Negation Detection in Clinical Reports Written in German

Viviana Cotik§, Roland Roller‡, Feiyu Xu‡, Hans Uszkoreit‡, Klemens Budde◇ and Danilo Schmidt◇
§Universidad de Buenos Aires, Buenos Aires, Argentina
§Universität des Saarlandes, Saarbrücken, Germany
§Language Technology Lab, DFKI, Berlin, Germany
vcotik@dc.uba.ar
‡Language Technology Lab, DFKI, Berlin, Germany
{firstname.surname}@dfki.de
◇Charité Universitätsmedizin, Berlin, Germany
{firstname.surname}@charite.de

Abstract

An important subtask in clinical text mining tries to identify whether a clinical finding is expressed as present, absent or unsure in a text. This work presents a system for detecting mentions of clinical findings that are negated or just speculated. The system has been applied to two different types of German clinical texts: clinical notes and discharge summaries. Our approach is built on top of NegEx, a well known algorithm for identifying non-factive mentions of medical findings. In this work, we adjust a previous adaptation of NegEx to German and evaluate the system on our data to detect negation and speculation. The results are compared to a baseline algorithm and are analyzed for both types of clinical documents. Our system achieves an F1-Score above 0.9 on both types of reports.

1 Introduction

Named entity recognition (NER) and relation extraction (RE) are central research topics in medical text mining. Clinical reports often contain a large number of expressions of negation and speculation. It is important to recognize whether extracted assertions (especially on medical conditions) describe these findings as factual, as contrafactual (absent) or as speculated (suspected). If, for instance, the report mentions *urolithiasis* (kidney stones) it surely matters, whether this medical condition has been diagnosed, rejected or merely suspected.

In comparison to many other text types, electronic health reports, radiology reports and other kinds of medical reports are often written in a rather telegraphic style. Furthermore they contain many technical terms as well as non-standard and ambiguous abbreviations (Kim et al., 2011). Many of those issues also appear in social media texts (Reitan et al., 2015). However, in the biomedical domain there are only very few annotated corpora available, due to data privacy issues. Therefore the curation or development of suitable data and tools for the clinical domain pose great challenges.

Various tools have been created for detecting negations and speculations in English medical reports. Probably the most popular one is NegEx (Chapman et al., 2001). The algorithm takes as input sentences with tagged *findings* and a list of negation and speculation terms called *triggers* and then determines whether the *finding* is within the scope of negation or speculation. In comparison to English, German clinical data differs in various characteristics which have to be taken into account for the successful application of an algorithm detecting non-factuality. First of all, German is a richly inflected language (e.g. *no* can be translated as *kein*, *keiner*, *keine* etc.). Furthermore, German includes *discontinuous triggers*, such as *kann ... ausgeschlossen werden ...*[1] (*can be ruled out*). Triggers may precede, but may also follow the negated expression, as presented in Table 1. Regarding this situation, Wiegand et al. (2010) state, that the detection of negation scope in German language is more difficult than in other languages, such as English.

[1]Dots indicate potential positions of the finding: (kann ... *finding*... ausgeschlossen werden, ... *finding*... kann ausgeschlossen werden)

Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016), pages 115–124, Osaka, Japan, December 12th 2016.

precede	**follow**
frei *von Beschwerden* (*free of symptoms*)	*beschwerde**frei*** (*without symptoms*)
nicht *klopfschmerzhaft* (*no percussion tenderness*)	*Hinweise für eine cerebrale* Metastasierung *gibt es derzeit* ***nicht***. (*There is no indication of a cerebral metastasis.*)

Table 1: Same negation triggers that might precede or follow a finding

Another interesting aspect of German negations are *surrounding triggers*, such as *lehnt ... ab (reject)* and *wies ... zurück (declined)*). In many cases it is possible to reduce/shorten triggers. However, in the case of given examples, a reduction would make the triggers too general, extending them to different meaning: *wies* (without *zurück*) for instance, could mean *to reject*, but also *to verify* in combination with the separated particle *nach*. Similar to English, negations can be directly bound to a target word as prefix or suffix, such as ***un****auffällig (****un****remarkable),* fett**frei** *(***non***fat) or motivations****los*** *(****without*** *motivation*).

