Spa-neg: an approach for negation detection in clinical text written in Spanish

Oswaldo Solarte-Pabón[1 [0000−0003−0315−2838]], Ernestina Menasalvas[2 [0000−0002−5615−6798]], and Alejandro Rodriguez-González[2 [0000−0001−8801−4762]]

[1] Universidad Politécnica de Madrid, Spain
oswaldo.solartep@alumnos.upm.es
[2] Universidad Politécnica de Madrid, Spain
{ernestina.menasalvas,alejandro.rg}@upm.es

Abstract. Electronic health records contain valuable information written in narrative form. A relevant challenge in clinical narrative text is that concepts commonly appear negated. Several proposals have been developed to detect negation in clinical text written in Spanish. Much of these proposals have adapted the Negex algorithm to Spanish, but obtained results indicated lower performance than Negex implementations in other languages. Moreover, in most of these proposals, the validation process could be improved using a shared test corpus focused on negation in clinical text. This paper proposes Spa-neg, an approach to improve negation detection in clinical text written in Spanish. Spa-neg combines three elements: i) an exploratory data analysis of how negation is written in the clinical text, ii) use of regular expressions best adapted to the way in which negation is expressed in Spanish, iii) tests, and validation using a shared annotated corpus focused on negation. Obtained results suggest that the combination of these elements improves the process of negation detection. The tests performed shown 92% F-Score using IULA Spanish, an annotated corpus for negation on clinical text.

Keywords: Negation Detection · Electronic health records · Clinical natural Language Processing.

1 Introduction

Negation detection is a challenging problem in the information extraction field. This is particularly important in systems aimed to extract knowledge from clinical information, where the detection of negation is key to understand symptoms, diagnoses, or treatments [5].

Negation detection is typically divided into two subtasks: cue identification and scope recognition. The cues are words or terms that express negation (e.g., without, not, nothing) [12]. The expressions *negation terms* and *negation phrases* can also be used to refer to negation cues. The scope is a text fragment affected by the corresponding negation cue in a sentence [14].

One of the most used algorithms to detect negation in medical records is Negex [6]. This algorithm has been recognized as one of the most useful approaches for the detection of negated medical concepts. However, it has been studied that can present deficiencies when terms that express negation appear contiguously several times, or when multiple instances of the negation affect the same medical concept [16]. Despite the acceptance that Negex has had, it can be improved. In the case of the Spanish language, several studies have been proposed to detect negation in clinical texts [8] [28] [18]. These proposals adapted the Negex algorithm to Spanish but obtained lower results than Negex implementations in English. Additionally, in these proposals, the validation process could be improved using a shared test corpus focused on negation in clinical texts.

This paper proposes Spa-neg, an approach to detect negation on clinical texts written in Spanish. Spa-neg combines three elements to improve negation detection: i) an exploratory data analysis to understand how negation appears in clinical texts, ii) adaptation of regular expressions based on the data analysis, and iii) a validation using corpus focused on annotating negation in clinical texts. Obtained results suggest that the combination of these elements improves the process of detecting negation. When performing the tests using a corpus annotated by experts and focused on negation in clinical texts, a 92% F-Score was obtained.

The rest of the article is organized as follows: section 2 describes proposals that address negation detection, with an emphasis on proposals for the Spanish language. Section 3 shows an exploratory data analysis of the negation. The Spa-neg proposal is described in detail in section 4. Section 5 describes the tests that were carried out to measure the performance of Spa-neg proposal. Finally, section 6 shows conclusions and future work.

2 Related Works

In the field of clinical text, several proposals have been developed for negation detection. One of the most relevant is Negex [6], an algorithm which has been widely used and adapted to several languages [4] [27] [9]. This algorithm uses two regular expressions and a set of negation terms that activate the negation. The Negex algorithm also uses a set of termination terms to indicate the negation scope [7]. Negex has been implemented in several clinical information extraction tools such as MetaMap [1] or Apache cTAKES [26].

In [3], a strategy that uses the syntactic properties of the sentence to calculate the scope is proposed. Another study similar to the previous one is proposed in [23], where it is described an algorithm that incorporates the use of a dependency tree within Negex.

