



MASTER'S THESIS

Negation Detection in Medical Documents Using Syntactical Methods

accomplished at the
Institute of Software Technology & Interactive Systems
of the Vienna University of Technology

under supervision of
Ao.Univ.Prof. Mag. Dr. Silvia Miksch and Mag. Dr. Katharina Kaiser

by

STEFAN GINDL
9925024

Weissenwolffgasse 12
1210 Wien

Vienna, 4. March 2008

*Dedicated to my mother.
Without you this would have never been possible.*

Abstract

Medical information is often stored in a narrative way, which makes the automated processing a difficult and time-consuming task. Persons responsible for the authoring of medical documents do not take care of a further processing with automated systems. So, information stored in medical writings is not directly usable for the processing with computers. Due to this, efforts have been made to transfer these narrative documents in a format easier processable with computers.

This matter of fact also applies to clinical practice guidelines (CPGs). As many medical documents, CPGs are written in a narrative speech as well, without regards to a computer-assisted processing. For the implementation of CPGs in medical facilities an automated processing is therefore desirable. An important fact is that a lot of information in CPGs is provided in a negated form, expressing that certain circumstances in patients or treatments are not available, existing or applicable. Although negated, this information is nevertheless very useful, since it can express the absence of certain conditions or diseases in patients. Moreover, negations can describe which treatment options should not be taken into account for a given patient, helping a practising physician or nurse in his/her decision process for the assortment of a proper treatment. Thus, a proper Negation Detection in CPGs is an important task for the automated processing of this type of medical documents. It helps to accelerate the decision making process and can support medical staff in their care for patients.

We developed algorithms capable of Negation Detection in CPGs. We use syntactical methods provided by the English language to achieve a precise detection of occurring negations. According to our results we are convinced that the involvement of syntactical methods can improve Negation Detection, not only in medical writings but also in arbitrary narrative texts.

Kurzfassung

Medizinische Daten werden in der Regel in natürlicher Sprache abgespeichert. Dies liegt daran, dass medizinisches Personal im Normalfall keine Rücksicht auf eine spätere Verarbeitung in automatisierten Systemen nimmt. Die Informationen sind daher nicht sofort für eine Verarbeitung mit Computern verfügbar. Aus diesem Grund wurden vielfach Anstrengungen unternommen, um die Daten in ein für Computer verarbeitbares Format überzuführen.

Dieser Umstand gilt auch für medizinische Leitlinien (sogenannte “Clinical Practice Guidelines” (CPGs)). Sie sind, wie alle medizinischen Dokumente, ebenfalls in natürlicher Sprache verfaßt, ohne Rücksichtnahme auf eine computergestützte Verarbeitbarkeit. Die Einführung von medizinischen Leitlinien im klinischen Umfeld macht allerdings eine automatisierte Verarbeitung wünschenswert. In diesem Zusammenhang muß weiters beachtet werden, dass ein großer Teil der Information in medizinischen Leitlinien in negierter Form vorhanden ist. Die Tatsache, dass diese negiert vorliegt, macht sie aber deswegen nicht irrelevant oder weniger beachtenswert. Im Gegenteil, oftmals werden auf diese Art und Weise ungeeignete Behandlungsverfahren gekennzeichnet, beziehungsweise Behandlungen, die bei gegebenen physiologischen Zuständen nicht durchgeführt werden sollen oder dürfen. Solche Informationen können den Behandelnden helfen, die richtigen Therapieverfahren anzuwenden oder auch gewisse Verfahren abzulehnen. Dementsprechend ist die Erfassung von Negationen (die sogenannte “Negation Detection”) in medizinischen Leitlinien ein essentieller Bestandteil für die automatisierte Weiterverarbeitung dieser Art von medizinischen Dokumenten. Entscheidungsprozesse werden somit beschleunigt und dies wirkt sich positiv auf die Behandlung der Patienten aus.

