Some Aspects of Negation Processing in Electronic Health Records

Svetla Boytcheva[1], Albena Strupchanska[2],
Elena Paskaleva[2] and Dimitar Tcharaktchiev[3]

[1]Department of Information Technologies, Faculty of Mathematics & Informatics,
Sofia University "St. Kl. Ohridski", 5 James Bauchier Str., 1164 Sofia
svetla@fmi.uni-sofia.bg
[2]Linguistic Modelling Department, Central Laboratory for Parallel Processing,
Bulgarian Academy of Sciences, 25A Acad. G. Bonchev Str., 1113 Sofia
{albena, hellen}@lml.bas.bg
[3]University Hospital of Endocrinology, Nephrology and Gerontology,
Medical University, 6 Dame Gruev Str., Sofia
dimitar@medicalnet-bg.org

Abstract

The presented paper discusses a hybrid approach for negation processing in Electronic Health Records (EHRs) in Bulgarian. The rich temporal structure and the specific combination of medical terminology in both Bulgarian and Latin do not allow the application of standard language processing techniques. The problem gets even worse due to the often use of specific abbreviations, analyses and clinical test data. Various expressions of negation often occur in EHRs. This raises many difficulties for language processing especially in semantic analysis. That is why we propose an approach that combines information extraction with deep semantic analysis, allows to determine the negation, negation scope and to treat it appropriately. We present a prototype of a system MEHR for automatic recognition of medical terms and some facts in EHRs in Bulgarian. This is the first step towards filling in a template concerning the patient status. Automatic extraction of facts needed for description of patient status in full is our ultimate goal. However in this paper we just focus on proper treatment of negation.

1. Introduction

Automatic generation of Patient's Chronicle – symptoms and diagnosis from EHRs is very challenging and ambitious task. It requires recognition of medical terminology, deep semantic analysis in certain domain relevant and important points, processing of temporal structure and discourse analysis. Even partial solution for generation of patient's chronicle requires intensive linguistic and domain knowledge as well as application of different language processing techniques. For languages other than English such knowledge resources are missing. So in the attempt to process EHRs in Bulgarian we have to rely on usage of limited resources and shallow processing techniques.

Because of the frequent usage of negated observations in important parts of EHRs, a crucial element in correct recognition of patient's symptoms and diagnosis is determining whether a particular symptom/diagnosis is present or absent in the patient. In this paper we will mainly focus on negation – its usage, classification and proper treatment.

The paper is organized as follows: The first section contains a brief introduction. Section 2 presents overview of some related work. Section 3 describes system architecture and main modules. Section 4 focuses on the problems in negation treatment and describes the most frequently used types of negations in EHRs. Section 5 presents the main steps of the system's work and illustrates them by an example. Section 6 contains a brief discussion about problems and unsolved cases. Section 7 explains further work and conclusion.

2. Related work

Several language processing systems, which extract and codify information from electronic health records in English, have been developed. A good overview, comparison and evaluation of such systems can be found in [1, 2]. We will briefly mention the main language techniques and resources which have been

used. Some systems try to parse the whole sentence (LSP, Ménélas), others try to process large segments of the sentence or local phrases only (RECIT, SPRUS). However for sentence processing the system MedLEE consecutively uses all methods mentioned above. All the systems use different amounts of syntactic, semantic, and domain knowledge and their combinations vary considerably. Some of them include knowledge concerning the sentence structure, others rely on semantic patterns/frames. It seems that the methods based on analysis of phrases rather than complete sentences show substantial increase in recall while incurring only a small loss in precision.