In this paper we present an adaptation of NegEx (Chapman et al., 2001) for German clinical notes and discharge summaries. Our work is based on a previous version of NegEx triggers translated to German (Chapman et al., 2013). We conducted the following modifications: 1) we corrected and extended the trigger set, 2) we extended the regular expressions to possible expansions, and 3) we classified the triggers according to their position relative to the findings. Our work differs from Chapman et al. (2013) in that we evaluate NegEx on German clinical texts. The evaluation is carried out on two types of clinical data sets (clinical notes and discharge summaries) and it is compared to a baseline algorithm.For evaluation purposes we created a gold standard. Our system outperforms the baseline on both document types and achieves a F1-Score of over 0.9.

The remainder of the paper is organized as follows. Section 2 presents previous work in the detection of negation terms in the medical domain. Section 3 presents the main contributions, by explaining the methods and the data sets used, by providing an analysis of length and types of negation and speculation terms and by describing the generation of our gold standard. Section 4 presents the results of evaluating each of the algorithms with the test data set. After a discussion of the obtained results, the paper ends with conclusions and an outlook on future work.

2 Previous Work

Negation detection in the biomedical domain is a well-studied problem. Various workshops and challenges have addressed this problem in the last years, such as the *Workshop on Negation and Speculation in Natural Language Processing in 2010*,[2] *CoNLL 2010 Shared Task: Learning to Detect Hedges and Their Scope in Natural Language Text* (Farkas et al., 2010), the *2010 i2b2 NLP challenge*, that focused on the negation and uncertainty identification (Uzuner et al., 2011) and *SEM 2012 Shared Task: Resolving the Scope and Focus of Negation* (Morante and Blanco, 2012). Díaz published a book about Negation and Speculation detection in medical texts (2014). S. M. Meystre and Hurdle (2008) present a review of information extraction in biomedical texts, which also addresses negation detection.

A widely used tool for negation and speculation detection is Negex (Chapman et al., 2001). The method uses a simple algorithm based on regular expressions to detect *triggers* that indicate negation or speculation. Next it uses a window of words preceding or following each relevant term to determine if the term is under the scope of negation or speculation or not. NegEx has been extended to Context (Harkema et al., 2009) and adapted to Swedish, French, Spanish and other languages with good results (Skeppstedt, 2011; Deléger and Grouin, 2012; Cotik et al., 2016; Stricker et al., 2015; Costumero et al., 2014; Afzal et al., 2014). Beside NegEx and Context a wide range of other methods exist, e.g. based on syntactic techniques (Huang and Lowe, 2007; Mehrabi et al., 2015; Sohn et al., 2012; Cotik et al., 2016) or machine learning techniques (Uzuner et al., 2009). However, in clinical context simple methods, such as NegEx work very reliably for the task they have been designed for.

[2] `http://www.clips.ua.ac.be/NeSpNLP2010/program.html`

Other research has been dedicated to clinical negation detection together with the detection of pathological entities in German texts. Bretschneider et al. (2013) classify sentences containing pathological and non-pathological findings in German radiology reports. Their approach uses a syntacto-semantic parsing approach. Gros and Stede (2013) present Negtopus, a system that identifies negations and their scope in medical diagnoses written in German and in English.

Chapman et al. (2013) translate NegEx triggers into Swedish, French and German. The work reports, among others, the frequency of occurrence of German triggers in an annotated corpus of German medical text (Wermter and Hahn, 2004), that, as far as we know, is not available for public use. Both publications, (Gros and Stede, 2013) and (Chapman et al., 2013), are related to our work. However, Negtopus focuses currently only on negation terms. It has been evaluated on a set of only 12 cardiology reports for German negation detection. NegEx with the German trigger set has not been evaluated and thus its performance is still unknown to us.

