Towards Cross Language Morphologic Negation Identification in Electronic Health Records

Ioana Barbantan(✉) and Rodica Potolea

Computer Science Departament, Technical University of Cluj-Napoca,
Gh. Baritiu St 28, Cluj-Napoca, Romania
{ioana.barbantan, rodica.potolea}@cs.utcluj.ro

Abstract. The current paper presents an approach for analyzing the Electronic Health Records (EHRs) with the goal of automatically identifying morphologic negation such that swapping the truth values of concepts introduced by negation does not interfere with understanding the medical discourse. To identify morphologic negation we propose the RoPreNex strategy that represents the adaptation of our PreNex approach to the Romanian language [1]. We evaluate our proposed solution on the MTsamples [2] dataset. The results we obtained are promising and ensure a reliable negation identification approach in medical documents. We report precision of 92.62 % and recall of 93.60 % in case of the morphologic negation identification for the source language and an overall performance in the morphologic negation identification of 77.78 % precision and 80.77 % recall in case of the target language.

Keywords: Cross language · Morphologic negation · Electronic health records · Dictionary

1 Introduction

The evolution of technology acquaints us frequently with specialized gadgets that can influence our everyday life. We gain access to devices that manage our eating and exercise habits, monitor our heart rate and inform us of the calories we burn and consume or translate our activity to statistical dimensions. The English language became ubiquitous; we use English terms when referring to our computer components and the actions we can carry out using them or when sharing our thoughts and feelings on the social media. The devices nowadays have English imprints on them; when entering a store we usually find the Open/Closed sign more often than the corresponding information for the native language in each country. For the young generation these aspects do not represent issues as a large percentage of the young population in every country is familiar with the English language. Issues arise when these devices are used by elderly people whose existence was not overwhelmed by the adoption of the English language and the rapid evolution of technology. Usually, the main topic of interest for this category of population is represented by the development of their health.

As technology evolves, the number of medical devices that we gain access to grows as well; nowadays, we can easily send our health status by means of these devices to

R. Agüero et al. (Eds.): MONAMI 2014, LNICST 141, pp. 417–430, 2015.
DOI: 10.1007/978-3-319-16292-8_30

our medical doctors that automatically fill up our electronic health records with this new information. The problem is what to do when the persons that need to use the devices are not familiar with the language in which the instructions are presented or the information displayed on them.

The EHRs capture the medical history and current condition with detailed information about symptoms, surgeries, medications, illnesses or allergies. They are an important source of new information and knowledge if exploited correctly. From these documents we can retrieve new ways of how diseases interact with each other, the influence of demographics on the patients' conditions and many more. But in order to do this, the documents need to be clear, carry trustworthy information and should be unambiguous. In most cases the EHRs are unstructured documents and may contain recurrent information. The problem we address in this paper is defining a strategy for identifying negation in EHRs, towards retrieving relations among medical concepts. We propose an approach for adapting for the Romanian language to our already established methodology for English [1]. In both languages negation is of syntactic and morphologic types and a correspondence between the negation concepts is easily noticeable.

The main contribution of the paper addresses the existing drawback of several negation identification approaches that do not consider negation represented using negation prefixes in both languages. We propose a strategy that includes negation prefixes in identifying and dealing with the negative concepts in the EHRs. In order to tackle morphologic negation, our strategy is based on interpreting the structure of the words and evaluating the existence of the words with and without prefix in the language by taking into account the definitions provided in a dictionary specific to the source respectively the target languages.

The rest of the paper is organized as follows. In chapter II we present similar systems dealing with EHRs and negation. In chapter III the EHRs are briefly introduced along with the two most common negation types. Our solution is detailed in chapter IV where we describe the RoPreNex strategy and present the experiments we performed in chapter V. The last two chapters include the conclusion of our work and future enhancements for the approach we propose.

2 Related Work

Both in Romanian and English languages the task of identifying negation is mostly focused on negation expressed with specific words like *nu, fara, nici* in Romanian or *not, without, nor,* the English corresponding words. Morphologic negation is disregarded and there are even cases when it is otherwise inferred. For example, the authors in [3] talk about negation prefix when dealing with the word *not* and refer to it as negation prefix, whereas they are dealing with syntactic negation based on Givon's negation classification [4]. They present a system that identifies the n-words that represent negative quantifiers in order to determine the negative concord for the Romanian language.