Leroy et al. [3] developed a shallow parser that captures relations between noun phrases in medical abstracts. The parser has a syntactic basis and it extracts relations between all noun phrases (NPs) regardless of their type. It uses AZ Noun Phraser with adjusted settings to extract medical nouns and NPs. The parser searches templates in the texts which are based on English closed-class words, i.e prepositions (by, of, in), negation, conjunctions (and, or) and auxiliary or modal verbs. The extracted relations can contain up to five arguments: relation negation, left-hand side (LHS), connector modifier, connector and right-hand side (RHS). For instance:
NOT : Hsp90 (LHS)—inhibit (connector)—receptor function (RHS)
The parser recognizes two types of negations: negation that precedes a verb phrase and negation that is part of a noun phrase. Please note that negation is just recognized but not processed further. If there exists a marker of negation in a template, the template is ignored.

Another attempt for negation recognition in medical texts has been made by Mutalik et al. [4]. They have created a Negfinder program consisting of a lexer and a parser. The lexer identifies a lot of negation signals and further classifies them on the basis of properties such as whether they generally precede or succeed the negated concept and whether they can negate multiple concepts. A single token is corresponding to each class. Taking into account the token and some grammar rules, the parser associates the negation signal with a single concept or with multiple concepts preceding or succeeding it. The parser relies on a restricted subset of context-free grammars and it partially parses the sentence focusing on the occurrence of concepts matched to the UMLS, negation signals, negation terminators, and sentence terminators. It treats most other words just as fillers. The output of the Negfinder shows that a large part of negations can be detected by a simple strategy but the reliability of detection depends on the accuracy of concept recognition, which is hampered by composite concepts and homonyms. The Negfinder has precision 91% and recall 96%. The authors conclude that "*in most cases, errors made by Negfinder are easily correctable by syntactic methods and involve minor modifications of the lexer or the parser. However, in some cases semantic methods may be required, such as better characterization of temporary composite concepts using noun phrase detection combined with a rich semantic model of the domain*."

Chapman et al. [5] propose a simpler algorithm (NegEx) than the one used in Negfinder which can detect large portions of negations. The algorithm relies on a set of negation phrases and regular expressions to negate UMLS terms. Although it can be easily extended, the algorithm has lower recall since it limits the number of words between the negation phrase and the UMLS term up to five. It performs with 84% precision and 78% recall. NegEx has even lower precision when the negation phrase is "not". The authors of [6] made an attempt to improve the precision in case of "not" by using Naïve Bayes and Decision Trees machine learning algorithms. They have analyzed a sample of sentences which NegEx inaccurately negated. The result of their analysis could be summarized into a simple rule whose addition to the NegEx increases the analysis was restricted to the specific negation phrase.

3. System Architecture

The mail goal of the MEHR (Maintaining Electronically Health Records) is to extract from EHRs in natural language (NL) all required information for automatic generation of Patient's Chronicle – symptoms and diagnosis. We propose an approach that combines information extraction with deep semantic analysis, allows to determine the

negation, negation's scope and to treat it appropriately. MEHR is a prototype of a system for automatic recognition of medical terms and some facts in EHRs in Bulgarian. This is the first step towards filling in a template concerning a patient's status. Automatic extraction of facts needed for description of patient's status in full is our ultimate goal. However in this paper we just focus on proper treatment of negation.

The MEHR architecture is strongly influenced by EHRs specific structure. The information in EHRs in Bulgarian is organized in 11 ordered predefined topics: Personal data, Anamnesis, Status, Examinations, Consultations, Debate, Treatment, Treatment results, Recommendations, Working abilities, Diagnosis. The average length of each EHR is about 2-3 pages.

MEHR system works in two modes – for filling symptoms scenarios templates and for filling diagnosis scenarios templates. In the current version of MEHR system we do not treat temporal structure and discourse, thus we can not determine relations between a diagnosis and all corresponding symptoms that caused this diagnosis.

MEHR system uses the following resources: lexicon, grammar rules, negation rules, terminology bank, shallow ontology of body parts and templates.