3 Methods

The adaptation of NegEx to German requires having a set of triggers written in German. In order to evaluate the new system, a gold standard data set is necessary, consisting of medical text with tagged *findings* and a classification of those *findings* as negated, speculated or affirmed.

3.1 Baseline Algorithm description

The baseline algorithm uses a small list of negation and speculation terms obtained from a previous annotation task of another dataset. If one of those terms co-occur in the same sentence with a previously tagged *finding*, we assume the *finding* is negated or speculated. If not we assume it is affirmed.

3.2 NegEx Algorithm description

NegEx (Chapman et al., 2001) takes as input sentences, each of them with a previously tagged *finding*, and a list of triggers (*negation and speculation terms*), and as output it determines whether the finding is negated, speculated or affirmed. Each trigger has a label assigned, which determines the scope of the negation or speculation. PREN and POST labels correspond to negation terms that occur before and after the finding respectively. The same occurs with PREP and POSP, referring to speculation terms. CONJ refers to trigger terms that terminate the scope of a negation or speculation and PSEU to pseudo-negations.[3] For more information refer to Chapman et al. (2001).

The algorithm takes the following decisions: if a finding appears more than once in the sentence, and one of the occurrences is negated, the algorithm assumes that all occurrences are negated. If there are many occurrences of the same trigger in the trigger list (with different labels), the algorithm uses the label according to this precedence list: PREN, POST, PREP and POSP.

3.3 Triggers

The translated NegEx triggers of Chapman et al. (2013) are publicly available and our work is based on them. However, due to various reasons the original translation has been adapted by us. First, in some cases the authors suggest alternative formulations and regular expressions for a trigger. Those alternatives were added to the trigger list and regular expressions were transformed into strings (e.g. *kein.*$\{0,2\}$*signifikant.*$\{0,2\}$*(aenderung.*$\{0,2\}|$*Veraenderung.*$\{0,2\}$*)* to *keine signifikante aendeurng* | *keine signifikanten anderungen*, etc.) (*no significant changes*). Next, a small set of triggers have been exchanged by using an alternative translation. Moreover, new triggers which appeared to be useful were also added to the list. Classification with respect to speculation, proper negation and pseudo-negations and direction of scope was also revised for all triggers (i.e. the appropriate labels were assigned). A set of 506 triggers was obtained.[4] In addition to our trigger set, tests were also performed with the triggers translated by Chapman et al. (2013) without modification. The set contains 167 triggers. Alternative translations and regular expressions were not considered.

[3] If a *finding* is under the scope of a PSEU trigger, NegEx assumes it is affirmed.

[4] The link to the trigger data set will be made available here: `http://macss.dfki.de`.

3.4 Creation of a German Negation and Speculation Gold Standard

The data used for the following experiments consists of anonymized German discharge summaries and clinical notes of the nephrology domain. Both types of documents (discharge summaries and clinical notes) are written by medical doctors and have significant differences. The clinical notes are rather short and are written by doctors during or shortly after a visit of a patient. Discharge summaries instead are written during a stay at the hospital. The document is more structured. It contains information about medical history, diagnosis, condition, medication etc. of the patient. Discharge summaries contain much more text compared to clinical notes and often contain longer and more well-formed sentences.

Both types of documents exhibit non-standard abbreviations, that might include findings and negations among them (e.g. *oB -ohne Befund (without finding)-, opB -ohne pathologischer Befund (without pathological finding)-*). Texts have morphemes representing negation, speculation or findings and positioned as prefix, suffix or in the middle of a word. Examples are *un**, like in *unangenehm (uncomfortable)*, *unklar(e—er—es) (not clear)*, *unverändert(e) (unchanged)* and **los* or **losigkeit*, like in *Appetitslosigkeit (anorexia)* and *Schlaflosigkeit (insomnia)* (both represent findings), *problemlos(e) (without problems)* (that represents the absence of a finding). Table 2 provides an overview of the annotated data set used to test our experiment.[5]

	discharge summaries	clinical notes
# number of documents	8	175
total amount of words	6221	6674
total amount of sentences	1076	1158
avg. words per document (std. deviation)	777.63 (322.14)	38.14 (30.49)

Table 2: Comparison of annotated data sources

In order to be able to evaluate the results of our NegEx adaptation, a manually annotated gold standard was required. The annotation was carried out using the brat rapid annotation tool.[6] Moreover, in order to decrease the time dedicated to manual annotation, the data was automatically pre-annotated using an annotation tool (Roller et al., 2016).

