Laboratory Project of ANLP-MAI for the course 2015-2016.

The Laboratory Project of ANLP-MAI for this course consists of building a system for detecting negation and its scope on medical discharges documents (EHR).

Detecting negations and their scopes in NL documents is an important issue for many NLP applications. Although detecting triggers ('no', 'not', 'neither', 'negative', etc.) indicating a potential negation is not difficult, accepting the case as a true negation and determining the scope of the negation is a challenging task. Consider, for instance, the following sentence: "The chest X-ray showed no infiltrates and EKG revealed sinus tachycardia." The negation phrase "no" applies to "infiltrates" but not to "sinus tachycardia". In many texts this issue is important but in EHR the correct detection of negated events (and their exact scope) is crucial.

Many negation detection systems exist, but most of them proceed only at lexical level while it seems at a glance that in some cases syntactic and even semantic analysis is needed for an accurate analysis of scope. Perhaps the most widely used system, providing a nice accuracy score, is NEGEX, Chapman et al, 2001.

What we propose is using NEGEX as a hard baseline and try to improve it by using syntactic and/or semantic analysis of the documents.

We can find attached:

1. A copy of the paper Chapman et al, 2001 describing NEGEX (10.1.1.11.4439.pdf)
2. A python implementation of NEGEX provided by Peter Kang and consisting of two python programs:
    a. negex.py
    b. wrapper.py
3. A description of the implementation, README.txt
4. A set of triggers used by NEGEX
5. A dataset of EHR with negation annotations, Annotations-1-120.txt, that can be used for learning and test

What you are asked to do is:

1. Carefully analyzing NEGEX performance focusing on its limitations
2. Propose a way of improvement of NEGEX accuracy by using syntactic and/semantic information
3. Implement your proposal (you are free to use external parsers as Freeling, Stanford, OPEN NLP, SENNA, etc…). NLTK can provide you useful material.
4. Compare your results with NEGEX
5. The laboratory case should be carried out by teams of two students. Individual works or 3 members teams should be approved by the tutor.

Wendy W. Chapman , Will Bridewell , Paul Hanbury , Gregory F. Cooper , Bruce G. Buchanan
A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries (2001)

Barcelona, 9th February 2016