MEHR architecture is shown on fig.1. The system consists of the following modules:

- A&C : Annotation and chunking
- Post Processing Module
- Negation Treatment Module
- Extractor
- Filling Scenario Templates Module

Each EHR is split into different topics and then it is sent as an input to MEHR. The system processes one separate EHR topic per time. The A&C module (programmed in Perl) annotates the text and extracts chunks from it. The annotation process is based on morphological analysis. It uses a lexicon, which consists of 50 000 lexems and contains all their wordforms. Words in the text are juxtaposed to lexicon entries and for each word the module finds the word's basic form (lexem) with its lexical and grammatical features. Chunks are sequence of words that forms a syntactic group. We have defined a nominal chunk as an adektive followed by a noun. We define an adektive as a word that syntactically behaves as an adjective which is coordinated with the succeeded noun. The adektives describe some attributes of the succeeded noun. Chunks are recognized by rules (regular expressions), which take into account morphological features of words and their mutual position. The module recognizes mostly nominal chunks (NPs).

The output of A&C system is a tagged text with information about recognized nominal chunks. The lexicon that uses A&C is expanded with medical terminology and frequently used words in EHRs. Post Processing Module using lexicon, verb frames for some domain important verbs and grammar rules determines some VP chunks. The negation treatment module inserts markers in the text for negated phrases and determines scope of negation by using negation rules. More detailed information about this process can be found in the next paragraph. The output of negation module is used as an input of Extractor. The extractor determines patient's symptoms or diagnosis with the help of Terminology Bank and Ontology for Diagnosis, Shallow Ontology for Body parts and Frozen Phrases Templates. In the current prototype the diagnosis extraction is in the initial stage of implementation. The Filling Scenario Templates module tries to fill all obligatory fields in Patient's Chronicle Template and some of optional fields if there is additional information.

4. Negation Treatment

We treat the negation in the context of a sentence, so we will briefly describe its semantics in this context. In Bulgarian as in other languages, it is possible either to negate the whole statement or to negate different components, arguments of the sentence predicate. In the former case of negation, the negation is general while in the latter it is partial. The negation is expressed by specific lexical means: particles that form complex predicate negation, pronouns and adverbs as well as verbs with general semantics of "absence", "lack", "inhibit" etc. We will give some general classification of negation in Bulgarian language and discuss the different expressions of negation in EHRs.

4.1. Expressions of negation in Bulgarian

4.1.1. Surface markers of negation

General negation i.e negation of the verb action. It is expressed by the negative particle "не"(not), which is considered as a part of the verb complex. The usage of particles "нито" (neither) and "ни" (nor) is just to repeat and exalt the negation. However, their usage together with the negation itself shifts the scope of the negation from general to partial i.e negation of the arguments of the main predicate. For instance, take into account the semantics of the following sentences: "He doesn't drinks". vs. "He doesn't drink neither wine nor whiskey." The particles that exalt the negation can be used independently in a positive sentence and they are semantically related to the succeeding word: "Neither drugs, nor any treatments can help him." Another way of negating a verb action in Bulgarian is by preceding it with the preposition "без" (without) followed by particle "да" (in fact this sequense correspond to a double conjunction). Since the semantics is the same MEHR treats "не" and "без да" in the same way.

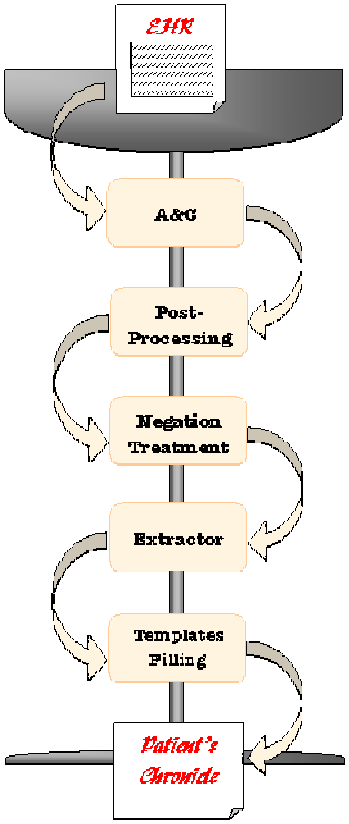

Fig. 1 MEHR Architecture

Partial negation. The scope of the partial negation is spread out to any other parts of the predicate-argument structure as well as to their attributes except for the predicate itself. The negation is again expressed by the usage of lexical means, which can negate:

- the presence/existence of some attributes of an object without being explicitly mentioned. The lexical means of expression are *(i)* negative adverbs and pronouns: "никъде, никога, никакъв" (nowhere, never, nobody) etc. In contrast to English, negative adverbs and pronouns always presuppose general negation as they are always used together with negation of the predicate, *(ii)* prepositions with meanings of an "absence" e.g. "без" (without). These prepositions negate the presence/existence of an attribute expressed by a noun. In this case the attribute is mentioned explicitly e.g. "дишане без хрипове" (breathing without crepitating).
- the presence of an argument of some predicate with focus on the absence itself and mentioning the argument explicitly. The means of expression here are the usage of lexems, in which the negation is part of their meanings e.g. "няма" (there is not, not exist) – negation of "има" (there is, exist); "липсвам" (absent) – negation of "присъствам" (be available); "отричам" (deny) – "признавам" (confess). The words that have a negative meaning but a positive form we will call inherent negatives. In this kind of negation the negative particle "не" (not) is missing at the surface level.

4.1.2. Negation of complex syntactic units

- In coordinative sentences the negative particle "не" (not) precedes each negated predicate i.e the number of simple sentences and the number of negative particles "не" is equal. The list of negated verbs is separated by commas or by coordinative conjunction "и" (and). It is also possible to have a contrastive conjunction "а" (but). Examples: "Не са били установени промени и не е хоспитализирана" (No changes were found and the patient has not been in hospital).
- Homogeneous parts of a sentence (members of coordination): when the negative particles "не" (not) or "без" (without) precede a chain of homogeneous sentence parts, they spread out the negation to the whole syntactic group not only to the nearest group's element. Example: "без хирзутизъм, аменорея" (without hirsurtism, amenorea).

- ❖ Heterogeneous parts of a complex sentence: when the negative particle "не" precedes the chain of heterogeneous sentence parts, it spreads out the negation to the nearest to it element.

4.2.Negation in EHRs and its treatment

Expressions of negation mostly occur in Anamnesis, Status and Debate sections of EHRs. These sections include crucial information about disease, patient status and ways of treatments. Anamnesis and Status contain descriptions of diseases with their symptoms (symptom complex of disease). The clinical chronicle is in the Debate (reasons for disease, patient's problems, treatments and results). Since these sections contain very important information for patient's chronicle correct interpretation of negation is needed.

We have analyzed negations in EHRs by using surface markers of negation and a concordancer. The most frequently used negation markers are shown in table 1.

Negation marker	Statistics	Total words
не (No)	350	55 000
без (without, missing)	200	
отрича (deny)	35	
липсва (missing, absent)	30	

Table1. Most frequent negation markers

Roughly we can categorize negation into two types: a direct negation and a distant one.

4.2.1. Direct negation

The negative markers directly precede some verb chunk or noun (nominal) chunk. In the first case the MEHR interprets the negation as an absence of action or state expressed by the verb – "не вижда" (not see). In the second case, the negation is interpreted as negation of the whole nominal chunk without specifying if it negates existence of some object or its attribute. Usually the noun of the nominal chunk is the subject of negation, however, in some cases (domain terms) the noun has less semantic importance than the adektive.

Example of direct negation:

*"...**без** шумова находка..."*

*"...**няма** патологични промени..."*

*"...**липсват** видими структурни дефекти..."*

*"...**отсъстват** хрипове..."*

In such cases we consider the whole noun chunk as the scope of negation without further deep processing. In future we plan to separate noun chunk into sub-chunks and to look at the most frequent sub chunks. We expect that the most frequent ones will be more meaningful and they can be potential candidates in determining the negation scope into chunks themselves.