Potential triggers were detected by using a small negation and speculation dictionary. Findings were pre-annotated using data of the UMLS[7] Methathesaurus. If a given string can be found in UMLS and its semantic type matches a set of predefined types (Anatomical Abnormality, Congenital Abnormality, Acquired Abnormality, Finding, Sign or Symptom, Pathologic Function, Disease or Syndrome, Mental or Behavioral Dysfunction, Neoplastic Process, Injury or Poisoning), then the string was annotated as a *finding* by the tool. After, the data was processed by a human annotator. Annotations wrongly made by the tool were removed or corrected and missing concepts were included. Furthermore, the annotator had to decide and annotate whether a given finding occurs in a positive, negative or rather speculative context. Finally, the annotations were corrected by a second -more-experienced- annotator to enhance the quality of the data.

Table 3 shows the number of findings that are affirmed, the number that are speculated and the number that are negated in the gold standard. The table shows, that both document types contain a large number of negations. It is interesting to note, that the ratio of affirmed and negated/speculated concepts is very different in both sets. While clinical notes contain approx. 25% more negations than affirmations, the data set contains hardly any speculations. On the other hand, the discharge summaries contain three times more affirmations than negations and speculations. However, the number of speculations is significantly higher compared to the clinical notes.

Table 4 and Table 5 present an analysis of the annotated negation and speculation terms for each document type. The tables depict the most frequent negation and speculation triggers in combination with

[5]The information was generated by applying a German tokenizer and a sentence splitter. All non alphabetical tokens were removed.

[6]`http://brat.nlplab.org/`

[7]`https://www.nlm.nih.gov/research/umls/`

type of finding	discharge summaries	clinical notes
affirmed	390	255
negated	106	337
speculated	22	4
findings (distinct)	518 (366)	596 (205)

Table 3: Number of affirmed, speculated and negated findings in the gold standard.

trigger order (i.e. the trigger comes before or after the finding) and its overall frequency. Furthermore, the tables present the mean word distance between trigger and finding, including standard deviation (std) and the overall information about how frequently a trigger occurs before (b) or after (a). Table 4, for instance, shows that *kein Nachweis (no evidence)* is used in 14.15% of the cases as negation trigger before the finding. Furthermore the table shows that the mean word distance between trigger and finding in the discharge summaries is 0.92 with a standard deviation of 1.42. In 97% of the cases the trigger occurs before the finding in the discharge summaries.

	discharge summaries	clinical notes
Trigger patterns *(translation, position, freq.)*	keine *(no, b, 35.85%)* kein *(no, b, 15.09%)* kein Nachweis *(no evidence, b, 14.15%)* ohne *(without, b, 9.43%)* kein Hinweis *(no indication, b, 5.66%)*	keine *(no, b, 64.47%)* kein *(no, b, 27.99%)* keine *(no, a, 3.46%)* kein *(no, a, 0.94%)* ohne *(without, b, 0.63%)*
mean distance (std)	0.92 (1.42)	0.40 (5.62)
position (b/a)	97% / 3%	94% / 6%

Table 4: Annotated negation terms

	discharge summaries	clinical notes
Trigger patterns *(translation, position, freq.)*	Verdacht *(suspicion, b, 30%)* fraglich *(doubtful, b, 10%)* am ehesten *(likely, b, 10%)* wahrscheinlich *(probable, b, 5%)* wahrscheinlich *(probable, a, 5%)*	? *(?, a, 100%)*
mean distance (std)	1.55 (1.64)	0 (0)
position (b/a)	80% / 20%	0% / 100%

Table 5: Annotated speculation terms

The tables show that the variation of triggers in the clinical notes is much smaller compared to the trigger variation in the discharge summaries. This can be explained by the telegraphic style of the clinical notes. In those reports, information is written very quickly, often while the patient is sitting next to the doctor. Due to time pressure and the internal use of the notes, verbose formulations are rare.