Um diese Aufgabe zu lösen haben wir Algorithmen entwickelt, die das Auffinden von Negationen in medizinischen Leitlinien ermöglichen sollen. Dazu benutzen wir syntaktische Verfahren, die die englische Grammatik gezielt ausnützen, um aufgetretene Negationen exakt aufzuspüren. Unsere gewonnenen Resultate zeigen, dass die Verwendung syntaktischer Methoden dabei hilft, die Identifikation von Negationen zu verbessern, wobei unsere Methoden nicht nur auf medizinische Dokumente, sondern auf jede Art von natürlichsprachlichem Text anwendbar sind.

Contents

Contents	iv
1 Introduction	1
1.1 Motivation	1
1.2 Research Question	2
2 Clinical Practice Guidelines	5
3 Related Work	8
3.1 Background	8
3.1.1 SNOMED - Systematized Nomenclature of Medicine - Clinical Terms [®]	8
3.1.2 International Classification of Diseases and Related Health Problems (ICD)	9
3.1.3 Unified Medical Language System (UMLS)	11
3.2 Evaluation	12
3.2.1 Sensitivity/Recall	12
3.2.2 Specificity	12
3.2.3 Positive Predictive Value (PPV)/Precision	13
3.2.4 Negative Predictive Value (NPV)	13
3.2.5 Correlation between Statistical Parameters	13
3.3 NegEx	13
3.3.1 Preprocessing of Selected Documents	14
3.3.2 Description of the Algorithms	14
3.3.3 Performance Analysis	15
3.3.4 Applications of NegEx	15
3.4 NegFinder	16
3.4.1 The Components Pipeline	16
3.4.2 Concept Finding	17
3.4.3 Input Transformation	17
3.4.4 The Lexical Scanner	17
3.4.5 The Parser	18
3.4.6 Evaluation	18
3.4.7 Applications of NegFinder	19
3.5 A Controlled Trial of Automated Classification of Negation from Clinical Notes	19
3.5.1 Method	19
3.5.2 Statistical Analysis	20
3.6 Negation Processing in Electronic Health Records	20
3.6.1 System Architecture	21
3.6.2 Details of Negation Identification	21

3.6.3	Evaluation	22
3.6.4	Conclusio	22
3.7	Automatic Mapping Clinical Notes to Medical Terminologies	22
3.7.1	Method	22
3.7.2	Preprocessing	23
3.7.3	Negation Detection	23
3.7.4	Evaluation	23
3.8	Concept Negation in Free Text Components of Vaccine Safety Reports	23
3.9	NegExpander	24
3.10	A Hybrid Approach	24
3.11	Discussion	26
4	NegHunter	27
4.1	MetaMap Transfer (MMTx)	27
4.2	The Negation Detection Algorithms	30
4.2.1	The Adverbial Negation	31
4.2.2	The Intra-Phrase Triggered Negation	36
4.2.3	The Prepositional Negation	37
4.2.4	The Adjective Negation	39
4.2.5	The Verb Negation	39
4.2.6	Assigning Prepositional Information	40
4.3	Summary	41
5	Evaluation	43
5.1	Distribution of Identified Negated Phrases	43
5.2	Analysing the Performance of NegHunter	44
5.2.1	Errors Related to MMTx	47
5.2.2	Discussion	49
6	Conclusion	50
	Bibliography	52

List of Tables

1.1	The training set of the used CPGs.	3
1.2	The test set of the used CPGs.	4
3.1	Comparison of the performance of the baseline algorithm and NegEx.	15
3.2	Recall and Precision of NegEx in Pathology Reports	16
3.3	Differences in sensitivity and specificity in the two different examining methods.	19
3.4	The procedure proposed by [Elkin et al., 2005] reaches high values in all statistical parameters.	20
3.5	Performance of the Hybrid Approach on 120 Radiology Reports	25
4.1	A Simple Past Active Sentence and the splitting into its phrase components	32
4.2	The classification of the phrase in a Simple Past Passive sentence	33
4.3	The classification of a phrase in Present Perfect Passive	34
4.4	In the active voice the phrase classification is similar for seven tenses.	34
4.5	The structure of Present and Past Progressive Active.	35
4.6	Future II and Conditional II in Active Voice spanning three Verb Phrases with their usage.	35
4.7	Simple Present and Simple Past in the Passive Voice	35
4.8	Present Perfect, Past Perfect, Future I and Conditional I Passive.	35
4.9	Present and Past Progressive Passive.	35
4.10	The passive form of Future II and Conditional II.	35
5.1	Evaluation results of the adverbial negation	45
5.2	Evaluation results of the intra-phrase triggered negation	46
5.3	Evaluation results of the prepositional negation	46
5.4	Evaluation results of the verb negation	46
5.5	Evaluation results of the adjective negation	47
5.6	Overall evaluation results	47
5.7	Recall and Precision in each negation class	47
5.8	Errors of MMTx related to their negation classes	48
5.9	Percentage of MMTx errors compared to overall errors	49