The MEHR extends the definition of a chunk into compound chunk where it combines nouns, noun chunks or both when there is a conjunction between them.

*"...**липсват** отоци и варикозни промени..."*

For determining the scope the system treats the compound chunk like a conjunction of simple ones:

*"...**липсват** отоци..."* and *"...**липсват** варикозни промени..."*

In cases where the negation marker is succeeded by a noun, if the noun belongs to our terminology bank or body parts ontology then the negation scope is the noun itself otherwise the negation scope is unsolved. In order to solve the latter case we have to recognize not only chunks but whole noun phrases (mostly nouns connected by prepositions).

4.2.2. Distant negation

The distance negation includes cases where the negative marker is at some distance from the subject of negation and it hampers the correct interpretation. At the surface syntactic level the negation is attached to some object but at the semantic level it relates to another (shifting) or spreads the negation to another (distribution). There is a relationship between distant objects and the treatment of negation strictly depends on it. Since the negative particle "не" (not) is the most frequent negation marker in the EHRs the following classification of relationships between distant negation objects will be related to it. At present MEHR takes into account the

relationship between two objects. Let's call them A and B by convention.

Predicate-argument relationship
B fills some valency role of A. In this case we believe that there is a distribution of the negation over A and B.
Examples:

"***Не съобщава*** *за минали заболявания.*"
"***Не е изслдван*** *ацетон в урината*"
"***.. да не се включва*** *дигоксинът*"

The treatment of negation depends strictly on the semantics of the negated verb. For the most frequent verbs in EHRs we have defined templates that describe verbs and their valency roles.
Example of templates:

X ***съобщава*** *за Y на Z*
X ***изследва*** *Y в Z*
X ***включва*** *Y в Z*

Such templates are defined for all inherent negatives (липсва, отрича, няма, отказва, изключва, отхвърля, спира) too. We have 18 templates for the verbs.

Modular relationship
A explains B by the usage of a modal verb. B is either a verb or a noun derived by a verb. In this case at a surface level the modal verb is negated but semantically the action/state of the main verb is negated. The MEHR treats this situation by applying a rule using a list of modal verbs in Bulgarian language. The negation scope is defined as B with some certainty factor depending on the semantics of the modal verb.
"***не се нуждае*** от *лечение*"

Copula relationship
A is a semantic empty verb used as a syntactic relation to B. In the restricted domain these empty verbs can be enumerated. MEHR uses a list of such verbs (наблюдава се, проявява се, установява се...)
"***не е установен*** *ацетон в урината*"

Anaphora
Since the MEHR tries to determine negation scope using shallow processing only, it doesn't recognize negation in the case of anaphora between A and B.
"препоръчано ***лечение с метизол*** *което* болната ***не е приемала***"

It is important to mention that surface negation markers that have same interpretation are grouped into clusters. For each cluster we associate some semantic treatment. In the example above all markers belong to the same cluster.

5. Extractor

Specially implemented Splitting module separates each EHR into 11 parts corresponding to the 11 topics mentioned above. We will focus on those of them that are important for MEHR system processing. Anamnesis, Status and Debate contain information about patient's symptoms (see fig.2). Anamanesis, Debate, Consulations and Dignosis contain information about patient's diagnosis. The Status part contains all current symptoms of the patient. After processing these parts some templates related to symptoms are filled in. They have the following structure:
(body part – sympt1, sympt 2 ,…).
The body parts are recognized by using a terminology bank and a shallow ontology.

After careful analysis of EHRs we have found that some phrases have special meaning and can serve as clues for the recognition of symptoms, key events etc. We call these phrases "frozen phrases". For distinguishing symptoms we have defined templates for symptoms frozen phrases. An example of such template is:

"постъпва {по повод на, с оплаквания от, в състояние на, във връзка с}"

The first part of a template is a verb that means „entering the hospital for treatment" and the second part is a list of synonym phrases after which symptoms are expected.

