learning) and we will evaluate their performance on a set of documents manually annotated in order to assess how well the automatic methods behave.

1) String Matching. We have first considered a simple model that tries to find an exact matching between the words in the text and the terms in the corresponding semantic resource (SNOMED CT and DeCS). A preprocessing stage removes first non-valid characters that may be present in the text documents and/or terms, and everything is initially transformed to lowercase for the purpose of comparison. Then, each term in the resource is compared with each ngram in the text, to try to find suitable matchings.

We tested the *re* Python library (Python Software Foundation, 2020), but in our experiments the execution times were high (between 38.96 and 206.53 seconds, on an HP Pavilion with Intel Core i7-8700 and 16 GB RAM, depending on the size of the document and whether SNOMED CT or DeCS was used). Finally, we used *FlashText* (Singh, 2017a; Singh, 2017b), which performed the task much more efficiently (in about $3,29\%$ of the time, on average); an object *KeywordProcessor* is built, containing all the dictionary entries, to enable a quick detection of text matches through the use of a *trie* data structure (Fredkin, 1960; Sahni and Mehta, 2018).

2) Approximate Detector with a Spanish Dictionary (String Matching with Lemmatization and Spell Correction using a Spanish Dictionary). An obvious shortcoming of the string matching approach is that even a slight change in the form of a word or group of words will lead to a mismatch. To alleviate this problem, with this method we first perform more preprocessing steps. More specifically, the preprocessing steps are the following: 1) transformation of all the text to lowercase (except in the case of words with all the letters in uppercase, which are assumed to be acronyms), 2) removal of stopwords (prepositions, determiners, etc.), 3) lemmatization of the words in the text and the terms (i.e., obtention of the lemma of each word; e.g., the lema of “rojo”, “roja”, “rojos” and “rojas” is “rojo”, which represents the red color in Spanish). FreeLing (Padró, 2008; Padró and Stanilovsky, 2012), which offers an API that can be called from Python, has been used to perform these tasks. With these tools, a tokenization of the input is applied, stopwords are removed, and the lemmas of the remaining words are obtained.

Besides, as some words in the input text data included typos (this is to be expected in documents written by doctors, as they usually have to write a considerable amount of text in a short time), we also applied a spell checker to correct the potential typos before trying to find a suitable match, based on the Levenshtein distance, that tries to obtain the most similar correct word (e.g., “hormigueo”, which is “tingling” in English, instead of the misspelled word “hormiguoe”); for this purpose, we used *symspellpy* (mammothb, 2019), a port of *SymSpell* (Garbe, 2019) for Python, along with a Spanish dictionary, obtained from (Dave, 2019), containing 1211000 entries. The terms detected after applying a spell checker are subject to some uncertainty, as the application of a spell checker automatically could actually lead to a term different than the one intended in the original text; for example, the word “reinitis” may appear in a text instead of “retinitis” (in English, also “retinitis”) but it could be corrected as “rinitis” (in English, “rhinitis”), which is a different disease.

We also considered other tools, such as the *NLTK (Natural Language Toolkit)* library (Loper and Bird, 2002; NLTK Project, 2020a) with its package *SnowballStemmer* (NLTK Project, 2020b), but it does not offer a lemmatization functionality; instead, it only allows to retrieve the lexeme of words, which is not as appropriate (e.g., the lexeme of both “hombre”/“man” and “hombro”/“shoulder” is “hombr”, even though “hombre” and “hombro” are two Spanish words with very different meanings). We also performed some tests with *spacy* (Explosion AI, 2020); although this tool incorporates a lemmatizer, we have noticed that some nouns are lemmatized obtaining an infinitive verb form, even though the morpholinguistic analysis correctly identifies the original word as a noun.

3) Approximate Detector using Lexical Medical Resources (String Matching with Lemmatization and Spell Correction using SNOMED-CT or DeCS). It is equivalent to the previous method but using either SNOMED-CT or DeCS as a dictionary for spell correction. The *python-Levenshtein* library (Haapala, 2019) has been used to compute the Levenshtein distance; a minimum threshold of 0.9 is applied to consider the equivalence between two words. The main disadvantage of this approach is the execution time needed (200-500 seconds, with the aforementioned HP Pavilion, depending on the length of the text): the complexity is $O(n*m)$, where n is the number of terms in the dictionary and m is the number of words in the text.

3.4 Experimental Comparison of the Annotation Methods

To compare the performance of the methods, we have performed tests with 30 real text documents randomly extracted from the input dataset, corresponding to anonymized clinical histories provided by the *Instituto Aragonés de Ciencias de la Salud* (*IACS*), which is the entity that promotes knowledge in Biomedicine and Health Sciences in the region of Aragón (Spain). In total, there are 1212946 documents, from which we finally considered a subset with the 1859 documents that contained more than 1500 characters. Even though the size of the dataset is not very large, we have to take into account that the text documents are rich in medical terms (the average number of medical terms is 36.7, with a standard deviation of 16.15). Moreover, we did not observe significant differences between the performance observed for individual documents (e.g., the average standard deviation of the precision and recall for individual documents is around 0.1). So, the results can be considered representative for this experimental evaluation. Enlarging the dataset is possible, but time consuming and subject to two limitations: the clinical histories need to be carefully anonymized, to guarantee the privacy of the patients, and the documents used for testing have to be manually annotated.