List of Figures

3.1	Illustration of the SNOMED-CT [®] components [SNOMED International, 2006]	10
4.1	Entity-Relationship-diagram of the textfeature package of MMTx [Divita, 2005]	28
4.2	Entities of MMTx's textfeature package used within <i>NegHunter</i>	29
5.1	Distribution of negated phrases	44

Chapter 1

Introduction

A large part of medical information is only available in natural language, which makes an algorithmic processing with computers a very difficult task. So to say, this information is “locked up” in files and databases in a format not suitable for automated processing [Hripcsak et al., 1995]. A lot of information in medical writings also underlies the concept of negation, which means, that certain treatment options or parts of a diagnosis, cannot or must not be applied for a specific patient. Negation Detection has been treated under the circumstances of pathology reports [Mitchell et al., 2004], discharge summaries or radiology reports [Chapman et al., 2001]. Still, the processing of negated terms in more complicated writings is an open topic. Negation detection demands a lot of knowledge about language itself to correctly identify negated words or terms in free text. Endeavour on this topic may help to develop computer assisted treatment strategies to improve patient care.

The detection of negated terms in clinical practice guidelines (CPGs) allows the faster filtering of relevant information. It may be important for a practitioner to decide, which treatment options are available for a patient and which are not. This information can also be combined with the occurrence of specific physiological conditions. Certain treatments cannot be conducted under certain physiological circumstances. A fast access to this information can contribute to a better treatment of patients. A former processing of guidelines including a Negation Detection supports a practitioner in his/her decision process. Otherwise, the human researcher would need to read through all possible options by himself, which can be a cumbersome and time-consuming process.

1.1 Motivation

Negation is an important part of inter-human communication. It can be used to invert concepts and to show refusal of opinions. Sometimes persons in life-threatening situations use the simple word “no” to express their helplessness. In the development of a person’s psyche and character a “no” implies that this person has developed his or her own will and is able to express this. A person who never says “no” will never be asked to give his or her view on a problem, since the answer is already known. The concept of negation is therefore a universal concept in the most languages if not in all languages.

The existence of negations in language has been philosophised by man since more than 2000 years. Even Aristotle played with this peculiarities and tried to find a classification of the different types of negation in natural language and found four classes of negation. He came to the result that negation can be splitted into four types which he named as correlation (e.g., double vs. half), contrariety (e.g., good vs. bad), privation (e.g., blind vs. sighted) and contradiction

(e.g., he sits vs. he does not sit) [Horn, 1989].

The right understanding of a negation demands a competent knowledge of the used language, since even one word can completely change the sense of the statement. The statement can then be inverted, weakened or amplified. The following simple example by Horn [Horn, 1989] shows this effects in negated sentences (as cited in [Mutalik et al., 2001]):

1. I'm not tired.
2. I'm not a bit tired. (which equals "I'm not at all tired.")
3. I'm not a little tired. (which equals "I'm quite tired.")

Even this example shows the complexity of a negation in natural spoken or written language. In the first sentence the negation is very clear and one cannot be confused in its meaning. The following two sentences generate subtle problems, because one of them states the complete absence of tiredness ("I'm not a bit tired."), whereas the other sentence describes a condition of exhaustion ("I'm not a little tired."). A german reader who is not very familiar to the english language can get problems with phrases when they are written down. The second and third sentence seem to be equal and the lack of mimic art or a tonal amplification leaves the reader confused. An english native speaker will in contrast to a non-native speaker have no problems with the right understanding of the semantic of this sentences.