On the other hand, [17] describes NegMiner, a proposal that aims to solve some of the weaknesses of the Negex algorithm. For example, when contiguous negation terms and multiple negation expressions appear in the same sentence. The author argues that Negex algorithm is still not sufficiently robust since many sentences are more complex to be correctly processed by just a couple of regular expressions. To solve these shortcomings, NegMiner proposes new rules to deal with contiguous negation terms and to detect various expressions of negation within the scope of a medical concept. Experimental results showed a higher performance than the Negex algorithm.

According to [12], machine learning approaches have also been used as a technique in detecting negation in medical texts. In [24], [2] and [13] are proposed machine learning-based classifiers to detect negation. Algorithms such as Support Vector Machines (SVM) and Conditional Random Fields were used. Tests were performed using the Bioscope corpus [30], which is focused on the English language.

Although the proposals mentioned above show significant advances and improvements in negation detection, most of these works have focused on the English language. According to [25], information extraction in the medical domain also represents its own challenges in languages other than English.

2.1 Negation detection on clinical text written in Spanish

One of the proposals to detect negation on clinical texts in Spanish was described in [8]. This proposal adapts Negex using an approach that first translates negation terms from Negex to Spanish, and then calculates the frequency of negated terms. Reported results show that when calculating the frequency of the terms that activate the negation, large differences were found between the terms used in English and Spanish. When Negex adaptation was evaluated, a large number of false positives were found, and an F-score value of 84% was obtained. This value was lower than the value obtained by Negex implementations in English. The authors suggest that one of the reasons for the high number of false positives might be because of the structure in which medical texts are written in Spanish differs from English. Therefore, the rules implemented in Negex do not always work properly for negation in Spanish. Others Negex adaptations to Spanish were also developed in [28] and [18]. In these proposals, a wider set of negation terms are used to indicate negation in Spanish. Just as the previous proposal, these are adaptations of the rules proposed by Negex.

In [10], an adaptation of Negex is proposed by adding syntactic properties such as part of speech (POS) tagger and a syntactic dependencies tree. In this proposal, the syntactic properties were used to perform a manual identification of patterns that are used to calculate the negation scope. An interesting finding to highlight in this proposal is that in the tests carried out, they reported better results when combining Negex with POS tagger properties. However, this adaptation only used the original rules proposed in Negex. This may represent a weakness in certain cases, such as those reported in [17].

In addition to syntactic properties, [20] proposes the use of pragmatic properties and grammar rules previously created to detect negation in radiology reports. Although they show promising results, one weakness is that, like the previous proposals, the performed tests lack the use of shared annotated corpus, focused on negation.

Finally, to address the lack of annotated corpus focused on the negation in Spanish, several studies have been carried out [19] [15] [22]. Specifically, [11] and [21] have developed an annotated corpus on the negation in medical texts in Spanish. These corpus are manually annotated by experts and help to validate tools that automatically detect negation. According to [29], a shared annotated corpus is very useful for developing natural language processing tools in the medical domain.

3 Exploratory data analysis of negation

This data analysis aims to obtain indicators to understand how negation is expressed in clinical texts written in Spanish. It contains extracted indicators such as length of sentences, number of negation terms by sentence, types of negation, and frequency of negation.

3.1 Datasets

– ***Dataset 1:*** a random sample of 3000 medical reports, which have been previously anonymized. This sample represents 20% of a database that contains around 15000 medical reports of a hospital in Spain.

– ***Dataset 2:*** The corpus "IULA Spanish", created by [21]. This corpus is a shared data set, in which experts have manually annotated negation in clinical text written in Spanish. It contains 3194 sentences taken from clinical documents. Each annotation includes the term that activates the negation and their scope. This corpus is available in:
http://eines.iula.upf.edu/brat/#/NegationOnCR_IULA/.