The results obtained are shown in Table 3. The best of the three approaches, in terms of F-measure, is the second method. Besides, we can see how the use of SNOMED CT can lead to better results. It should be noted that the use of the approximate detector using lexical medical resources leads to higher recall values but also to a decreased precision, particularly when using SNOMED CT (that has a higher number of terms). This is because the probability of incorrectly detecting a similar word increases using

Table 3: Automatic annotation of clinical histories: experimental results.

	String matching		Approximate detector using a Spanish dictionary		Approximate detector using lexical medical resources	
	SNOMED CT	DeCS	SNOMED CT	DeCS	SNOMED CT	DeCS
Precision	**0.82**	0.72	0.77	0.65	0.56	0.62
Recall	0.63	0.53	0.74	0.66	**0.81**	0.69
F-measure	0.71	0.61	**0.75**	0.66	0.66	0.65

the approximate matching; this is also characteristic of the traditional tradeoff between precision and recall (Buckland and Gey, 1994). Given the results obtained, the second method could be considered (and extended, if needed) as a basis to develop a system that can facilitate the task of annotation of texts.

Indeed, most undetected terms are due to the use of acronyms and abbreviations used by doctors to refer to diagnostics, therapies, corporal structures, and diseases (e.g., HVI, VD, HTP, etc.). These acronyms appear frequently in the clinical histories but their appearance in the lexical medical resources is scarce, which leads to a decreased recall (i.e., false negatives). Some acronyms are not standardized and they may even have different meanings (e.g., “TEP” can mean “Tromboembolismo Pulmonar” / “Pulmonary Embolism” or “Triángulo de Evaluación Pediátrica” / “Triangle of Pediatric Evaluation”). Therefore, the recall can be improved by defining a suitable dictionary of acronyms and incorporating context-dependent detection methods to disambiguate the correct meaning of certain acronyms. Another important source of false negatives is the presence of commercial names of drugs, which appear in the clinical histories but not in the lexical resources used (SNOMED CT and DeCS), where the active pharmaceutical ingredients may instead be present; the complementary use of data sources like DrugBank (https://www.drugbank.ca/) or DrugCentral (http://drugcentral.org/) could be considered to tackle this problem.

Concerning the precision, SNOMED CT and DeCS contain some detected words that have not been manually annotated as medical terms in the clinical histories used for experimentation. Several false positives correspond to terms whose identification would change the real meaning of the word (e.g., “pico” in a text with the meaning of “peak value” rather than a part of anatomy, which is“beak” in English, “Urgencias” representing the area of a hospital that receives patients that may have an emergency issue, which in English is “Emergency Department”, rather than an “urgency” in the medical sense, “base” representing the lower part of something, which is “basis” in English, rather than a chemical substance, which is “base” in English, etc.). In some cases, false negatives arise because the terms were not considered representative enough from a medical point of view, but without implying a significant mistake.

4 AUTOMATIC DETECTION OF RECOMMENDATIONS FROM CLINICAL GUIDES IN SPANISH

We have also developed a classifier whose goal is detecting, given a medical text, if a certain text fragment is providing a recommendation (a suggestion based on medical evidence) or just other information. For this purpose, we focus on Spanish clinical guides (“Guías de Práctica Clínica del Sistema Nacional de Salud”/“Clinical Practice Guidelines of the National Health System”), which collect recommendations and scientific evidences for clinical treatments in different circumstances, assessing the risks and benefits of the different approaches. These guides are periodically updated to reflect new knowledge on the topic covered and they are available in PDF format from a public website (IACS, 2018). Specifically, we have considered 65 clinical guides for our experiments. First, we used a developed tool that transforms the PDF files of the clinical guides into text files and formats them appropriately, taking the structure of the clinical guides into account. With this tool, we obtained 58864 sentences from the clinical guides.

4.1 Methods Considered for the Detection of Recommendations

Proposing new classification methods is not our goal at this point. Rather, we would like to assess the feasibility of applying known techniques for recommendation classification in this context. Therefore, for experimental evaluation, a baseline and four supervised machine learning classifiers (Caruana and Niculescu-Mizil, 2006) frequently used in the context of natural language processing (NLP), applied over a vector representation of the texts using the metric *Term Frequency – Inverse Document Frequency (TF-IDF)* (Lan et al., 2007), have been implemented and evaluated:

1) Verb Categorization (used as a baseline), which is based on a compiled list of verbs that are frequently used in medical recommendations in natural language. We have selected 45 different verbs (such as "justificar" / "justify", "mejorar" / "improve", "solucionar" / "solve", "aliviar" / "alleviate", "curar" / "heal", "ayudar" / "help", etc.). A text is estimated to be a recommendation if it contains one of the verbs in the list. A lemmatization process is applied to avoid mismatches due to the presence of verbs in conjugated forms.
2) A Support Vector Machine (*SVM*) (Hearst et al., 1998), which tries to determine the best hyperplane that separates the training data in the target classes. We performed tests with different values for the soft margin parameter C (0.1, 0.5, and 1.0), as well as tests with both a linear kernel and a polynomial kernel.
3) Multinomial Naive Bayes (Kibriya et al., 2004), which applies the Bayes theorem to estimate the class of a document based on the words it contains and it is based on the assumption that all the predictors (words) are independent.
4) Random Forest (Breiman, 2001), where several decision trees are built (based on different training sets) and their predictions are combined. The goal is to reduce the variability of the model and increase its precision, at the expense of higher latency and memory consumption as well as a decreased interpretability (compared with single decision trees). We have performed tests with both 1000 and 2000 estimators (decision trees).
5) K-Nearest Neighbors (Chakrabarti et al., 2008), where the class of an instance is estimated based on the predictions of the k nearest neighbors, weighting the predictions depending on the distance to the given instance. We have performed tests with two different distance metrics (the Euclidean distance and the Manhattan distance) and different values of the number of neighbors k ($k = 3$ and $k = 5$).

The verb categorization approach has been implemented as a Python script, based on the use of a dictionary of *recommendation verbs*. The other methods are implemented using the Python library *scikit-learn* (Pedregosa et al., 2011).

4.2 Experimental Comparison of the Recommendation Detection Methods

To evaluate a classification approach, we need a labelled data set, so a process of manual detection of recommendations by humans was followed: five persons (two family doctors and three computer scientists) analyzed each a subset of the documents to find recommendations. It is important to stress that the identification of a sentence as a recommendation or not may depend on the subjectivity of the person; for example, given the sentence "La presentación de TAG más complejos y graves en el inicio, el fracaso en completar el tratamiento y la cantidad de tratamientos intermedios durante el período de seguimiento se asocian con peores resultados de la TCC a largo plazo" ("The presentation of more complex and serious TAG at the beginning, failure when completing the treatment and the amount of intermediate treatments during the follow-up period are associated with worse results of the TCC in the long term"), present in one of the clinical guides, was classified as a recommendation by some persons while others considered that exposing these results did not implicitly convey any recommendation. As a similar example, we also observed disagreement in the interpretation of the text "Un modelo integrado en el que los médicos de familia son apoyados por especialistas, que durante 8 semanas (4-8 sesiones) ayudan a los pacientes a desarrollar habilidades cognitivo-conductuales a través de relajación, reconocimiento de pensamientos ansiogénicos y de falta de autoconfianza, búsqueda de alternativas útiles y entrenamiento en acciones para resolución de problemas, técnicas para mejorar el sueño y trabajo en casa" ("An integrated model where family doctors are supported by specialists, who during 8 weeks (4-8 sessions) help the patients to develop cognitive and behavioral abilities through relaxation, acknowledgement of anxiogenic and lack of self-confidence thoughts, search of useful alternatives and training in actions for problem solving, techniques to improve sleep and work at home").

In order to assess an *agreement score for the human detectors*, 100 texts were randomly selected and they were classified by the 5 persons mentioned above (i.e., they identified recommendations and no recommendations), calculating the score as the percentage of agreements among all the annotators over the total number of texts. In this way, we obtained an agreement score of 63%, which may not seem very high but it is due to the fact that, as explained in the previous paragraph, the interpretation of a sentence as a recommendation or not may depend on the subjectivity of the person reading the sentence. The agreement score between the two doctors is 79% and the agreement score considering only the annotations of the three Computer Scientists is 70%. Considering the annotations of the two doctors, the Cohen's kappa is 0.581, which indicates a moderate agreement.

To compare the different techniques, each person labelled a subset of the documents used for testing and we applied a k-fold cross validation with $k = 5$.

Table 4 summarizes the experimental results in terms of precision, recall and F-measure. The method that provides the best results is the one using a random forest, with no clear impact when passing from 1000 to 2000 trees, so the one with 1000 trees is considered the best approach, among the ones compared, due to its higher simplicity. The second best method is the one using SVM with a linear kernel and $C = 1$. Next in the rank is the kNN approach with $k = 3$ or $k = 5$ using the Euclidean distance as the distance metric. Multinomial Naive Bayes achieves an intermediate performance, worse than the kNN approach with the Euclidean distance but better than some variants of SVM (linear SVM with $C = 0.1$ and the polynomial SVM approach tested) and the kNN approach with the Manhattan distance. The verb categorization approach (used as a baseline) is the one obtaining the worse results along with the polynomial SVM approach as well as the kNN approach with k=3 and the Manhattan distance.

The overall performance of most methods is quite acceptable, especially if we take into account that the score agreement between human annotators is 63%. Due to the existing subjectivity, directly comparing the classification performed by the system with the one proposed by a user is not completely fair. Indeed, the percentage of failures (in terms of false negatives and false positives) of the methods usually fall below the disagreement score among humans (37%), and therefore they could be explained by the subjectivity when interpreting sentences as recommendations. Besides, an F-measure of 0.82, achieved by the random forest methods, is a quite good result for practical applications, as it means that in general doctors can reliably use this method as a support tool to find recommendations quickly.