A cross-lingual approach for document summarization is proposed in [5]. The authors evaluate how a cross language approach could help in spreading the news around the world when dealing with ordinary but not breaking news that are easily propagated among websites. The authors use as source language Romanian and translate the summarized information into English. The translation is performed using a bidirectional English-Romanian translation tool. The authors evaluate the performance of their approach by asking a set of questions to judges. The questions are regarding the Romanian summarizations and then the same set of the questions were asked for the translated summaries. An accuracy of 43 % is reported in the case of giving correct answers for the summarized documents. Most of the questions that could not have been answered are due to the fact that the translated summaries were not clearly understood.

Negation in medical documents is subject of interest in the medical domain as the diagnosis process the stated and missing or denied symptoms are weighted differently. There are several approaches dealing with identifying and labeling negation, like NegEx [6], Negfinder [7], or the tool presented in [8] developed for the BioScope negation annotated corpus. The main drawback of these tools is the absence of treating morphologic negation, hence leaving out several negated terms, especially when dealing with medical documents.

One reason for not considering morphologic negation is motivated by the authors in [9] by the few occurrences of these terms or by considering the prefixes as not determining negations [10]. In [7] negation is defined only when the negation terms negate subjects or object concepts (no, without, negative) and specify that even though there are concepts that have negative connotations (like akinesia) they are disregarded and report these cases as miscellaneous errors. These approaches are valid when dealing with data that is not domain dependent or in cases when the negation algorithm is meant to find all concepts that can be determined by a single negative identifier. In the case of medical records, (domain dependent documents) the negations are prevalent as in the medical language negation prefixes are broadly used. In medical documents is it expected that negation is clearly formulated as these documents should be clear and carry as few ambiguous terms as possible.

The analysis of morphologic negation is presented as a future enhancement for the work of the authors in [9], where they predict a growth in the performance of identifying the scope and focus of negation by removing the prefix and determining the validity of the obtained word. They introduce how negation in Natural Language is characterized and present an approach of automatically determining the scope and focus of negation. The frequency of the negation-bearing words in the corpus they use leads to considering negation only the determiners not and n't. The scope of negation was identified with 66 % accuracy.

Negation in medical documents is approached by Averbuch et al. in [11] that report that including negation in information retrieval improves precision from 60 % to 100 % with no significant changes in recall. They also state that the presence of a medical concept in the record, like a symptom, does not always imply that the patient actually suffers from that condition as the symptom can be negated.

Capturing word's semantics and relationships with the help of a dictionary in order to categorize a text is presented in [12]. The text is disambiguated and represented as features using the concepts and hypernymy relations in WordNet. The authors compare the results of text categorization when using a bag of words approach for document representation and when using the WordNet information for selecting the features.

They evaluate the methods on two datasets and notice that the WordNet approach exceeds in all test conditions the bag of words approach.

Negation has application in sentiment analysis when the opinion (positive, negative or neutral) is in question. In sentiment analysis the goal is to identify the polarity of assertions that can be positive or negative [13]. Usually this is done using specific words for the polarity categories [14]. The BioScope negation annotated corpus is used for evaluation in order to extract the polarity of the sentences using a Conditional Radom Field approach and a dependency parser [9]. The authors report achieving a 75.5 % F1 score on the BioScope corpus, a medical corpus and 80 % F1 score when using a product reviews corpus.

3 Theoretical Background

This section attempts to set the background of the work presented in this paper by introducing the main concepts we operate with: EHRs and the need to structure them, the role of handling negation in EHR concept extraction and structuring and the cross language strategies.

3.1 Electronic Health Records

The EHRs are unstructured or semi-structured text documents carrying medical information about patients. The HL7 standard announces as their main purpose being to provide the medical history and current condition [15]. The access to these medical documents is restricted to the medical personnel and the sharing of the information from these documents needs authorization.

The content of the EHRs can be organized into categories like symptoms, procedures, surgeries, medications, illnesses or allergies. If exploited correctly, the EHRs can offer information about future epidemics, or can be used to predict the status of patients having similar conditions.