3.2 Results from dataset 1

– **Length of the sentences:** Figure 1 shows the distribution of sentences where negation appears and its length measured in the number of tokens. A token is the result of the tokenization of the sentence text in atomic elements such as words, numbers, or acronyms. The length of sentences is distributed between 1 and 198 tokens. The first quartile corresponds to sentences with a length of 6 tokens, the second quartile to 9 tokens, and the third quartile to 17 tokens.
This distribution is a positive asymmetric since the higher frequencies are below the mean. In other words, to express negation, the use of short sentences rather than long sentences is more frequent. For example, only the range between 1 and 9 tokens contains 50% of the negated sentences. In addition, after ten tokens, the frequency decreases rapidly, and sentences of greater length are rare.

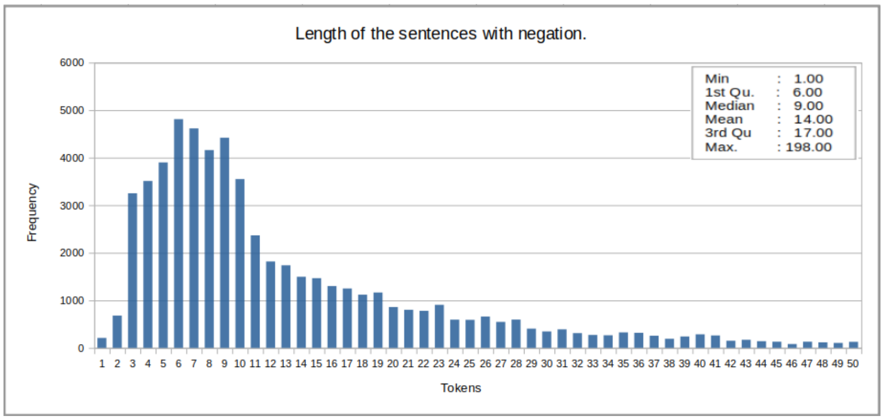

Fig. 1. Length of sentences with negation (Dataset 1)

– **Negation types:** The number of sentences that match each type of negation was obtained under the definitions shown in table 1. Negation types describe different ways to express negation in clinical texts. Negation terms are words that express negation (e.g., *"no, sin, nunca, negativo para"*). In section 4.2 there are examples of sentences related to these types.

Table 1. Different ways to express Negation

Negation Type	Description	Frequency
Contiguous Negation	When the same negation term is repeated several times consecutively on the sentence.	28%
Not contiguous negation	When the sentence contains one or more different negation terms.	43%
Short sentence negation	When the sentence contains a single negation term, and the number of tokens in the sentence is in the first quartile.	25%
Double negation	Sentences where occurs the double negation.	4%

– **Number of negation terms:** Table 2 shows the number of negation terms that appear in sentences in a contiguous way. The results are shown in 4 ranges, separated by the length of the sentences and the quartile (Q1, Q2, Q3) to which they belong. Table 3 shows the number of negation terms that appear in sentences in a not contiguous way. According to tables 2 and 3, the distribution of negation terms in contiguous negation is different, compared with not contiguous negation.

Table 2. Number of contiguous negation terms

Tokens Range	***Number of Negation terms***						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. [1 – Q1]		99%	1%				
2. (Q1 – Q2]		77%	23%				
3. (Q2 - Q3]		61%	32%	5%	4%		
4. >Q3		48%	27%	12%	8%	3%	2%

Table 3. Number of Not contiguous negation terms

Tokens Range	***Number of Negation terms***						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. [1 – Q1]	100%						
2. (Q1 – Q2]	95%	5%					
3. (Q2 - Q3]	93%	7%					
4. >Q3	91%	9%					

3.3 Results from dataset 2

Figure 2 shows the distribution of sentences where negation appears in *"IULA Spanish" dataset*. The length is distributed between 1 and 130 tokens. The first quartile corresponds to sentences with a length of 7 tokens, the second quartile to 11 tokens, and the third quartile to 20 tokens. Similar to results from dataset 1, this distribution is a positive asymmetric since the higher frequencies are below the mean. This data set also shows that to express negation, the use of short sentences rather than long sentences is more frequent.