The analysis of the data and the development of the trigger set were performed in an independent way (annotated negation and speculation terms were not added as triggers).

4 Results

In this section we present the negation and speculation detection results of our NegEx adaptation (which we call *OTS* -our trigger set-) and the comparison against the original NegEx triggers provided by Chapman et al. (2013) (which we call *NTS* -NegEx trigger set-) and against our baseline. Results are presented in Table 6 and Table 7 and evaluated by using Accuracy, Precision, Recall and F1. In this case

True Positive (TP) refers to terms negated by the Gold Standard and correctly predicted by the methods. Furthermore, each table indicates the number of correctly and wrongly predicted instances.

dataset	discharge summaries			clinical notes		
algorithm	Baseline	NegEx		Baseline	NegEx	
triggerset	–	NTS	OTS	–	NTS	OTS
TP	103	65	99	333	123	328
FP	46	9	13	55	10	19
TN	366	403	399	204	249	240
FN	3	41	7	4	214	9
Accuracy	0.91	0.96	0.96	0.90	0.62	0.95
Precision	0.69	0.88	0.88	0.86	0.92	0.95
Recall	0.97	0.61	0.93	0.99	0.36	0.97
F1	0.81	0.72	0.91	0.92	0.52	0.96

Table 6: Performance on the negation detection task for both datasets with NegEx and with the baseline. TP refers to True Positive results, FP to False Positive, TN to True Negatives and to False Negatives. NTS refers to NegEx original triggers and OTS to our trigger set.

dataset	discharge summaries			clinical notes		
algorithm	Baseline	NegEx		Baseline	NegEx	
triggerset	–	NTS	OTS	–	NTS	OTS
TP	9	0	11	1	0	2
FP	14	0	7	5	5	8
TN	482	496	489	587	587	584
FN	13	22	11	3	4	2
Accuracy	0.95	0.96	0.97	0.99	0.98	0.98
Precision	0.39	0	0.61	0.17	0	0.2
Recall	0.41	0	0.5	0.25	0	0.5
F1	0.4	0	0.55	0.2	0	0.29

Table 7: Performance on the speculation detection task for both datasets with NegEx and with the baseline.

Table 8 shows the negation and speculation triggers that appear more than four times, taking into account discharge summaries and clinical notes.

5 Discussion

The results show, that the baseline algorithm provides promising results for the negation detection task. This might have to do with the fact that in German many of the triggers can be used before or after the finding (see Table 1). However, the results show, that in all cases the NegEx adaptation achieves better results compared to the baseline algorithm. In particular the negation and speculation detection applied to the discharge summaries leads to much better results than using the baseline algorithm. This can be explained by the fact that the discharge summaries include a larger variety of triggers, which are not covered by the baseline, but covered by the German trigger set. Moreover, discharge summaries have longer and more complex sentences, that include *CONJ* triggers, which end the scope of negation. However, the results show, that both algorithms achieve better results using the clinical notes. We believe the reason is related to the fact that clinical notes have much shorter and simpler sentences than the ones of discharge summaries. The test with the original German trigger set achieves lower results than our NegEx adaptation and our baseline. The results improve and are similar to ours (F1=0.92 for discharge summaries and 0.94 for clinical notes) if the trigger *keine* is added to NTS.

trigger type	trigger	translation	number of occurences
negation	keine, kein	no	471, 226
	ohne	without	49
	nicht	not	50
	noch	still/yet	40
	aber	but	18
	jedoch	but/however	15
	bis auf	except for	11
	entfernt	removed	7
speculation	verdacht	suspicion	13
	ehesten, eher	rather	13,8
	nicht sicher	not sure	5
	?	?	14

Table 8: Negation and speculation triggers used more than four times. Both kind of reports are taken into account.