The standard indicates that the EHR information should be captured using standardized code sets or nomenclature or even unstructured data. When the raw data is unstructured there are no straightforward mechanisms to infer knowledge out of data. When the presence of a symptom is noticed in an EHR, a semantic analysis is required to determine whether the occurrence is an affirmative or a negated one.

3.2 Negation

When performing the anamnesis for a patient the medical doctors are interested whether the patient suffers or not from different symptoms, and based on the responses from the patient, a diagnosis is established. When querying a data source for patients with similar conditions, it is important for the machine to distinguish between the negated and affirmed symptoms. Negation can be expressed in different ways as the patient can state he has fever or he is afebrile. In this case it is important to treat all types of negations that occur in documents.

Givon classifies negation as syntactic negation in the case of explicit negation and morphologic negation when using prefixes [4]. Expressing the symptomatology of a patient the following three sentences can be used.

- *The patient has no symptoms.*
- *The patient is asymptomatic.*
- *The patient doesn't have symptoms.*

As underlined in the three examples, negation can be expressed using explicit terms like no and n't but can also be expressed with a prefix a (*aymptomatic*).

Syntactic Negation is introduced by specific negation terms like *no, without, deny, not, rule out* in case of the English language or *fara, neaga, eliminia* in case of Romanian are commonly found in natural language (the two lists of terms are not correspondent). Unlike morphologic negation that is associated with a single word, syntactic negation can determine several words like in the case of an enumeration of symptoms: *The patient presented without fever, neck pain or tiredness*.

Morphologic Negation is expressed with specific negation prefixes placed in front of the words to alter their meaning by swapping their truth value. They support enhancement of the vocabulary by increasing the number of words. The prefixes can also be used in learning new terms as presented in the study in [16]. The separation of a word into prefix and root form helps in understanding the meaning of the words. Table 1 captures the negation prefixes in the English and Romanian languages and their correspondence.

Table 1. Correspondence of English and Romanian negation prefixes.

English negation prefix	Romanian negation prefix	Meaning	English example	Romanian example
In, il, im, ir	In, i	Negative prefixes	Insufficiency	Insuficient
A, an	A	Not, without, lacking	Afebrile	Afebril
Non		Absence, negation	Nonsurgical	Nechirurgical
Dis	Des, dez, de, ne	Negation, removal, expulsion	Discontinue	Discontinuu
Anti	Anti, Contra	Opposing, against	Anti-inflammatory	Antiinflamator
Un	Ne	Not, reversal, cancellation, deprived of	Uncomplicated	Necomplicat

3.3 Cross Language Approach

We define our approach for solving specific Natural Language Processing (NLP) tasks in a new language (we will call it target language) once a solution is set up in some other language (we will call it source language). In our case the source language is English, the target language is Romanian, and the task is negation identification in

EHRs. However, the approach is not limited to specific languages and/or tasks. Once in a source language an efficient solution for text processing has been identified, the approach defines a way of adapting it to the target language.

A cross language strategy to perform sentiment analysis for identifying subjective sentences is proposed in [17]. While translating queries from English to Indonesian the authors in [18] show that using a collection of dictionaries rather than a single dictionary significantly improves the results.

The English language covers a large amount of everyday life subjects, so, we should try to benefit in any possible way from this. The concepts used in Computer Science especially but also in other scientific and professional fields tend to be English. The medical domain makes no exception. Terms like *bypass* or *follow-up* became familiar to every single one of us [19].

4 Methodology for Morphologic Negation Identification in EHRs

In our work so far, we have implemented several strategies for identifying negation in medical documents (needed to further structure the EHRs). In [20], we employed a vocabulary of terms and a binary bag of words feature vector, while in [1] we replaced the vocabulary obtained with a dictionary of the English language. Our current work proposes cross-language strategy that deals with identifying morphologic negation in a target language based on an already established strategy for a source language.

4.1 Cross Language Strategy for Morphologic Negation Identification

Given the source language solution for negation identification, we define a methodology for adapting the solution for a target language. However, the methodology makes no restriction to the choice of languages. Provided the linguistic resources for other languages are available, the same strategy may be applied. Our specific goal is to instantiate the cross language methodology that identifies morphologic negation in both the source and target languages using the linguistic resources in the corresponding language. The source language is English, for which we have proposed and evaluated the PreNex strategy [1] and the target language is Romanian, which represents the subject of our current approach.

