Negation detection in Norwegian medical text: Porting a Swedish NegEx to Norwegian
Work in progress

Andrius Budrionis∗, Hercules Dalianis†, Kassaye Yitbarek Yigzaw∗, Alexandra Makhlysheva∗, Taridzo Chomutare∗

∗Norwegian Centre for E-health Research, University Hospital of North Norway
Tromsø, Norway
†DSV/Stockholm University
Kista, Sweden
Andrius.Budrionis@ehealthresearch.no, Hercules.Dalianis@ehealthresearch.no,
Kassaye.Yitbarek.Yigzaw@ehealthresearch.no,
Alexandra.Makhlysheva@ehealthresearch.no,
Taridzo.Chomutare@ehealthresearch.no

Abstract
This paper presents an initial effort in developing a negation detection algorithm for Norwegian clinical text. An evaluated version of NegEx for Swedish was extended to support Norwegian clinical text, by translating the negation triggers and adding more negation rules as well as using a pre-processed Norwegian ICD-10 diagnosis code list to detect symptoms and diagnoses. Due to limited access to the Norwegian clinical text the Norwegian NegEx was tested on Norwegian medical scientific text. NegEx found 70 negated symptoms/diagnoses in the text combined of 170 publications in the medical domain. The results are not completely evaluated due to the lacking gold standard. Some challenging erroneous tokenizations of Norwegian words were found in addition to the need for improved preprocessing and matching techniques for the Norwegian ICD-10 code list. This work pointed out the weaknesses of the current implementation and provided insights for future work.

1. Introduction
Free-text is an integral part of clinical documentation. Regardless of the maturity of the electronic health record (EHR) systems and initiatives to structure free-text information, it is still highly prevalent and widely accepted by clinicians. The nature of the free-text fits well with the level of uncertainty in diagnosis-setting process and allows clinicians to document patient history in a way, which is difficult to put into a rigid structure.

While free-text is a recognised and often preferred way of clinical documentation, it presents major challenges for data reuse. Better use of health data accumulated in EHRs is a important aim for future health care worldwide and Norway is no exception (Direktoratet for e-helse, 2017). Retrieval of knowledge from free-text documentation is a complex task. Individualistic patterns for describing patient status, various dialects and spelling errors are well-known challenges to be addressed.

Negated symptoms and diagnoses are part of diagnosis-setting process and are common in medical text. Such concepts have to be identified and handled accordingly in the knowledge retrieval process ensuring that they are not mixed with the positive symptoms and diagnoses, (Groopman, 2007).

Negation detectors (NegEx) are available in several languages, however, to our knowledge, Norwegian NegEx has not yet been developed. A Swedish version of NegEx has been presented earlier (Skeppstedt, 2011) and yielded satisfactory precision and recall when tested on Swedish medical text. This paper presents the development and initial tests of the Norwegian NegEx using earlier presented Swedish work as a starting point.

2. Related research
The existing NegEx algorithms and dictionaries are language-specific and translation is not always sufficient to adopt them in a new language. The most of effort in the field was put into developing negation detection algorithms for English (Chapman et al., 2001; Mehrabi et al., 2015; Peng et al., 2018; Ou and Patrick, 2015). However, tools for processing other languages, such as Swedish (Skeppstedt, 2011), French, German (Chapman et al., 2013), and Chinese (Kang et al., 2017) exist.

The English version of NegEx obtained a precision of 84.5% and a recall of 82.4% when tested on discharge summaries (Chapman et al., 2001). The Swedish version was ported from the English original version of NegEx and obtained a precision of 87.9% and a recall of 91.7% for negation cues on Swedish clinical text. (Skeppstedt, 2011). Tanushi et al. compared various approaches to negation detection (Swedish NegEx, PyConTextNLP and SynNeg). PyConTextNLP is an extension of the English NegEx and SynNeg is based on a syntactic parser that considers sentence boundaries. All three systems produced similar results; SynNeg performed better on long and complex sentences (Tanushi et al., 2013).