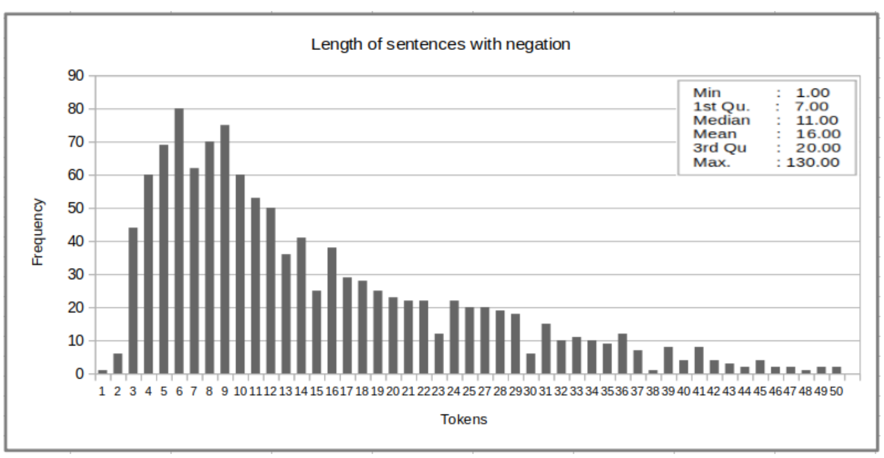

Fig. 2. Length of sentences with negation (Dataset 2)

Table 4 shows a summary of the indicators obtained from the exploratory data analysis for both analyzed datasets. There are considerable similarities between these datasets. For example, in both cases is more frequent the use of short sentences instead of long sentences for expressing negation. Moreover, in both cases, the negation is expressed in different ways, such as double negation, contiguous, and not contiguous negation.

Table 4. Summary of indicators from exploratory data analysis

Indicator	***Dataset 1***	***Dataset 2***
Negation Frequency	27%	32%
First quartile	6.00	7.00
Median	9.00	11.00
Mean	14.00	16.00
Third quartile	17.00	20.00
Contiguous Negation	28%	30%
Not contiguous negation	43%	40%
Short sentence negation	25%	27%
Double Negation	4%	3%

4 Proposal for negation detection in clinical texts

Spa-Neg proposal aims to improve negation detection in clinical texts written in Spanish. For this purpose, Spa-neg takes advantage of statistical information extracted from exploratory data analysis of negation. Spa-neg is divided into three components: Negation indicators, Negation detection, and Scope calculation.

4.1 Negation Indicators

This component contains indicators extracted from exploratory data analysis which are used to improve negation detection, as shown below:

– **Types of negation indicator** revealed that there are different ways of expressing negation in clinical texts. For this reason, our proposal includes rules that are better adapted to the way in which negation is expressed in Spanish. These rules are shown in section 4.2.

– **The number of negation terms indicator** showed that the way in which contiguous negation occurs is different from not contiguous negation. This indicator is used to calculate the negation scope. Spa-neg first checks if negation happens in a contiguous way, and depending on the outcome, calculates the negation scope. The scope calculation is explained in section 4.3.

– **The sentence length indicator** is used to estimate the value of the short sentence heuristic. This value is estimated based on the first quartile position in the length distribution of sentences. This value indicates that a sentence is considered as a *short sentence* if its length belongs to the first quartile, and it has only one negation term. Short sentence heuristic is also used to calculate the scope. An example of short sentence is: *"**No** se palpan masas"*.

4.2 Negation detection component

Spa-neg uses five regular expressions to detect negation, as described below: regular expressions 1 and 2 are adapted from the Negex proposal. Regular expressions 3 and 4 are an adaptation for Spanish based on [16] proposal. Moreover, we add the fifth regular expression to deal with cases where double negation is presented in clinical texts written in Spanish:

1. <pre-negation term>*<UMLS term>
2. <UMLS term>*<post-negation term>

The symbol * represents an unspecified number of tokens on the sentence. The UMLS (Unified Medical Language System) term, represents medical concepts affected by the pre-negation or post-negation term. Spa-neg contains a dictionary with the list of pre-negation and post-negation terms.

In the sentence *"**No** lesiones cutaneas relevantes"*, the negation is detected using the first regular expression because the word *"No"* is considered as a pre-negation term.