The resources we employ in our analysis consist in a dataset of EHRs available in English. The strategy starts by translating them into Romanian using an online translation service in order to obtain standardized documents and a reliable comparison. We propose a dictionary based approach where we identify morphologic negation. The dictionaries we use are WordNet[1] for English and DexOnline[2] for Romanian.

1 http://wordnet.princeton.edu/.

2 http://dexonline.ro/.

4.2 Dictionary Based Negation Extraction

In our previous approach, we proposed identifying morphologic negation from English medical documents, namely from the MTsamples [2]. We proposed a dictionary based approach [1] that exploits the meaning of the words by using an English language dictionary, namely WordNet and proposed several rules that semantically exploit the meaning of the words.

The rules for negation identification used in the PreNex approach are:

Definition recurrence rule: the root of a prefixed word is contained in the prefixed word's definition.

Definition content rule: both the root of the prefixed word and the prefix word are defined in WordNet and the definition of the prefixed word contains a negation identifier.

Hyphen rule: the prefix is followed by hyphen or space – the case is handled by removing the special character and sending the entity to be analyzed with the previous rules.

Compound words: progressively build a word from consecutive letters on an n-gram basis; remove the prefix and perform an analysis of the root. If the word can be split into two words with definitions in WordNet, we consider the word negated with negation prefix.

The rules were added progressively and as it can be noticed in Table 2, at each step improvements in the precision and recall were obtained. The first rule identifies negation prefixes with a precision of 95.07 % and a small recall of 29.09 %. The following rules added introduce degradation in precision, but the decrement of 2.45 % becomes irrelevant in relation to the increase of 64.51 % achieved for recall. We report as final results in the case of the PreNex strategy, precision of 92.62 % and recall of 93.60 %.

Table 2. PreNex performance.

PreNex rules	Precision (%)		Recall (%)		F-measure (%)	
(1): R1	95.07		29.09		44.55	
(2):(1)+R2	91.56	-3.55	74.78	45.69	82.35	37.77
(3):(2)+R3	92.81	1.25	89.01	14.22	90.87	8.54
(4):(3)+R4	92.62	-0.19	93.60	4.59	93.11	2.23
Overall improvement	-2.45		64.51		48.55	

The false positives introduced by the proposed solution are "infusion", "absolute", "intensity" or "another".

4.3 Dataset

In order to evaluate our approach we used a dataset of English EHRs provided by [2]. These are semi-structured documents that contain medical information about hospitalized patients. They present the evolution of the patient from the point they were admitted in the hospital to the point of discharge. The documents capture the symptoms, medical history, the procedures performed and the administered medication. There are cases when the patient is required to return to the hospital for a follow-up examination, in which case the conditions and details about the appointment are also established in the document.

As our current goal is to analyze medical documents for the Romanian language, we propose an adaptation of the morphologic negation identification proposed for English. As we want to make sure that the documents we send for analysis are compliant with the medical standards, we translated the documents to Romanian. Also, the amount of available annotated EHRs for Romanian is not satisfactory for a reliable analysis. In order to obtain the Romanian version of the EHRs, we used an online translation tool to obtain the correspondence of the medical documents between the two languages. There were cases where the translation tool employed could not translate all terms due to the fact that the words were not found in the English dictionary or in the English-Romanian dictionary. This issue was encountered in the case of the word *nontender or nonfasting* which are domain specific terms, in our case, the medical one.

The proposed methodology of evaluating the negation identification for the Romanian language follows the steps in Fig. 1. First, the corpus of documents used in the English language is translated for a reliable comparison. Then we preprocess the documents and apply the proposed negation identification rules. The last step employed is the proposed adapted strategy for identifying negated concepts.

Fig. 1. Input data preparation.

4.4 RoPreNex Algorithm

In the current approach we propose identifying morphologic negation in medical documents written in Romanian, by adapting the strategy proposed for the English language. As presented in Table 1, a correspondence between the English and Romanian negation prefixes is noticeable.