Attempts to improve the precision and recall of the original NegEx (English) are described in the literature. For in-
stance, Mehrabi et al. included analysis of dependencies between negation terms and other concepts in the sentence aiming to decrease the number of false positives in the original NegEx. The attempt was only partially successful and presented higher precision and recall scores only in one out of three selected corpora (Mehrabi et al., 2015). The findings highlight interoperability concerns, which may be caused by varying practices in documenting patient condition, spelling errors and potentially incomplete dictionary of negation terms. A recent work by Peng et al. presented significant improvement to the original NegEx performance (on average 9.5% higher precision and 5.1% higher F1-score) when tested on two corpora containing radiology reports (Peng et al., 2018). The improvement was achieved utilising patterns on universal dependencies that help to identify the scope of negation triggers.

3. Methods and Data

The Swedish version of NegEx (Skeppstedt, 2011) was ported to Norwegian and evaluated on Norwegian medical scientific text in the domain of gastrointestinal surgery, while we are waiting for access to clinical text from the EHR. The reason to choose gastrointestinal surgery is our future focus on the analysis of free-text notes documenting this type of surgery in patient records in the University Hospital of North Norway.

The Norwegian medical scientific text was downloaded from the Tidsskrift Den norske legeforening and transformed from portable document format (PDF) to pure UTF-8 coded text. Specifically scientific publications in the field of gastrointestinal surgery were chosen, in total 170 articles containing 294,745 words.

3.1 Porting the Norwegian NegEx from Swedish

Swedish and Norwegian are closely related languages almost completely comprehensible for speakers from both language groups. They have similar grammar, however spelling is rather different. Norwegian bokmål is a preferred written standard for about 90% of the population and its spelling is derived from Danish language.

Since Norwegian and Swedish grammars are similar, for creating the Norwegian version of NegEx, Swedish version of NegEx was taken as a basis. The grammatical differences between Swedish and English are described by Skeppstedt (Skeppstedt, 2011). The list of Swedish negations triggers was translated to Norwegian. Some expressions were added in form of new negation rules in cases when a phrase could be translated in several ways, spelled differently or have both a direct and a reversed word order, and verbs in passive and active voices. Moreover the pseudo negation "ikke minst" (Eng. "at least") was added in a form of a new pseudo negation rule. Further, the English version of NegEx was also taken into account when exact match between Norwegian and Swedish was not available.

Starting with the Swedish version of NegEx with 40 negation rules, additional 26 negation rules were added to a total of 67 negation rules in the Norwegian version of the NegEx. The distribution was as follows: 15 Norwegian POST-negation rules (for instance triggers such as, negativt, manget, bekreftet (Eng. "negatively", "missing", "de-"), 34 PREN-negation rules (for instance triggers such as, aldri, ikke, nekte (Eng. "never", "not", "deny") and 18 PSEUDO-negation rules (for instance triggers such as, ikke utelukker, ikke forårsket, ikke bekreftet (Eng. "not excluded", "not caused", "not confirmed"). POST-negation rules are used to find negated term after the negation trigger and PREN-negation rules are used to find negated term before the negation trigger.

NegEx requires a list of symptoms and diagnoses to be matched to the medical text for negation. A Norwegian version of ICD-10 from year 2017 was used for this purpose. The Norwegian ICD-10 list contains 19,597 codes and their descriptions in free-text. The ICD-10 list was pre-processed by removing stop-words using the Norwegian Snowball stop-word list and manually removing long modifiers adhoc, for example Annen spesifisert... (Eng: Other specified). Finally, single-word symptoms and diagnoses, enabling matching between matching list and the analysed text, were also extracted manually and gave a final list of 19,628 terms.

In addition to the ICD-10 list, 23 significant words from the gastrointestinal surgery domain (Table V in (Soguero-Ruiz et al., 2016)) were added into the domain list. NegEx pre-processed automatically both lists and compiled them into total 19,651 terms.

An example of the Norwegian NegEx correctly functioning on a medical text is presented in the Figure 1.