In the sentence *"Urocultivos de control* **negativos**.", the negation is detected using the second regular expression because the word *"Negativos"* is considered as a post-negation term. Regular expressions 1 and 2 best fit in cases of not contiguous negation.

3. <pre-contiguous negation term>* <UMLS term>
4. <UMLS term>* <post-contiguous negation term>

The regular expression 3 improves negation detection in sentences that contain multiple contiguous pre-negation terms. This case occurs when the same negation term is repeated several times consecutively, as it shows in the next example.

*"**No** dolor en el tórax, **no** tos **ni** expectoración, **no** dolor abdominal"*.

The regular expression 4 improves negation detection in sentences that contain several contiguous post-negation terms, as it shows in the next example.

*"Análisis de proteínas 0.2 g/l **negativo**, c.cetonicos **negativo**, bilirrubina **negativo**, urobilinogeno +- mg/dl, **negativo**, leucocitos pendiente realización."*

Finally, we add regular expression 5 to detect double negation in clinical text written in Spanish.

5. <pre-negation term>*<pre-negation term>* <UMLS term>

Double negation happens when two negation terms have detected that change the meaning of what is being negated. For example, the word *"descartar"* is often used to negate concepts within medical texts. However, this word can also be used to change the meaning of negation if it is combined with other terms that also activate the negation.

In the sentence: *"**No** se puede **descartar** tumor en el pulmón."*, there are two negation terms: *"No"* and *"descartar"*. In this case, these two terms should not be considered separately but should be analyzed together. Therefore, this sentence should not be detected as negated. Rule 5 allows for dealing with these cases.

4.3 Scope Calculation

When negation was detected with a pre-negation term, the scope is to the right to this term (forward in the sentence). If the negation is detected with a post-negation term, the scope is to the left to this term (backward in the sentence).

When negation is detected with a pre-negated term, the scope is calculated using the algorithm shown in figure 3. This algorithm receives two inputs: the sentence where negation was detected and metadata about negation detection (Neg-Data). This metadata contains the conditions and the way in which negation was detected. The algorithm requires that the sentence has been previously tokenized and POS (*Part of speech)* tagged.

The algorithm first checks whether negation was detected in a contiguous way. In this case, the negation scope will be given by the next negation term. The position of the next negation term indicates the end of the negation scope being analyzed at this time.

Next, the algorithm checks if negation was detected in a short sentence. If this is true, it is used the short sentence heuristic value. In this case, the negation scope is given by the end of the sentence. The short sentence heuristic improves negation detection because it avoids possible errors in searching for a termination term.

Next, the algorithm searches for a termination term, which indicates the end of the negation scope. Termination terms have been previously defined on the Spa-neg dictionary.

Finally, if none of the above conditions is true, the algorithm uses POS tagging properties to calculate the negation scope. In this case, the scope is determined by a token labeled with any of the following categories: conjunction, punctuation mark, or a verb. POS tagging is useful when negation is detected in a sentence that does not contain a termination term.

Algorithm 1: Scope Calculation

```
inputs: sentence, Neg-Data;
output: scope;

if (Neg-Data.is-contiguos = true) then
    scope=Neg-Data.next-position;
else
    if (sentence = shortSentence) then
        scope = sentence.end;
    else
        stopTerm = findTerminationTerm()
        if (isValid(stopTerm) = true) then
            scope = stopTerm.end;
        else
            token= findTokenUsingPOS();
            scope = token.end;
        end
    end
end
```

Fig. 3. Algorithm to calculate negation scope.

If negation is detected using a post-negated term, the scope is determined using an algorithm similar to the one showed in figure 3, but in this case, the scope is searched backward from the position where the post-negated term appears.

5 Tests and validation

To validate the proposed approach, we used the “IULA Spanish” corpus. It is proposed by [21] and was described previously.