We propose the RoPreNex algorithm to identify morphologic negation in Romanian medical documents, presented in Fig. 2. The rules consider the existence of the words and their root form in the DexOnline dictionary and also the content of their definition. DexOnline represents the Romanian language dictionary and consists of a collection of Romanian dictionaries. The dictionary interlinks the words with their definitions and has also integrated synonyms and the newest words that appeared in the language after 2004, when the integration of the dictionary on paper completed.

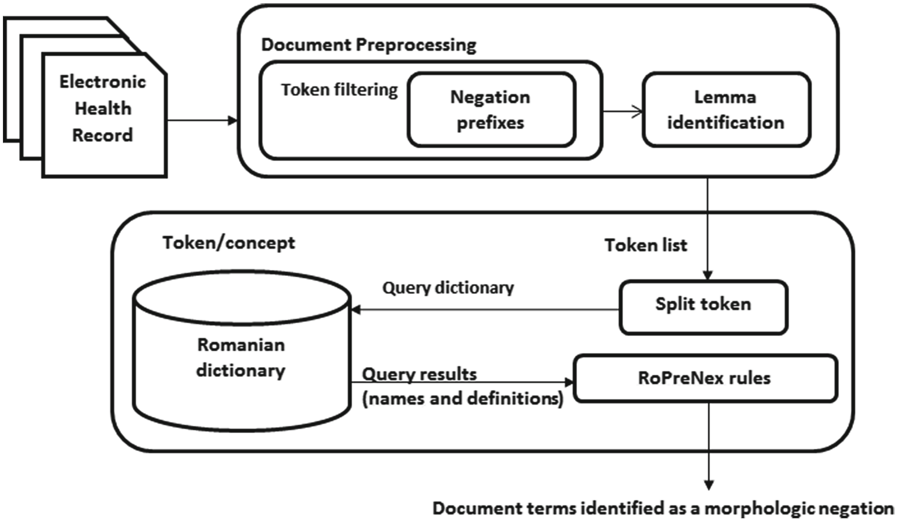

Fig. 2. RoPreNex flow.

There are several changes that had to be applied to our PreNex algorithm. As the DexOnline dictionary also contains definitions for the words that are not frequently used in the language (like regionalism or rural expressions) we must include an additional verification step such that these words do not interfere with our search. In case of the rural expressions, the words are truncated or the first letters may be removed in which case we might deal with a false case of root word.

The Romanian lemmatization tools are less efficient than the ones existing for English. Moreover, they should be accessed via a web service, which induces time overhead (also a less reliable solution), so we propose a lemmatization process that is able to bring the words in the documents to their dictionary form. The approach works as follows. For each word in the documents that was selected as possible negated concept, we remove its prefix and before determining its truth value, we preprocess it. The approach we propose considers the termination of the words. Usually the difference between the words in the document and their dictionary form appears in the added termination that announces the inflections (e.g. *e* is added for the plural of nouns or *em* in the case of verb tenses). When a match between the preprocessed word from the document and the words in the dictionary, we send to the negation identification rules the currently preprocessed word.

The RoPreNex strategy is presented in the algorithm below. For the morphologic negation identification task we proposed the following rules *Literal words, Definition content, Undefined prefixed word*.

Literal words. The DexOnline dictionary also contains definitions for the words that are not used in common language like regionalism or rural expressions. In this case the words are shortened and the first letters are removed when expressing the words, case in which they could falsely represent prefixes. This rule is a preprocessing step applied for the words in the dictionary. The following rules are the actual negation identification rules.

Definition content. The definition content rule identifies negation based on the definition of the word. First, we identify the prefix, remove it from the word, and obtain the root of the word. If the root and the prefixed word exist in DexOnline, we check whether the word's definition contains at least one negation identifier.

Undefined prefixed word. The undefined prefixed word rule is applied in the cases when the prefixed word is not defined in the dictionary as it could represent a domain specific term. In this case we remove the prefix and determine whether the root of the word is defined in the dictionary.