A search for negations in natural language is therefore a difficult task, but the search for them in a medical scope can be easier. Medical language is much more restricted than narrative speech, a physician will not use stylistic elements like double negation extensively to write reports or patients histories. Thanks to this fact a negation detection in a medical scope is supposed to be less difficult because of this restrictions.

Negation detection contributes to an information reduction when searching in databases. Search results can be minimized to the entities searched for, otherwise even negated entities would be returned in queries. This work contains already existing algorithms for the task of Negation Detection. It shows, how possible approaches can look like. We also present our approach of Negation Detection in the area of CPGs.

1.2 Research Question

In this work we examine the question, which methods best fit for a Negation Detection in CPGs. As medical guidelines are not written in colloquial prose the assumption was obvious that occuring negations would not be of the same complicated type as they are in person-to-person language. Guidelines may lack the often vast usage of stylistic elements such as the double negation. This fact may make it easier to come up to algorithms which deliver a proper result.

In order to answer the asked question the research work was divided into following five sections:

Examining existing studies on Negation Detection. Initially we examined existing works related to Negation Detection. This part of our work is relevant to decide, whether existing algorithms or methods can be used for our research.

Selection and analysis of a set of guidelines for developing and testing of our method.

We have chosen a set of 20 CPGs which all deal with the topic "cancer". We used four of these guidelines to get an overview over the occuring negations (see Table 1.1). With this information we developed methods to classify and identify negations.

Development of algorithms for Negation Detection in CPGs. After the hand-reading of the guidelines we have decided a proper Negation Detection of CPGs. We have analyzed, if already existing methods are applicable or if we had to develop algorithms tailored to CPGs.

Implementation. This part of our work contains the implementation of our methods, to see, if the approach is successful. Furthermore, a refinement was carried out to make the methods fit to the circumstances of the field of CPGs.

Evaluation. The evaluation presents the results obtained from our implementation of the algorithms. We provide statistical parameters used in the area of Natural Language Processing. We used the remaining 16 CPGs (see Table 1.2) selected before for evaluating our algorithms.

Table 1.1: The training set of the used CPGs.

-
1. Breast cancer. [Singapore Ministry of Health, National Committee on Cancer Care, 2004]
 2. Cervical cancer. [Singapore Ministry of Health, 2004a]
 3. Chemotherapeutic management of stage IV non-small cell lung cancer. [Socinsky et al., 2003]
 4. Chronic cough due to lung tumors: ACCP evidence-based clinical practice guidelines. [Kvale, 2006]
-

In the following section we initially describe CPGs, their characteristics, and how they are used in medical facilities. Then we present related work to the subject of Negation Detection in the medical scope. Furthermore, we explain, what our approach to Negation Detection in CPGs is and how we have come to it. In Section 5, we carry out an evaluation to give measurements of the performance of our methods. In the last section we subsume the results of our work and give future directions for our approach.

Table 1.2: The test set of the used CPGs.

-
1. Colorectal cancer. [Singapore Ministry of Health, 2004b]
 2. Cutaneous melanoma. A national clinical guideline. [Scottish Intercollegiate Guidelines Network (SIGN), 2003a]
 3. Epithelial ovarian cancer. A national clinical guideline. [Scottish Intercollegiate Guidelines Network (SIGN), 2003b]
 4. Guideline for the management of cancer pain in adults and children. [Miaskowski et al., 2005]
 5. Long term follow up of survivors of childhood cancer. A national clinical guideline. [Scottish Intercollegiate Guidelines Network (SIGN), 2004]
 6. Lung cancer. Palliative care. [Kvale et al., 2003]
 7. Management of hepatocellular carcinoma. [Bruix et al., 2005]
 8. Management of patients with lung cancer. A national clinical guideline. [Scottish Intercollegiate Guidelines Network (SIGN), 2005a] Management of transitional cell carcinoma of the bladder.
 9. A national clinical guideline. [Scottish Intercollegiate Guidelines Network (SIGN), 2005b]
 10. Practice parameters for colon cancer. [Otchy et al., 2004]
 11. Small cell lung cancer. [Simon and Wagner, 2003]
 12. Staging evaluation - Hodgkin's disease [Wolkov et al., 2005]
 13. The diagnosis and treatment of lung cancer. [National Collaborating Centre for Acute Care, 2005]
 14. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of multiple myeloma: an evidence-based review. [Hahn et al., 2003]
 15. The role of postoperative chemoradiotherapy for advanced squamous cell carcinoma of the head and neck. [Winqvist et al., 2004]
 16. Use of epoetin in patients with cancer: evidence-based clinical practice guidelines of the American Society of Clinical Oncology and the American Society of Hematology. [Rizzo et al., 2002]
-