5 CONCLUSIONS

The automatic processing of health documents can bring significant benefits to existing health systems, for example by helping doctors to find relevant practice recommendations or key terms. In this paper, we have tackled the problem of applying text mining to health-related documents written in Spanish, which is a big challenge, as most resources, tools, and experiences have been developed for English documents. Specifically, we have tackled the problem of automatic annotation of clinical histories with medical terms, as well as the problem of detecting recommendations in clinical guidelines. Based on the experimental evaluation performed, the methods evaluated can be used as a basis for further research, as we could expect further improvements by sophisticating the techniques applied or extending and fine-tuning them for the specific use cases considered. Our work, based on a real-world case study, contributes to increasing the scarce literature providing experimental evaluations with medical documents in Spanish.

Snapshots of a preliminary prototype of a decision support system application that we are developing can be seen in Figures 1 and 2. The text to be analyzed can be entered by the user directly or obtained by using an implemented tool that extracts the text from PDF files. On the top part of Figure 1 we show the original text, with the terms detected shown between "__" and in bold, and ended with an annotation in brackets to indicate the class associated to the term detected. For example, "antihistamínico" ("antihistamine" in English) has been detected as a term belonging to class "sustancia" ("substance") and "urticaria" has been also detected as a term belonging to classes "trastorno" (that could be translated as "outbreak" in this context) and "anomalía morfológica" ("morphological anomaly"). The middle part of Figure 1 indicates how the text shown is categorized by the classifier (in this case, as a recommendation, suggestion or evidence). Finally, in the bottom part of Figure 1 the terms detected and their classes are summarized, to provide a quick overview. Figure 2 shows another example of output for a different input text that has not been detected as a recommendation, which is correct. For clarity and demonstration purposes, we show here a short piece of text, but we have tested the annotator with larger texts corresponding to clinical histories of patients (e.g., see Appendix 5).

As future work, we plan to consider a number of improvements to the methods proposed in this paper and to extend our current experimental evaluation, which has already shown promising results that support the feasibility and interest of the proposals presented and contributed to the scarce amount of experiences with medical texts in Spanish. One of the directions we want to pursue is to analyze ways to perform a context-dependent analysis of the text (e.g., by identifying general topics at a paragraph or section level, that can be used to later evaluate the probability that a given word refers to a certain medical term, especially in the case of misspelled words). We would also like to analyze how additional strategies to deal with acronyms can improve the results. In the case studies presented in this paper, we have focused on clinical histories and medical guides, but the evaluated methods may behave differently (and require significant adaptations) when applied to health-related texts with different structure and typology (like scientific articles), where the application of text mining

Table 4: Detection of recommendations in clinical guides: experimental results.

	Precision	**Recall**	**F-measure**
Verb categorization (baseline)	0.64	0.49	0.56
SVM linear, C=1	0.84	0.79	0.81
SVM linear, C=0.5	0.85	0.75	0.80
SVM linear, C=0.1	0.75	0.46	0.57
SVM polynomial (degree=2), C=1	0.46	0.65	0.54
Multinomial Naive Bayes	0.81	0.73	0.77
Random Forest (1000 estimators)	0.86	0.78	**0.82**
Random Forest (2000 estimators)	0.86	0.78	**0.82**
kNN (k=3, Euclidean distance)	0.77	**0.84**	0.80
kNN (k=5, Euclidean distance)	0.78	0.83	0.80
kNN (k=3, Manhattan distance)	**0.94**	0.39	0.55
kNN (k=5, Manhattan distance)	0.91	0.43	0.58

Figure 1: Prototype of a decision support system: output sample (recommendation).

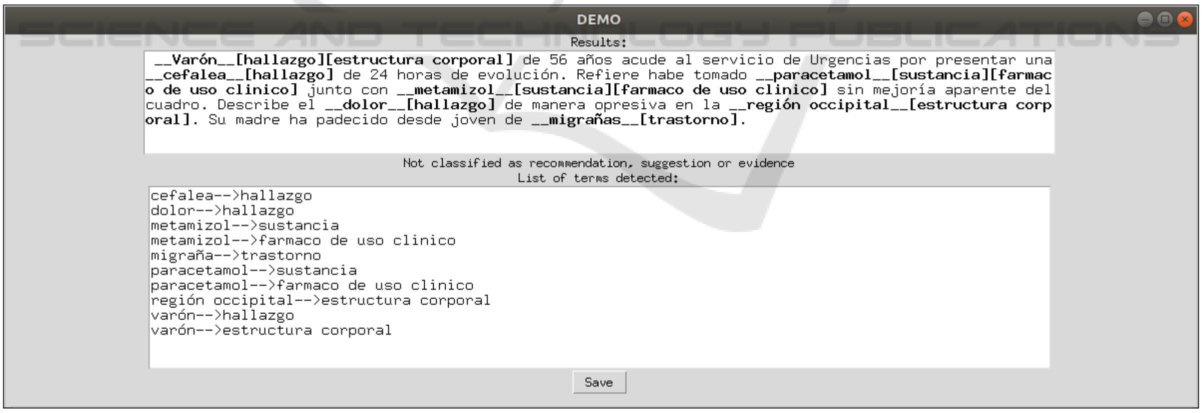

Figure 2: Prototype of a decision support system: output sample (no recommendation).

could provide benefits to other types of end users (like researchers). Finally, performing a large-scale experimental evaluation with these and other proposed methods (e.g., using deep learning, if a large set of data could be compiled) would help to better validate the significance of the results and refine the proposed techniques.