Obtained results were measured using Precision, Recall and F-score in the task of negation detection and scope calculation.

$$\textbf{Precision} = \frac{\text{Negations and their scope correctly detected}}{\text{Number of detected negations}} \tag{1}$$

$$\textbf{Recall} = \frac{\text{Negations and their scope correctly detected}}{\text{Total negations in the corpus}} \tag{2}$$

$$\textbf{F score} = 2 * \frac{\text{Precision*Recall}}{\text{Precision + Recall}} \tag{3}$$

Table 5 shows the results in the task of negation detection and their scope. To validate our approach, we carried out four tests, as is explained below.

- ***Test 1:*** This test executed an adaptation of the Negex algorithm to Spanish using only regular expressions 1 and 2, as proposed in Negex.
- ***Test 2:*** In this execution, rules 3 and 4 were added to the Negex adaptation that was carried out in the previous test.

- ***Test 3:*** In this test, rule 5 is added to detect double negation cases.
- ***Test 4:*** Finally, in this test, in addition to using the five rules mentioned above, the use of POS Tagger labeling is added, as shown in Figure 3. This test includes all the ideas proposed in this paper.

Table 5. Negation detection and its scope results

	Test 1: Negex Adaptation Rules: 1 y 2	Test 2: Rules 1, 2, 3, 4	Test 3: Rules 1, 2, 3, 4, 5	Test 4: Rules: 1, 2, 3, 4, 5 + POS Tagger
Precision	0.76	0.86	0,90	0.91
Recall	0.82	0.90	0.93	0.95
F score	0.78	0.87	0.90	0.92

In test 1, an F-score value of 78% was obtained, which is the lowest of all tests. It is because, in this test, the algorithm only takes into account rules 1 and 2, as proposed in Negex. These rules are not sufficient to correctly detect cases of sentences that have multiple contiguous negation terms. The sentence:

*"Análisis de proteínas 0.2 g/l **negativo**, c.cetonicos **negativo**, bilirrubina **negativo**, urobilinogeno +- mg/dl, **negativo**, leucocitos pendiente realización."*,

is an example where test 1 fails to calculate the scope because there are multiple post-negation terms and there is no term that indicates scope termination for each instance of the detected negation.

In test 2, rules 3 and 4 were added to detect cases where contiguous negation terms appear, as in the previous example. By adding these rules, the results were significantly improved, and an 87% F-score was obtained. These rules are useful since in clinical texts written in Spanish, is very common the use of sentences that contain negation terms expressed contiguously.
In test 3, we add rule 5, to allow detecting cases where double negation occurs. By adding rule 5, an F-score of 90% was obtained in the tests performed. It indicates that, by correctly detecting the double negation, the results improved by 3% compared to the previous test.

The results obtained in test 4, include all the ideas proposed in this paper. In this case, 91% accuracy, 95% recall, and a 92% F-Score were obtained. These results are superior to those obtained in the previous tests and were achieved by a combination of elements such as the use of 5 regular expressions that improved negation detection and the algorithm shown in Figure 3 to calculate the scope. This algorithm takes advantage of the use of heuristics obtained from the statistical analysis of the negation.

The algorithm shown in Figure 3 also takes advantage of the syntactic properties of the sentence, such as POS Tagger labeling. POS tagging is very useful when negation is detected in a sentence which does not contain any term that indicates the end of the scope.

6 Conclusions and future work

In this paper, we proposed Spa-neg, an approach to negation detection in clinical texts written in Spanish. This approach used three elements to improve negation detection: an exploratory data analysis to understand how negation is expressed in clinical texts, regular expressions best adapted to the way in which negation is expressed in Spanish, and a shared test corpus focused on negation in clinical texts. This corpus was useful in the validation and testing process.

The use of additional regular expressions improved negation detection in the Spanish language because these regular expressions were better adapted to the way in which negation is written in clinical texts. It is important to mention that detecting negation in a contiguous way, and detecting double negation improved negation detection in cases where Negex adaptations to Spanish can present deficiencies.

To calculate the negation scope, our approach Spa-neg, took into account the following: the regular expressions with which negation was detected, the indicators obtained from an exploratory data analysis, termination terms that indicate the end of the scope and POS tagger properties of the sentence. POS tagger, in particular, is very useful when negation happens in a long sentence, which does not contain any words that indicate scope termination.