Chapter 2

Clinical Practice Guidelines

Clinical practice guidelines (CPGs) are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [Field and Lohr, 1990]. They represent the currently best known practice in a given medical case, they are, so to say, the “state-of-the-art” of a treatment.

The decision options collected in CPGs are gathered according to the best known evidence about a disease. Therefore, multidisciplinary expert groups together discuss the problem and all contribute to a CPG. Yet this is not the end of the development of a CPGs. As medical knowledge also grows and changes in time, CPGs also need to be reviewed from time to time in order to bring them into an actual state. This fact must be treated with special care, since CPGs claim to be the “state-of-the-art”, the currently best known knowledge in a certain medical field.

In the following paragraph we give an introduction into the guideline development process

The development of CPGs [Shekelle et al., 1999] describe the appropriate stepwise development process of CPGs as follows:

1. **Defining the CPG’s topic:** The first step in the development process is the selection of the CPG’s topic. In this step it is important to exactly define which disease is treated. A refinement is inevitable or else the subject matter would be treated too broadly.
2. **Determining the development group:** For a certain topic it is important to gather experts with the right expertise. Moreover, the employment of multidisciplinary groups is also of great significance. Experts from different special fields contribute different opinions for a comprehensive treatment of a disease. For example, the treatment of cancer does not only require surgical knowledge but the knowledge of the usage of drug regimens (e.g., chemotherapy, use of epoetin) as well. Moreover, financial consideration should also be merged with the naked treatment recommendations. Finally, the definition of group leaders is important for the organisation of the group, so that it effectively reaches the given aims.
3. **Analysis of the evidence:** It is important to ensure that all available evidence is collected, and therefore a systematic review of the evidence is the appropriate way of proceeding. To evaluate the available evidence, the underlying study design needs to be considered, as well as studies confirming the relevance of the evidence. Finally, a summarization of the aggregated evidence must be carried out.
4. **Compiling evidence and writing the CPG:** The evidence needs to be prepared for

the deployment in medical facilities. Clinical circumstances as well as policy or financial aspects must be considered. The applicability of the developed CPG must be ensured.

5. **Maintenance of the CPG in the practical environment:** After the introduction of the CPG in a medical facility it needs to be reviewed at discrete points in time, and in case of obsolescence an update is necessary. This ongoing revision of a CPG is of special importance, since CPGs, as mentioned earlier, are thought of as the “state-of-the-art” of a certain medical topic.

As one can see, the development of CPGs is a rather complicated and sophisticated process. However, this is necessary to ensure that CPGs remain the best known medical practice.

Although the best known practice is aggregated in CPGs, this does not mean that they must be performed in each and every case. The last decision, which treatment a patient should be given always depends on the medicating practitioner as well as the patient him-/herself. Only the practitioner has all information of the physiological state of a patient and only the patient is in the position to decide, whether he/she appreciates the treatment option or not. So, CPGs are also an aid for the patients themselves as well as their relatives.

CPGs are developed for certain circumstances, for example particular diseases or conditions in which potential patients can be. Moreover, CPGs can have several target groups, for instance physicians, nurses, patients, or students. In this way they aid the way of deciding, which treatment option properly fits to a patient’s condition and disease. They are helpful for clinicians who “may also use guidelines to answer specific clinical questions arising out of their day to day practice” [Feder et al., 1999]. For patients, CPGs represent information material, so that they can better estimate what is coming up to him/her until his/her (possible) convalescence. If targeted to students, who do not have the experience of a practicing physician, CPGs serve as important teaching material. Nevertheless, in most cases the target group of CPGs will be physicians and nurses.