Algorithm notations:

ω – the possible prefixed word with negation prefix

ω̃ – the root of ω

ρ – the prefix of ω { anti, dez, des, de, ne, in, a, im, contra }

definition(ω) – the definition of a word

defined(ω) – the word is defined in dictionary

literal(ω) – the word is in its literal form

The algorithm of determining whether ω is prefixed with a negation prefix works as follows:

```
Input = ω –possible prefixed word with negation prefix
ρ = prefix of ω
ω̃ = Remove ρ from ω
//Literal words
If (literal(ω)) (literal(ω̃))
    ω <- lemma(ω)
    ω̃ <- lemma(ω̃)
    //R1. Definition content
    If (definition(ω) && definition(ω̃))
    If (definition(ω) contains one negation identifier)
          Then ω is negated with prefix
    Endif
    Endif
    Else //R2. Undefined prefixed word
    If (!defined(ω) and defined(ω̃))
          Then ω is negated with prefix
    Endif
Endif
END
```

5 Morphologic Negation Evaluation RoPreNex

This section presents the experiments performed with the proposed morphologic negation identification strategy for Romanian medical documents, and a comparison with PreNex, our original solution for English [1]. We evaluate our approach on the translated MTsamples dataset, presented in more detail in Sect. 4.3.

Table 3. RoPreNex algorithm rules coverage.

Rule	% of covered cases
Definition content	78.57
Undefined prefixed word	21.43

In our proposed strategy, most of the prefix negated concepts are identified by the *Definition content* rule, as most of them are defined in both representations: as prefixed word and root form, Table 3, line 2. A smaller percentage of the prefixed concepts are not defined in the dictionary, and for this case we had to introduce the second rule *Undefined prefixed word,* Table 3, line 3 which covers 21.43 % of the correctly identified concepts.

5.1 RoPreNex Performance

The rules we propose for identifying morphologic negation in Romanian medical documents are promising as can be seen in Table 4, last line. The performance of each rule is presented also in Table 4. Even though the *Undefined prefix word* rule has a small value for recall, as it misses some of the words it was supposed to identify, the great value for precision makes it an important rule for our approach. Making a tradeoff between the quality and number of correctly identified negated concepts, we report an overall performance of precision of 77.78 % and recall of 80.77 %. The PreNex strategy outperforms the current RoPreNex strategy with only 14.85 % in case of precision and 12.83 % in case of recall, as presented in Table 5. To the best of our knowledge a tool for dealing with morphologic negation has not been developed for the Romanian language, that is why we report our results to our English version of the algorithm.

The performance of our proposed strategy is satisfactory taking into account the fact that the documents we evaluated were translated and not original and the fact that we did not include any language specific methodologies for text analysis.

Table 4. RoPreNex performance.

Rule	Precision (%)	Recall (%)
Definition content	76.74	84.62
Undefined prefixed word	81.82	69.23
Overall performance	77.78	80.77

Table 5. Performance of negation identification strategy for Romanian and English strategies.

Approach	Precision (%)	Recall (%)
Romanian approach (RoPreNex)	77.78	80.77
English approach (PreNex)	92.62	93.60

5.2 Discussion

The reported performance of our proposed approach is promising, and we encountered the following problems addressed at the level of translation tool employed, at the word level, and at dictionary level.

Translation level issues. The translation tool we employed in our approach did not manage to perform a one-on-one translation. There were cases when in the translated document we encountered English words like *nontender* or *nonfasting*.

Word level issues. Other issues we came across were related to the fact that for the Romanian language we could not employ a lemmatizer that could help in normalizing the words such that we could obtain their dictionary format. Using our lemma implemented approach we managed to increase the recognition rate, but when the inflectional form of the word changes the root's structure it still remains an issue that has to be addressed. We also found cases when the words are shortened in the source language and the translation tool could not translate the word in the target language, like in the case of the word *noncontrib*.

Dictionary level issues. The DexOnline dictionary we used in our approach is populated with most of the words that exist in the Romanian language and also with the newest terms entered in the language. But, there still are cases when the dictionary fails to capture information about specialized terms like *atraumatic*.

The false positives introduced by our algorithm are usually represented by words that have a negation specific prefix but are not actually negated words. Like in the case of the word *informat* whose English correspondent is *informed*. In this case the word matches all rules we defined as the word and its root are both defined in the dictionary, but do not represent a negated entity.

6 Conclusions and Further Work

In this paper we propose a methodology for identifying morphologic negation in Romanian medical documents. It is the adaptation of our proposed solution on English documents, towards a cross-language NLP strategy for medical documents mining. The results we report for identification are promising as to our best knowledge there are no similar approaches for identifying morphologic negation for the Romanian language. The results obtained by our proposed methodology are precision of 77.78 % and recall of 80.77 % %, which are only 14.85 % in case of precision and 12.83 % in case of recall below the corresponding English results. We consider our methodology reliable when applied on medical documents as the false positives that are introduced are not medical related concepts. Our current work consists in enhancing the identification performance of morphologic negation for the Romanian language.