Considering the 506 triggers of our data, only 27 occur in the clinical reports (see the ones used more than four times in Table 8). This makes us infer that the translation effort could be avoided in further adaptation of NegEx to other languages. Other works arrived to similar conclusions (Cotik et al., 2016).

Reviewing the errors, we found that syntactic analysis could improve our results. For instance, in *kein starker Krampf (no strong cramp)*, *Krampf* is under the scope of *kein (no)*, a *PREN* trigger, but *no* is actually addressing to *strong* and not to *cramp*. The use of Part of Speech tagging or dependency parsing information could help us avoid this error. Moreover, the original NegEx speculation triggers did not help us to find speculation. In fact with those triggers no speculation terms have been detected (see Table 7). Thus, a number of speculation triggers have been added to OTS. Triggers were taken from general German knowledge and from the transformation of some of the original negation triggers to their corresponding speculation triggers (e.g. *Ohne Verdacht* -without suspicion- originated *Verdacht -suspicion-*). In particular, we added the trigger *?* as a speculation term occurring after the finding, since we knew it is frequently used in the clinical notes to express uncertainty. Some False Negative results were generated by the abundance of acronyms, some of them indicating negation of findings (e.g. in oB *-ohne Befund, without finding-*, B *-Befund, finding-* was annotated as negated, but we don't have *o-ohne, whithout-* as a trigger). In all cases negation detection achieves better results than speculation detection. This might be due to the fact that there is much greater variety of triggers for indicating speculation than triggers for indicating negation. Additionally, we detected some missing triggers. In some cases two classifications of the triggers (e.g *nicht*) were possible (see Table 1). For those triggers we missed some correct classifications, where the trigger appeared in the less frequent order (for example *Lymphozele nicht mehr sichtbar , Lymphocele not visible anymore)* was classified as positive, since *nicht* was in the trigger list as a PREN trigger. See also trigger preference list in Section 3.2.

Parenthesis and commas were not included as CONJ triggers in our trigger set. After evaluating FP and FN results (see Tables 6 and 7) tests were performed including them. Including parenthesis and commas as triggers reduces the number of false positives. Consider for example those cases that use the trigger *nicht*: *Hat Nitrendipin nicht vertragen (**Flush**) (Did not tolerate Nitrendipin (flush)). Befinden seit Entlassung nicht gebessert, hat weiterhin **Diarrhoe*** (Condition has not been improved since discharge, has still diarrhoea). In the previous examples the findings *Flush* and *Diarrhoe* are out of the scope of negation and therefore misclassified. We also could avoid false negatives in speculation detection in cases such as *keine Oedeme (...) (serom?) (no edema (...) (serum?))*, because with our trigger set *serum?* is under the scope of *kein*. In a subsequent test, we included parenthesis and commas as CONJ triggers, which increased F1 of clinical notes to 0.98 and F1 of discharge summaries to 0.94 for negations and F1 of clinical notes to 0.62 (with a recall of 1) and F1 of discharge summaries to 0.58 for speculation.

As explained above, clinical notes are much shorter than discharge summaries. The language is less verbose, often just consisting of sequences of noun phrases with some embedded prepositional phrases. Discharge summaries in contrast contain more verbs and full sentences. Thus it is not surprising when our analysis of triggers shows that the term *kein(e) -no-* as a negative determiner is much more often used in clinical notes (571 vs. 128) whereas the sentence negation nicht (not) occurs more often in discharge summaries (32 vs 18).