ACKNOWLEDGEMENTS

This work has been supported by the project TIN2016-78011-C4-3-R (AEI/FEDER, UE) and the Government of Aragon (COSMOS group, reference T64_20R). We thank the Health Sciences Institute in Aragón for providing us with real anonymized datasets.

REFERENCES

Aggarwal, C. C. and Zhai, C., editors (2012). *Mining Text Data*. Springer.

BIREME (2020a). DeCS. http://decs.bvsalud.org/. Last access: August 24, 2020. BIREME (Latin American and Caribbean Center on Health Sciences Information).

BIREME (2020b). DeCS Web Services. http://wiki.reddes.bvsalud.org/index.php/Servicios_DeCS. Last access: August 24, 2020. BIREME (Latin American and Caribbean Center on Health Sciences Information).

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Buckland, M. and Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, 45(1):12–19.

Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *23rd International Conference on Machine Learning (ICML 2006)*, pages 161–168. ACM.

Castaño, J., Gambarte, M. L., Park, H. J., Avila Williams, M. d. P., Pérez, D., Campos, F., Luna, D., Benítez, S., Berinsky, H., and Zanetti, S. (2016). A machine learning approach to clinical terms normalization. In *15th Workshop on Biomedical Natural Language Processing*, pages 1–11. Association for Computational Linguistics.

Chakrabarti, S., Cox, E., Frank, E., Gting, R. H., Han, J., Jiang, X., Kamber, M., Lightstone, S. S., Nadeau, T. P., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. J., and Witten, I. H. (2008). *Data Mining: Know It All*. Morgan Kaufmann Publishers Inc.

Chen, T., Dredze, M., Weiner, J. P., Hernandez, L., Kimura, J., and Kharrazi, H. (2019). Extraction of geriatric syndromes from electronic health record clinical notes: Assessment of statistical natural language processing methods. *JMIR Medical Informatics*, 7(1):e13039:1–e13039:12.

Coppersmith, G., Leary, R., Crutchley, P., and Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights*, 10:1–11.

Cornet, R. and de Keizer, N. (2008). Forty years of SNOMED: a literature review. *BMC Medical Informatics and Decision Making*, 8(S1).

Costumero, R., Lopez, F., Gonzalo-Martín, C., Millan, M., and Menasalvas, E. (2014). An approach to detect negation on medical documents in Spanish. In *Brain Informatics and Health*, pages 366–375. Springer.

Dave, H. (2019). FrequencyWords – Repository for Frequency Word List Generator and processed files. https://github.com/hermitdave/FrequencyWords. Last access: August 24, 2020.

Explosion AI (2016–2020). spaCy. https://spacy.io. Last access: August 24, 2020.

Fredkin, E. (1960). Trie memory. *Communications of the ACM*, 3(9):490–499.

Garbe, W. (2019). SymSpell. https://github.com/wolfgarbe/SymSpell. Last access: August 24, 2020.

Gholamrezazadeh, S., Salehi, M. A., and Gholamzadeh, B. (2009). A comprehensive survey on text summarization systems. In *Second International Conference on Computer Science and its Applications (CSA 2009)*, pages 1–6.

Haapala, A. (2019). Python-Levenshtein. https://github.com/ztane/python-Levenshtein. Last access: August 24, 2020.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support Vector Machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28.

IACS (2018). Guías de Práctica Clínica del Sistema Nacional de Salud / Clinical Practice Guidelines of the National Health System. https://portal.guiasalud.es. Last access: August 24, 2020. Instituto Aragonés de Ciencias de la Salud (IACS).

Jiang, J. (2012). Information extraction from text. In *Mining Text Data*, pages 11–41. Springer.

Kibriya, A. M., Frank, E., Pfahringer, B., and Holmes, G. (2004). Multinomial Naive Bayes for text categorization revisited. In *Australasian Joint Conference on Artificial Intelligence (AI 2004)*, volume 3339 of *Lecture Notes in Computer Science*, pages 488–499. Springer.

Lan, M., Tan, C. L., Su, J., and Low, H. B. (2007). Text representations for text categorization: A case study in biomedical domain. In *International Joint Conference on Neural Networks (IJCNN 2007))*, pages 2557–2562. IEEE.

Liao, X. and Zhao, Z. (2019). Unsupervised approaches for textual semantic annotation, a survey. *ACM Computing Surveys*, 52(4):66:1–66:45.