Han var ved innkomst hemodynamisk upåvirket og hadde ikke tegn til [NEGATED]peritonitt[NEGATED]. (Eng: "At arrival he was not affected hemodynamically and had no signs of [NEGATED]peritonitis[NEGATED]").

Table 1. Results of applying NegEx on a corpus containing text from 170 scientific medical publications are presented in Table 1. 

<table>
<thead>
<tr>
<th>Negation types</th>
<th>ICD-10</th>
<th>Standard</th>
<th>Helsefaglige</th>
<th>Kodeverket-ICD-10</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST-negation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PREN-negation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSEUDO-negation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: An example of machine based proper negation detection.

Results of applying NegEx on a corpus containing text from 170 scientific medical publications are presented in Table 1. 


4. Results

When executing NegEx on the 170 articles, the system found 70 negated symptoms/diagnoses. NegEx was also executed on a smaller sub corpus (1/4 of the original corpus) that gave in total 15 automatic negation labels, manually labelling gave 29 negations. No precision and recall measurements has been calculated since the final labelling is not ready yet.

5. Lessons learned

Lack of labelled corpora containing clinical Norwegian text, which could be used for the evaluation of the algorithm to validate the NegEx, was the first challenge to be addressed. Therefore, medical literature published in a scientific medical journal was used as a starting point. Labelling of this corpus is currently in progress, and needs to be carried out during several labelling rounds jointly with clinically trained personnel. Labelled corpus will allow us to evaluate the performance of the NegEx in terms of widely accepted precision and recall measures.

The automatically identified negations were manually analysed by the authors. This process resulted in additions to the dictionary of negation triggers and extra negation rules in the NegEx algorithm. Parts of the corpus were manually inspected, identifying negations, which were not captured by the NegEx. But also erroneous tokenization of the text such as ikke-operable metastaser (Eng. "inoperable") must be tokenized as ikke-operable [NEGATED]metastaser[NEGATED] (Eng. "in-operable [NEGATED]metastases") became machine labelled as ikke-operable metastaser. The automatically identified negations were manually inspected, identifying negations, which were not captured by the NegEx. But also erroneous tokenization of the text such as ikke-operable metastaser (Eng. "inoperable") must be tokenized as ikke-operable [NEGATED]metastaser[NEGATED] (Eng. "in-operable [NEGATED]metastases") meaning that ikke-operable (Eng. "inoperable") does not become a negation trigger to metastaser.

Such error analysis highlighted the flaws of the current algorithm, which are to be addressed in the next version of the Norwegian NegEx. The current algorithm uses exact string-matching strategy, looking for an exact match between the text in the corpus and the ICD-10 code list of symptoms and diagnoses.

Currently we are also experimenting with approximate string-matching techniques delivering a more robust mechanism to capture the negated symptoms/diagnoses. The clinical findings mentioned in the text, in many cases, are incomplete or phrased differently in comparison to the ones in the ICD-10 list. Therefore, many of them were not captured by the NegEx, hence one method is to pre-process the ICD-10 code list using bi- or tri-grams, similar to the approach by [Skeppstedt, 2011, Skeppstedt, 2011].

Table 1: Results of Norwegian NegEx applied on Norwegian medical scientific text in gastrointestinal surgery.

<table>
<thead>
<tr>
<th>Description</th>
<th>Words</th>
<th>Symptoms/diagnoses</th>
<th>Negations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results on whole corpus</td>
<td>294,745</td>
<td>1,835</td>
<td>70</td>
</tr>
<tr>
<td>Manually labelled sub corpus</td>
<td>75,614</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>Automatically labelled sub corpus</td>
<td>75,614</td>
<td>526</td>
<td>15</td>
</tr>
</tbody>
</table>

Future work will hence address both the aforementioned weaknesses of the algorithm as well as labelling of the electronic patient records from gastrointestinal surgery.

Acknowledgements

Great thanks to Maria Skeppstedt for good advices regarding the programming of the Norwegian NegEx algorithm. This work was funded by the Helse Nord grant (HNF1395-18) to the Norwegian Centre for E-health Research.

References