Our NegEx adaptation for negations yields very good results. Although not easily comparable (because of being applied to different languages and types of medical reports), they are better than the ones obtained by the original algorithm for English clinical texts and to the adaptations done to Swedish and Spanish (in this last case only for clinical notes, discharge summaries results are similar to results obtained for Spanish). They also outperform results obtained on 12 German cardiology reports by Gros and Stede (2013). We believe that the fact of having short sentences with simple syntactic structures helps us to get good results. It should also be considered that our data set is highly redundant (some negations or negation types occur frequently). In order to improve results an hybrid method combining syntactic analysis could be used.

6 Conclusions

This paper presented negation and speculation detection of medical findings reported in German clinical data. Two approaches were introduced: A dictionary look-up algorithm, that was taken as a baseline and an approach based on a revised version of an existing German NegEx trigger set. Tests were also performed with triggers that were previously translated to German. The system has been tested on two different data sets, German discharge summaries and German clinical notes. In both cases the German NegEx system outperforms the baseline and achieves an F1-Score above 0.9. Furthermore this work presented an analysis of negations and speculations existing in both document types. The analysis shows, that physicians tend to use a structurally simple and precise language. Therefore the degree of lexical variation in expressing negation is very low. However, applying NegEx to other text types might turn out to be more challenging.

As Chapman et al. (2013) state, the translation of triggers to another languages has faced a number of issues. German is a language with agglutinative features, where a morpheme representing negation can be added to a word. NegEx does not address this fact. German is an inflected language, so a single term can be translated to many others, because of gender and number agreement. This increased the size of our trigger set.

One of the challenges of working with medical language is the need for careful anonymization. Texts also exhibit large numbers of technical terms and non-standardized and ambiguous abbreviations. All of this raises the efforts needed for corpus curation and annotation raising the demand for gold-standard data that can be shared.

7 Future Work

We plan to detect negation that is represented by bound morphemes (prefix or suffix) of relevant content words. If a lexeme *lf* stands for a medical finding according to the UMLS thesaurus, *lf+"los"* (*without*) should be considered as a negation of the finding, e.g., *schlaflos* (without sleeping), but also *lf+"los"* or *lf+"losigkeit"* could be included in the thesaurus (e.g. *Appetitslosigkeit (anorexia)* and *Schlaflosigkeit (insomnia)*), and in this case the presence of suffix or infix *los* does note indicate the absence of a finding.

We also intend to investigate the benefits of employing syntactic analyses to improve the results. Especially for the clinical notes, chunk parsing technology will have to be adapted in order to cope with the nature of this text sort.

Acknowledgements

This research was partially supported by the German Federal Ministry of Economics and Energy (BMWi) through the project MACSS (01MD16011F) and by the Saarland University through the B5 - SFB 1102 'Information Density and Linguistic Encoding' project.

References

Zubair Afzal, Ewoud Pons, Ning Kang, Miriam Sturkenboom, Martijn Schuemie, and Jan Kors. 2014. ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. *BMC Bioinformatics*, 15(373):1–12.

Claudia Bretschneider, Sonja Zillner, and Matthias Hammon. 2013. Identifying pathological findings in German radiology reports using a syntacto-semantic parsing approach. In *Proc of the 2013 Workshop on Biomedical Natural Language Processing (BioNLP 2013)*, pages 27–35.

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

Wendy W. Chapman, Dieter Hillert, Sumithra Velupillai, Maria Kvist, Maria Skeppstedt, Brian E. Chapman, Mike Conway, Melissa Tharp, Danielle L. Mowery, and Louise Deléger. 2013. Extending the NegEx Lexicon for Multiple Languages. In *Proceedings of the 14th World Congress on Medical and Health Informatics*, pages 677–681, Copenhagen, Denmark.

Roberto Costumero, Federico López, Consuelo Gonzalo-Martín, Marta Millan, and Ernestina Menasalvas. 2014. An Approach to Detect Negation on Medical Documents in Spanish. In *Brain Informatics and Health*, volume 8609, pages 366–375.

Viviana Cotik, Vanesa Stricker, Jorge Vivaldi, and Horacio Rodriguez. 2016. Syntactic methods for negation detection in radiology reports in Spanish. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing (BIONLP 16)*, Berlin, Germany. Association for Computational Linguistics.