Loper, E. and Bird, S. (2002). NLTK: The Natural Language Toolkit. arXiv, cs/0205028.

mammothb (2019). symspellpy – Python port of SymSpell. https://github.com/mammothb/symspellpy. Last access: August 24, 2020.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Marimon, M., Gonzalez-Agirre, A., Intxaurrondo, A., Rodríguez, H., Martin, J. L., Villegas, M., and Krallinger, M. (2019). Automatic de-identification of medical texts in Spanish: the MEDDOCAN track, corpus, guidelines, methods and evaluation of results. In *Iberian Languages Evaluation Forum (IberLEF 2019)*, volume 2421, pages 618–638. CEUR Workhop Proceedings.

Marrero, M., Sánchez-Cuadrado, S., Urbano, J., Morato, J., and Moreiro, J.-A. (2010). Sistemas de recuperación de información adaptados al dominio biomédico. *Información biomédica*, 19(3):246–254.

Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., and Gómez-Berbís, J. M. (2013). Named Entity Recognition: Fallacies, challenges and opportunities. *Computer Standards & Interfaces*, 35(5):482–489.

Névéol, A., Dalianis, H., Velupillai, S., Savova, G., and Zweigenbaum, P. (2018). Clinical natural language processing in languages other than English: opportunities and challenges. *Journal of Biomedical Semantics*, 9(1).

NLTK Project (2020a). NLTK. https://www.nltk.org. Last access: August 24, 2020.

NLTK Project (2020b). NLTK – nltk.stem package. https://www.nltk.org/api/nltk.stem.html. Last access: August 24, 2020.

Padró, L. (2008). FreeLing. http://nlp.lsi.upc.edu/freeling. Last access: August 24, 2020.

Padró, L. and Stanilovsky, E. (2012). FreeLing 3.0: Towards wider multilinguality. In *Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 2473–2479. European Language Resources Association (ELRA).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pons, E., Braun, L. M. M., Hunink, M. G. M., and Kors, J. A. (2016). Natural language processing in radiology: A systematic review. *Radiology*, 279(2):329–343.

Python Software Foundation (2020). re – regular expression operations in Python. https://docs.python.org/3/library/re.html. Last access: August 24, 2020.

Sahni, S. and Mehta, D. P. (2018). *Handbook of Data Structures and Applications, Second Edition*. Chapman and Hall/CRC.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47.

Simpson, M. S. and Demner-Fushman, D. (2012). Biomedical text mining: A survey of recent progress. In *Mining Text Data*, pages 465–517. Springer.

Singh, V. (2017a). FlashText Python module. https://flashtext.readthedocs.io. Last access: August 24, 2020.

Singh, V. (2017b). Replace or retrieve keywords in documents at scale. arXiv:1711.00046.

SNOMED International (2020). SNOMED CT. http://www.snomed.org/snomed-ct/why-snomed-ct. Last access: August 24, 2020.

Spasic, I., Ananiadou, S., McNaught, J., and Kumar, A. (2005). Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics*, 6(3):239–251.

Trieschnigg, D., Pezik, P., Lee, V., de Jong, F., Kraaij, W., and Rebholz-Schuhmann, D. (2009). MeSH up: effective MeSH text classification for improved document retrieval. *Bioinformatics*, 25(11):1412–1418.

Wu, C., Xia, F., Deleger, L., and Solti, I. (2011). Statistical machine translation for biomedical text: Are we there yet? In *AMIA Annual Symposium*, pages 1290–1299. American Medical Informatics Association (AMIA).

Yim, W., Yetisgen, M., Harris, W. P., and Kwan, S. W. (2016). Natural language processing in oncology. *JAMA Oncology*, 2(6):797.

APPENDIX: EXAMPLES OF TEXTS ANNOTATED USING SNOMED-CT AND DeCS

In this appendix, we show two examples of texts annotated by using the annotation tool developed in this work. The first text has been annotated considering the dictionary created with data from SNOMED-CT and the second text with the dictionary of terms of DeCS. It should be noted that some texts contain typos, as they are real texts written by doctors during their daily practice (no proof-reading has been applied to correct the potential mistakes; only some sensitive data, such as the age of a patient, have been removed from the original texts). In the examples, we use different colors to represent the terms that are correctly annotated (shown in light green), terms that are incorrectly detected but that are not really relevant (shown as strikethrough text), and terms not detected but that should have been detected as relevant (shown in **bold light red**).