We consider improving the performance of this strategy by first preprocessing the documents by employing a spell checker for each language or a distance measure algorithm that could correct the form of the misspelled words. Another task that we consider is represented by treating the abbreviations and word shortenings as the medical terms can appear with a code mostly in the case of the diseases.

Acknowledgments. The authors would like to acknowledge the contribution of the COST Action IC1303 - AAPELE.

References

1. Barbantan, I., Potolea, R.: Exploiting Word Meaning for Negation Identification in Electronic Health Records, IEEE AQTR (2014)
2. MTsamples: Transcribed Medical Transcription Sample Reports and Examples. Last accessed on 23.10, 2012
3. Iordachioaia, G., Richter, F.: Negative concord in Romanian as polyadic quantification. In: Muller, S. (ed.) Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar Georg-August-Universitat Gottingen, pp. 150–170. CSLI Publications, Germany (2009)
4. Givon, T.: English Grammar: A Function-Based Introduction. Benjamins, Amsterdam, NL (1993)
5. Orasan, C., Chiorean, O.A.: Evaluation of a cross–lingual romanian–english multi–document summariser. In: Proceedings of LREC 2008 Conference, Marrakech, Morocco (2008)
6. Chapman, W., Bridewell, W., Hanbury, P., Cooper, G.F., Buchanan, B.G.: A simple algorithm for identifying negated findings and diseases in discharge summaries. J. Biomed. Inform. **34**(5), 301–310 (2001)
7. Mutalik, P.G., Deshpande, A., Nadkarni, P.M.: Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. J. Am. Med. Inform. Assoc. **8**(6), 80–91 (2001)
8. Vincze, V., Szarvas, G., Farkas, R., Móra, R., Csirik, J.: The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. In: Natural Language Processing in Biomedicine (BioNLP) ACL Workshop Columbus, OH, USA (2008)
9. Blanco, E., Moldovan, D.: Some issues on detecting negation from text. In: Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference (2011)
10. Councill, I.G., McDonald, R., Velikovich, L.: What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In: Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, Uppsala, pp. 51–59 (2010)
11. Averbuch, M., Karson, T.H., Ben-Ami, B., Maimond, O., Rokachd, L.: Context-sensitive medical information retrieval, In: Proceedings of AMACL, pp. 282–286 (2003)
12. Elberrichi, Z., Rahmoun, A., Bentaalah, M.A.: Using WordNet for text categorization. Int. Arab. J. Inf. Technol. **5**(1), 16–24 (2008)

13. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the 14. Association for Computational Linguistics (ACL), Philadelphia, pp. 417–424 (2002)
14. Rokach, L., Romano, R., Maimon, O.: Negation recognition in medical narrative reports. Inf. Retrieval **11**(6), 499–538 (2008)
15. Fischetti, L., Mon, D., Ritter, J., Rowlands, D.: Electronic Health Record – system functional model, Chapter Three: direct care functions (2007)
16. Stahl, S.A., Shiel, T.G.: Teaching meaning vocabulary: productive approaches for poor readers. Read. Writ. Q. Overcoming Learn. Difficulties **8**(2), 223–241 (1992)
17. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-03) (2003)
18. Hayuran, H., Sari, S., Adriani, M.: Query and document translation for english-Indonesian cross language IR. In: Peters, C., et al. (eds.) Evaluation of Multilingual and Multi-modal Information Retrieval. Lecture Notes in Computer Science, vol. 4730, pp. 57–61. Springer-Verlag, Heidelberg (2007)
19. Frinculescu, I.C.: An overview of the english influence on the Romanian medical language. Sci. Bull. Politehnica Univ. Timişoara Trans. Mod. Lang. **8**, 1–2 (2009)
20. Barbantan, I., Potolea, R.: Towards knowledge extraction from electronic health records - automatic negation identification. In: International Conference on Advancements of Medicine and Health Care through Techonology. Cluj-Napoca, Romania (2014)