Louise Deléger and Cyril Grouin. 2012. Detecting negation of medical problems in French clinical notes. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 697–702.

Noa P C Diaz. 2014. Negation and speculation detection in medical and review texts. In *SPLN*, volume 13.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and Their Scope in Natural Language Text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 1–12, Uppsala, Sweden.

Oliver Gros and Manfred Stede. 2013. Determining Negation Scope in German and English Medical Diagnoses. In *Nonveridicality and Evaluation*, pages 113–126.

Henk Harkema, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. 2009. ConText: An Algorithm for Determining Negation, Experiencer, and Temporal Status from Clinical Reports. *Journal of biomedical informatics*, 42(5):839–851.

Yang Huang and Henry J Lowe. 2007. A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports. *Journal of the American Medical Informatics Association*, 14(3):304–311.

Youngjun Kim, John Hurdle, and Stphane M Meystre. 2011. Using UMLS lexical resources to disambiguate abbreviations in clinical text. *Annual Symposium proceedings*, 2011:715–722.

Saeed Mehrabi, Anand Krishnan, Sunghwan Sohn, Alexandra M Roch, Heidi Schmidt, Joe Kesterson, Chris Beesley, Paul Dexter, C Max Schmidt, Hongfang Liu, et al. 2015. DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx. *Journal of biomedical informatics*, 54:213–219.

Roser Morante and Eduardo Blanco. 2012. Sem 2012 shared task: Resolving the scope and focus of negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, SemEval '12, pages 265–274, Stroudsburg, PA, USA. Association for Computational Linguistics.

Johan Reitan, Jørgen Faret, Björn Gambäck, and Lars Bungum. 2015. Negation scope detection for twitter sentiment analysis. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 99–108, Lisboa, Portugal. Association for Computational Linguistics.

Roland Roller, Hans Uszkoreit, Feiyu Xu, Laura Seiffe, Michael Mikhailov, Oliver Staek, Klemens Budde, Fabian Halleck, and Danilo Schmidt. 2016. A fine-grained corpus annotation schema of german nephrology records. In *Proceedings of the Clinical Natural Language Processing Workshop*, Osaka, Japan, December. Association for Computational Linguistics.

K. C. Kipper-Schuler S. M. Meystre, G. K. Savova and J. F. Hurdle. 2008. Extracting information from textual documents in the electronic health record: A review of recent research.

Maria Skeppstedt. 2011. Negation Detection in Swedish Clinical Text: An Adaption of NegEx to Swedish. *Journal of Biomedical Semantics*, 2(3):1–12.

Sunghwan Sohn, Stephen Wu, and Christopher G Chute. 2012. Dependency parser-based negation detection in clinical narratives. *AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science*, 2012:1–8.

Vanesa Stricker, Ignacio Iacobacci, and Viviana Cotik. 2015. Negated Findings Detection in Radiology Reports in Spanish: an Adaptation of NegEx to Spanish. In *IJCAI - Workshop on Replicability and Reproducibility in Natural Language Processing: adaptative methods, resources and software*, Buenos Aires, Argentina.

Özlem Uzuner, Xiaoran Zhang, and Tawanda Sibanda. 2009. Machine Learning and Rule-based Approaches to Assertion Classification. *Journal of the American Medical Informatics Association : JAMIA*, 16(1):109–115.

Özlem Uzuner, Brett R. South, Shuying Shen, and Scott L. DuVall. 2011. 2010 i2b2/VA Challenge on Concepts, Assertions, and Relations in Clinical Text. *Journal of the American Medical Informatics Association : JAMIA*, 18(5):552–556.

Joachim Wermter and Udo Hahn. 2004. An Annotated German-Language Medical Text Corpus as Language Resource. In *Proc 4th Intl LREC Conf*, pages 473–476.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, NeSp-NLP'10, pages 60–68, Stroudsburg, PA, USA. Association for Computational Linguistics.