Example 1: Annotation Using SNOMED-CT

Input Text: LUMBOCIATICADescripción de la(s) exploración(es): – EXPLORACIÓN: **RM de columna lumbosacra**, secuencias en ponderación T1 sagital, secuencia DIXON sagital y T1 y T2 plano axial. Hallazgos: Pérdida de la lordosis lumbar con rectificación. Abombamientos de platillos generalizados, pero con correcta altura de cuerpos vertebrales. Alineación anteroposterior conservada. Moderados signos espondilósicos con incipiente **osteofitosis** de predominio anterior. Salidas difusas circunferenciales discales. Disminución generalizada de intensidad de señal a nivel discal en T2 indicativo de deshidratación, mucho más evidente en los últimos niveles lumbares. Esclerosis interapofisaria asociada. – NIVEL L2-L3: bandas parcheadas de hiperseñal en T1 y T2 en platillos indicativos de cambios degenerativos tipo II. Salida difusa circunferencial discal. Leve deshidratación discal. Ligera esclerosis interapofisaria. – NIVEL L3-L4: salida difusa circunferencial discal. Leve deshidratación discal. Ligera esclerosis interapofisaria. – NIVEL L4-L5: salida difusa circunferencial discal. Pequeña hernia posteromedial del núcleo pulposo. Marcada deshidratación discal. Disminución del espacio intersomático. Marcada esclerosis interapofisaria, con hipertrofia. Se asocia con hipertrofia ligamentaria que disminuyen el calibre transverso del canal. En su conjunto se reconoce ligero compromiso de recesos laterales y de ambos forámenes secundario. – NIVEL L5-S1: Grandes bandas de hiperseñal en T2 y T2 en platillos, fundamentalmente en el inferior de L5, que indican cambios degenerativos tipo II. Importante disminución del espacio intersomático. Hiperintensidad de señal discal en T1 y T2, que se suprime con la secuencia saturación grasa, que indica recambio degenerativo graso discal. Marcada hipertrofia interapofisaria. Osteofitosis y protrusión ~~disco~~ osteofitaria posteromedial. Ligera disminución del calibre del canal. Diagnóstico: Nombre Responsable 1: [1] – Fecha de Firma: ZARAGOZA, N.Colegiado: Categoría Profesional 1: – Informe de Resultados de Pruebas de Imagen. Servicio de Radiodiagnóstico. Fecha de Impresión: – Pérdida de la lordosis lumbar con rectificación. – Signos espondilósicos con salidas difusas circunferenciales discales. Deshidratación es discales asociadas y esclerosis interapofisaria. – L2-L3: cambios degenerativos y II en platillos. – L4-5: disminución del espacio intersomático, esclerosis e hipertrofia interapofisaria ligamentaria con disminución de calibre transverso del canal. Compromiso de recesos laterales. Pequeña hernia posteromedial y de núcleo pulposo. – L5-S1: cambios degenerativos tipo II en platillos. Recambio graso discal. Marcada hipertrofia interapofisaria.

Results and Comments: 1) only the term "disco" ("disc") is a false positive in this text; although it could be considered a relevant medical term, in the text it is not used with the meaning attributed by the class associated to the term in SNOMED-CT ("drug, medicine") but rather as a part of the human spine; 2) in this case, the term "espondilosis" is detected only after applying a spell checker over the word "espondilósicos" (otherwise, it would have not been detected). The terms detected after applying a spell checker are subject to some uncertainty (as the application of a spell checker automatically could actually lead to a term different than the one intended in the original text), although in this case the detection is correct.

Example 2: Annotation using DeCS

Input Text: SINDROME CORONARIO AGUDO Paciente intervenido triple Bypass coronario AMI a DA y vena safena a Dx yDp .– se copiainforme, (se encuentra en OMI) le indican que han solicitado ~~consulta~~ en Cardiologia, pero no consta en su historico. Motivo del Alta: Curación o mejoría. Motivo inmediato del ingreso: Paciente de años de edad que ingresa procedente de Hospital Miguel Servet para cirugía coronaria urgente. Anamnesis: Antecedentes personales: Dudosa alergia a **Amoxcilina-Clavulanico** . Exfumador. No **HTA** . **DM tipo 2** (ADO). Dislipemia. Poliquistosis renal y ectasia pielocalicial derecha. Esteatosis ~~hepática~~. Bocio. Diagnosticado de **SAOS** . Reflujo Gastroesofágico. Hipoacusia. Intervenido de septoplastia. Colelitiasis. Colecistectomía. ~~Historia~~ Cardiológica: Estudiado por dolor torácico atípico en Medicina Interna, y Cardiología, con ergometría no sugerente de isquemia con 10 METS de carga en . El acude a ~~Urgencias~~ por clínica de ángor de ~~reposo~~ de algunas horas de duración, sin componente postural, y desencadenados por esfuerzo hace unas semanas. A su llegada a ~~Urgencias~~, nuevo dolor, realizando ECG que evidencia pseudopositivización de onda T, que desaparece tras comenzar pc de SLN + mínima elevación de TnUS (troponina ~~pico~~ 180), decidiendo ~~ingreso~~ en UCI. El se realiza coronariografía que evidencia enfermedad multivaso, es presentado en sesión ~~médico~~-quirúrgica decidiéndose cirugía en el ~~ingreso~~. Exploraciones Complementarias: **Ecocardiograma** : Cavidades cardiacas y Aorta ascendente de dimensiones normales. **HVI** ligera. Contractilidad global conservada, sin apreciar alteraciones segmentarias. Patrón de ~~relajación~~ disminuida, sin ~~elevación~~ de las **PTDVI** . **Válvulas** estructural y funcionalmente normales (VAo trivalva) Contractilidad normal del **VD** . **Cava** y **suprahepáticas** no dilatadas, sin inversión de flujos y normocolapso inspiratorio. No ~~signos~~ indirectos de **HTP** . No afectación pericárdica Cateterismo: Tronco: Sin lesiones. DA en segmento proximal presenta estenosis crítica y luego estenosis significativa respectivamente. 1ra diagonal: 1 mm con lesión significativa ostial. 2da diagonal. lesión significativa ostial. Lesión significativa en tercio distal de CX que in-

Table 5: Example of annotation with SNOMED-CT.