Negation detection in clinical texts is difficult to achieve since it can be expressed in different ways. For this reason, it is important to use an annotated and shared corpus for improving the validation and testing process. In our case, a shared corpus, "UILA Spanish" was used.

With regard to future work, we are planning to explore other approaches that can be applied in negation detection in medical texts. One example might be to include the use of semantic properties of sentences to improve results.

References

1. Aronson, A.R., Lang, F.M.: An overview of MetaMap: Historical perspective and recent advances. Journal of the American Medical Informatics Association **17**(3), 229–236 (2010). https://doi.org/10.1136/jamia.2009.002733
2. Attardi, G., Cozza, V., Sartiano, D.: Detecting the scope of negations in clinical notes. Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015 (5), 14–19 (2016)

3. Ballesteros, M., Francisco, V., Díaz, A., Herrera, J., Gervás, P.: Inferring the scope of negation in biomedical documents. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **7181 LNCS**(PART 1), 363–375 (2012)
4. Barigou, B.N., Barigou, F., Atmani, B.: Handling negation to improve information retrieval from french clinical reports. Journal of E-Learning and Knowledge Society **14**(1), 11–31 (2018). https://doi.org/10.20368/1971-8829/1455
5. Budrionis, A., Dalianis, H., Yigzaw, K.Y., Makhlysheva, A., Chomutare, T.: Negation detection in norwegian medical text : Porting a swedish negex to norwegian work in progress. Seventh Swedish Language Technology Conference (2018). https://doi.org/oai DiVA.org:su-162338
6. Chapman, W.W., Bridewell, W., Hanbury, P., Cooper, G.F., Buchanan, B.G.: A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics **34**(5), 301–310 (2001). https://doi.org/10.1006/jbin.2001.1029
7. Chapman, W.W., Hillert, D., Velupillai, S., Kvist, M., Skeppstedt, M., Chapman, B.E., Conway, M., Tharp, M., Mowery, D.L., Deleger, L.: Extending the NegEx lexicon for multiple languages. Studies in Health Technology and Informatics **192**(1-2), 677–681 (2013). https://doi.org/10.3233/978-1-61499-289-9-677
8. Costumero, R., Lopez, F., Gonzalo-Martín, C., Millan, M., Menasalvas, E.: An approach to detect negation on medical documents in Spanish. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **8609 LNAI**, 366–375 (2014)
9. Cotik, V., Roller, R., Xu, F., Uszkoreit, H., Budde, K., Schmidt, D.: Negation Detection in Clinical Reports Written in German. Proceedings of the 5th Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016) (BioTxtM), 115–124 (2016), http://www.aclweb.org/anthology/W16-5113
10. Cotik, V., Stricker, V., Vivaldi, J., Rodriguez, H.: Syntactic methods for negation detection in radiology reports in Spanish (December 2017), 156–165 (2016). https://doi.org/10.18653/v1/w16-2921
11. Cruz, N., Morante, R., Maña López, M.J., Mata Vázquez, J., Parra Calderón, C.L.: Annotating Negation in Spanish Clinical Texts pp. 53–58 (2017). https://doi.org/10.18653/v1/w17-1808
12. Cruz Díaz, N.P., Maña López, M.J.: Negation and Speculation Detection **13** (2019). https://doi.org/10.1075/nlp.13, http://www.jbe-platform.com/content/books/9789027262950
13. Cruz Díaz, N.P., Maña López, M.J., Vázquez, J.M., Álvarez, V.P.: A machine-learning approach to negation and speculation detection in clinical texts. Journal of the American Society for Information Science and Technology **63**(7), 1398–1410 (2012). https://doi.org/10.1002/asi.22679
14. De Albornoz, J.C., Plaza, L., Diaz, A., Ballesteros, M.: UCM-I: A rule-based syntactic approach for resolving the scope of negation. *SEM 2012 - 1st Joint Conference on Lexical and Computational Semantics **1**, 282–287 (2012)
15. Donatelli, L.: Cues, scope, and focus: Annotating negation in Spanish corpora. CEUR Workshop Proceedings **2174**, 29–34 (2018)
16. Elazhary, H.: NegMiner: An automated tool for mining negations from electronic narrative medical documents. International Journal of Intelligent Systems and Applications **9**(4), 14–22 (2017). https://doi.org/10.5815/ijisa.2017.04.02
17. Elazhary, H.: NegMiner: An automated tool for mining negations from electronic narrative medical documents. International Journal of Intelligent Systems and Applications **9**(4), 14–22 (2017). https://doi.org/10.5815/ijisa.2017.04.02