The analysis also shows us that phrases often occur after negation markers but serve as a conjunction between the markers and the negated objects e.g. "*от страна на, данни за, за наличие на*". These phrases we call stop/empty phrases. For each such phrase we have a template too

6. Example

Let's consider that we have as input in the MEHR system the following text form Status

part of EHR of a patient (negation words are bold).

TEXT

Шия със запазена подвижност. Щитовидна жлеза и периферни лимфни възли **не** се палпират увеличени. Крайници със запазени пулсации, **липсват** отоци и варикозни промени. Сърдечна дейност: ритмична, ясни тонове, **без** шумова находка.

The A&C module annotates the given text and extracts chunks from it, using the following resources: rules presented as Regular expressions for recognition of nominal chunks, grammar rules and lexicon extended with medical terminology. Below are presented the annotations sentence by sentence of the input text:

ANNOTATION

Sentence 1:

Шия{шия.N+F:s,шия.V+IPF+T:R1s}
със{със.PREP}
запазена{запазен.A+GR:sf,запазя.V+PF+T+NSE:Psf}
подвижност{подвижност.N+F:s}.

Sentence 2: Щитовидна{щитовиден.A:sf}
жлеза{жлеза.N+F:s} и{и.CONJ}
периферни{периферен.A+GR:p}
лимфни{лимфен.A:p}
възли{възел.N+M:p} не{не.PC}
се{се.PC,себе.PRO+RFL:SA}
палпират{палпирам.V+PF+T+NSE:R3s:I2s:E2s:E3s}
увеличени{увелича.V+PF+T+NSE:Pp}.

Sentence 3: Крайници{крайник.N+M:p}
със{със.PREP}
запазени{запазен.A+GR:p,запазя.V+PF+T+NSE:Pp}
пулсации{пулсация.N+M:p}
липсват{липсвам.V+IPF+I:R3p}
отоци{оток.N+M:p} и{и.CONJ}
варикозни{варикозен.A:p}
промени{променя.V+PF+T+NSE:R3s:I2s:E2s:E3s,промяна.N+F:p}

Sentence 4: Сърдечна{сърдечен.A+GR:sf}
дейност{дейност.N+F:s} -
{gb}ритмична{ритмичен.A+GR:sf},
ясни{ясен.A+GR:p}
тонове{тон.N+M:p}, без{без.PREP}
шумова{шумов.A:sf}
находка{находка.N+F:s}.

Below are listed found nominal chunks form A&C module and some VP chunks (??) founded from Post Processing Module .

FOUND CHUNKS:

Запазена_подвижност
Щитовидна_жлеза
Периферни_лимфни_възли
Запазени_пулсации
Варикозни_промени
Сърдечна_дейност
Ясни_тонове
Шумова_находка

After that the negation treatment module inserts markers <NEG> in the text for negated phrases and determines scope of negation using negation rules presented as Regular expressions. Then scope of the negation is marked with initial marker <NEG> and final marker </NEG>.

The extractor determines patient's symptoms or diagnosis with the help of Terminology Bank and Ontology for Diagnosis, Shallow Ontology for Body parts and Frozen Phrases Templates.

INTERNAL INTERPRETATION

Sentence 1: <Body Part = Noun> PREP <STATUS=NP chunk>.

Sentence 2: <Body Part = NP Chunk> & <Body Part = NP Chunk> <NEG> <Frozen phrase> < STATUS = ADJ> </NEG>.