Detected term	Associated class (or associated classes)
cambio degenerativo	anomalía morfológica
canal	estructura corporal
cuerpo vertebral	estructura corporal
deshidratación	trastorno
~~disco~~	fármaco de uso clínico
disminución	anomalía morfológica
esclerosis	anomalía morfológica
exploración	procedimiento
grasa	sustancia, estructura corporal
hallazgo	hallazgo
hernia	anomalía morfológica
hipertrofia	anomalía morfológica
lordosis	trastorno
núcleo pulposo, L5-S1	estructura corporal
protrusión	anomalía morfológica
se reconoce	hallazgo
signo	hallazgo
espondilosis (trastorno)	trastorno

volucra ostium de rama marginal. Arteria Intermedia: Estenosis en límite de la significancia en segmento proximal. CD: Estenosis ligera tercio proximal. DP con estenosis ostial en límite de la significancia y lesión en tercio medio significativa. Procedimientos Terapéuticos: Fecha de la intervención: . Cirujano. Se realiza bajo CEC triple bypass coronario: AMI a DA y vena safena a Dx y DP. En quirófano inestabilidad hemodinámica con bradicardia extrema que precisa de entrada urgente en C.

Results and Comments: 1) most undetected terms are acronyms and abbreviations used by doctors (e.g., HTA, HVI, PTDVI, VD, HTP, etc.); 2) incorrectly detected terms correspond to terms that are not relevant in the medical context of the text or whose identification would change the real meaning of the word (e.g., "relajación" in the text has a different meaning that a social phenomenon, "pico" in the text has the meaning of "peak value" rather than a part of anatomy, "Urgencias" in the text represents the area of a hospital that receives patients that may have an emergency issue rather than an "urgency" in the medical sense, etc.).

Table 6: Example of annotation with DeCS (1/2).

Detected term	**Associated class (or associated classes)**
alergia	ENFERMEDADES - SALUD PÚBLICA
anamnesis	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
bocio	ENFERMEDADES - SALUD PÚBLICA
bradicardia	ENFERMEDADES
CEC	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
cardiología	DISCIPLINAS Y OCUPACIONES
cateterismo	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
cirugía	DISCIPLINAS Y OCUPACIONES
cirujano	DENOMINACIONES DE GRUPOS - ATENCIÓN DE SALUD

Table 7: Example of annotation with DeCS (2/2).

Detected term	**Associated class (or associated classes)**
colecistectomía	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
colelitiasis	ENFERMEDADES
~~consulta~~	ATENCIÓN DE SALUD
dislipemia	ENFERMEDADES
dolor	ENFERMEDADES - PSIQUIATRÍA Y PSICOLOGÍA - FENÓMENOS Y PROCESOS
dolor torácico	ENFERMEDADES
ECG	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
ectasia	ENFERMEDADES
~~elevación~~	FENÓMENOS Y PROCESOS
enfermedad	ENFERMEDADES - SALUD PUBLICA
ergometría	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS
estenosis	ENFERMEDADES
hemodinámica	FENÓMENOS Y PROCESOS
~~hepático~~	ORGANISMOS
hipoacusia	ENFERMEDADES
~~historia~~	HUMANIDADES
hospital	ATENCIÓN DE SALUD - VIGILANCIA SANITARIA - SALUD PÚBLICA
~~ingreso~~	ATENCIÓN DE SALUD - SALUD PÚBLICA
inversión	ATENCIÓN DE SALUD - SALUD PÚBLICA
isquemia	ENFERMEDADES
lesión	ENFERMEDADES - SALUD PÚBLICA
medicina interna	DISCIPLINAS Y OCUPACIONES
~~médico~~	DENOMINACIONES DE GRUPOS - SALUD PÚBLICA - ATENCIÓN DE SALUD
paciente	DENOMINACIONES DE GRUPOS
~~pico~~	ANATOMÍA
procedimiento terapéutico	TÉCNICAS Y EQUIPOS ANALÍTICOS, DIAGNÓSTICOS Y TERAPEÚTICOS - VIGILANCIA SANITARIA
quirófano	ATENCIÓN DE SALUD - VIGILANCIA SANITARIA
reflujo gastroesofágico	ENFERMEDADES
~~relajación~~	ANTROPOLOGÍA, EDUCACIÓN, SOCIOLOGÍA Y FENÓMENOS SOCIALES
~~reposo~~	ANTROPOLOGÍA, EDUCACIÓN, SOCIOLOGÍA Y FENÓMENOS SOCIALES
~~signo~~	ENFERMEDADES
síndrome coronario agudo	ENFERMEDADES
troponina	COMPUESTOS QUÍMICOS Y DROGAS
UCI	ATENCIÓN DE SALUD - VIGILANCIA SANITARIA
~~urgencia~~	ENFERMEDADES - ATENCIÓN DE SALUD