18. Github: Negex-mes: Negex para textos médicos en espanol. https://octoverse.github.com/ (Santamaria, J)
19. Jiménez-Zafra, S.M., Cruz Díaz, N.P., Morante, R., Martín-Valdivia, M.T.: Neges 2018: Workshop on negation in Spanish. Procesamiento de Lenguaje Natural **62**, 21–28 (2019). https://doi.org/10.26342/2019-62-2
20. Koza, W., Filippo, D., Cotik, V., Vanessa, S., Ricardo, M.G.: Automatic detection of negated findings in radiological reports for spanish language: Methodology based on lexicon-grammatical information processing. Journal of Digital Imaging **1**, 19–29 (2019)
21. Marimon, M., Vivaldi, J., Bel, N.: Annotation of negation in the IULA Spanish Clinical Record Corpus pp. 43–52 (2017). https://doi.org/10.18653/v1/w17-1807
22. Martí Antonín, M.A., Taulé Delor, M., Nofre, M., Marsó, L., Martín Valdivia, M.T., Jiménez Zafra, S.M.: La negación en español: análisis y tipología de patrones de negación. http://rua.ua.es/dspace/handle/10045/57750 (2016)
23. Mehrabi, S., Krishnan, A., Sohn, S., Roch, A.M., Schmidt, H., Kesterson, J., Beesley, C., Dexter, P., Max Schmidt, C., Liu, H., Palakal, M.: DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx. Journal of Biomedical Informatics **54**, 213–219 (2015). https://doi.org/10.1016/j.jbi.2015.02.010, http://dx.doi.org/10.1016/j.jbi.2015.02.010
24. Morante, R., Blanco, E.: Sem 2012 shared task: Resolving the scope and focus of negation. SEM 2012 - 1st Joint Conference on Lexical and Computational Semantics **1**, 265–274 (2012)
25. Névéol, A., Dalianis, H., Velupillai, S., Savova, G., Zweigenbaum, P.: Clinical Natural Language Processing in languages other than English: Opportunities and challenges. Journal of Biomedical Semantics **9**(1), 1–13 (2018). https://doi.org/10.1186/s13326-018-0179-8
26. Savova, G.K., Masanz, J.J., Ogren, P.V., Zheng, J., Sohn, S., Kipper-Schuler, K.C., Chute, C.G.: Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. Journal of the American Medical Informatics Association **17**(5), 507–513 (2010). https://doi.org/10.1136/jamia.2009.001560
27. Tanushi, H., Dalianis, H., Duneld, M., Kvist, M., Skeppstedt, M., Velupillai, S.: Negation Scope Delimitation in Clinical Text Using Three Approaches: NegEx , PyConTextNLP and SynNeg. Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa 2013) **1**(1), 387–397 (2013)
28. TIJCAI 2015: Negated Findings Detection in Radiology Reports in Spanish: an Adaptation of NegEx to Spanish (2015)
29. Velupillai, S., Suominen, H., Liakata, M., Roberts, A., Shah, A.D., Morley, K., Osborn, D., Hayes, J., Stewart, R., Downs, J., Chapman, W., Dutta, R.: Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. Journal of Biomedical Informatics **88**(May), 11–19 (2018). https://doi.org/10.1016/j.jbi.2018.10.005, https://doi.org/10.1016/j.jbi.2018.10.005
30. Vincze, V., Szarvas, G., Farkas, R., Móra, G., Csirik, J.: The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics **9**(SUPPL. 11), 38–45 (2008). https://doi.org/10.1186/1471-2105-9-S11-S9