Sentence 3: <Body Part = Noun> - PREP < STATUS1 =NP chunk > & <NEG> < STATUS2=Noun>& < STATUS3=NP Chunk> </NEG>

Sentence 4: Template: **:** <Body Part = NP Chunk> : < STATUS1> & < STATUS2> & <NEG> < STATUS3=NP Chunk> </NEG>

The Filling Scenario Templates module tries to fill all obligatory fields in Patient's Chronicle Template and some of optional fields if there is additional information. For instance the obligatory fields in the templates include information such as: Patient's name, age, sex, address, list of previous diagnosis and etc. The optional fields can include the status information of some patient's body parts which data are not directly related to the current patient's diagnosis or these data are not important for the current diagnosis recognition. However the status for some body parts is very important for diagnosis

recognition that is why they are marked in the scenario template as obligatory fields.

The scenario templates are presented in MEHR system in XML format.
The following table shows a part from the scenario template about the patient with diabetes diagnosis. These data present information about body parts and their status. Those data that were recognized as negated in the input text are marked in the Status field with marker NEG. If there is information concerning presence or absence of several symptoms for one and same body part then in the scenario template for each pair "body part – status" is allocated a separate field.

RESULT

Body part	**Status**
шия	запазена подвижност
щитовидна жлеза	NEG палпира увеличена
периферни лимфни възли	NEG палпира увеличена
крайници	със запазени пулсации
крайници	NEG отоци
крайници	NEG варикозни промени.
сърдечна дейност	ритмична
сърдечна дейност	ясни тонове

7. Discussion

The presented system has been tested on about 70 EHRs containing 55 000 words in total. Although the relatively low number of negation markers in processed EHRs, their correct interpretation is very crucial as:

- negation markers appear more often in those parts (Anamnesis, Status, Debate) of EHRs that contain important information concerning patient's status.
- Very often negation marker precedes a coordination phrase so one marker is used for negating of not only one item but of a sequence of items.

The results of analysis show that about 57% of negations were recognized correctly, 28% were recognized incorrectly and about 15% were not recognized at all. In most cases incorrect recognition actually means inaccurate interpretation of negation scope. Careful analysis of EHRs and the choice of frozen phrases templates have the highest influence on interpretation process. In the example below the first two items could be interpreted correctly by a suitable template but the third one needs deeper semantic analysis.

Example:

"**няма** *оплаквания от крайниците"*
"**няма** *оплаквания от болки в крайниците"*
"**няма** *оплаквания от години"*

8. Conclusion and further work

We proposed an approach for treatment of negation in EHRs that uses shallow processing (chunking only) in combination with deep semantic analysis in certain points. The choice of proper templates for recognition and interpretation of negation influences the system performance.

As further work we plan to refine the chunking algorithm, to enlarge the number of templates and to expand the knowledge resources of the system (lexicon, ontology etc.).

References

[1] Friedman C, Hripcsak G, Shablinsky I. *An evaluation of natural language processing methodologies*. In Chute CG, ed. Proceedings 1998 AMIA Annual Symposium. Phil: Hanley & Belfus, 1998, p. 855-859.

[2] Friedman, C. and Hripcsak, G. *Natural language processing and its future in medicine*. Academic Medicine. 1999;74(8), p.890-895.

[3] Leroy, G., Chen, H., and Martinez, J.D. *A Shallow Parser Based on Closed-class Words to Capture Relations in Biomedical Text*. Journal of Biomedical Informatics (JBI) vol. 36, pp 145-158, June 2003.

[4] Mutalik PG, Deshpande A, Nadkarni PM. *Use of general-purpose negation detection to augment concept indexing of medical documents*. J Am Med Inform Assoc 2001; 8:598–609.

[5] Chapman, W. W., W. Bridewell, et al. (2001). *A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries*. Journal of Biomedical Informatics 34(5): 301-310.

[6] Goldin, Ilya M., Chapman, Wendy W. *Learning to Detect Negation with 'Not' in Medical Texts*. Proc. Workshop on *Text Analysis and Search for Bioinformatics* at the 26th ACM SIGIR Conference (SIGIR-2003). Eds. Eric Brown, William Hersh and Alfonso Valencia.