An important topic in CPGs is the methodology used to analyze the evidence and to assess the quality and strength of the evidence. This means, what kind of research has been carried out to constitute the evidence of recommendations in a guideline. In this point, CPGs suffer from a lack of standardization. Each organization developing CPGs features a different kind of grading schema. This makes a cross-comparison a confusing task [Atkins et al., 2004b]. Efforts have been carried out to tackle this problem. For example, the GRADE working group¹ aims a standardization of the grading schemas for the levels of evidence [Atkins et al., 2005, Atkins et al., 2004a]. Another attempt is to subsequently provide a meta schema based on the different existing schemas available in CPGs [Kaiser et al., 2007]. In this approach the different schemas using different symbols for the levels of evidence are mapped to a meta schema, where all grading schemas are represented by the same symbols, thus making a comparison possible.

The section containing the actual instructions of a CPG is the recommendations section. Here, the intended user can gather information about treatment options, ways of diagnosis in certain cases, and what kind of evidence is provided for that specific recommendation. Although this section contains a lot of instructions, these instructions are not a must in each and every case. The strict obedience of a recommendation does not always lead to good results in the treatment of patients. The physician has to decide for him-/herself, if this treatment option is reasonable for a specific patient with specific physiological characteristics. Recommendations in CPGs can also be meant in a negative way. This makes sense when a certain treatment must not or should not be accomplished for a given patient. In this case it is meaningful to provide a foregoing sorting out or presentation of those recommendations not advisable. A Negation

¹<http://www.gradeworkinggroup.org> (last assessed December 29, 2007)

Detection helps to achieve this kind of information. Thus, treatment decision can be accelerated, since the physician needs not manually read through all sort of recommendations.

The authors of CPGs also give information about potential unforeseeable side effects caused by the implementation of a CPG. For example, some treatments for cancer can cause oncologic diseases by themselves. This fact needs to be regarded, when a CPG is launched in a medical facility. The so-called “Strength of recommendation” gives information on this topic. It describes, how securely a treatment option can be accomplished for certain circumstances. It provides aid for the decision if the specific treatment should be carried out or if there are factors which delimitate the benefits of the treatment. In other words, the strength of recommendation expresses the trade-off between the benefits and potential harms of a therapy according to the underlying level of evidence, which can range from a very high level (achieved by meta-analysis of randomized clinical trials (RCTs)) to a low level (which indicates, that an evidence supporting the recommendation is absent). A distinct “Benefits/Harms of implementing the guideline recommendations” section often also provides supplementary information to this kind of questioning.

For the implementation of a CPG in a medical facility there is also a section for information about this topic. This supports the responsible persons in the implementation process and errors in the implementation can be avoided. In general, it is easier to launch CPGs in medical facilities with a manageable target group. The larger and the more complex the organization’s structure is, the more difficult is an implementation of CPGs. With growing complexity of the organization the barriers of implementation also grow [Feder et al., 1999].

As one can see, the development of a CPG is a very complex process. It can only be done by the cooperation of experts from several disciplines and by studying already existing knowledge about diseases and their options of diagnosis and treatment. The creation of a CPG is also a very sensible process, since false medical instructions can lead to mistakes in the treatment of patients and may lead to complications or even the death of persons. Although a CPG contains the best known knowledge on a certain health topic, the practitioner is always responsible for him-/herself to decide, if a specific recommendation is applicable for a patient or not. As mentioned earlier, they are intended to assist, not to dictate a physician.

Negation Detection contributes to a faster processing of CPGs and allows the responsible persons to achieve relevant information more quickly. As mentioned earlier, negative recommendations can be especially emphasised foregoing, or treatment options not advisable for a certain disease in a patient with a certain condition can be displayed. The proper detection of negations is therefore a contribution to the improvement of the implementation and dissemination of CPGs and their deployment in medical facilities

Chapter 3

Related Work

Although the subject of negation has already been discussed by the ancient Greek there have been little achievements of an automated negation understanding and processing. Specifically in the field of medicine there are only few studies concerning this topic. The two most important studies relating to an automated negation detection in medical documents have been carried out by the University of Pittsburgh and the Yale University School of Medicine. Furthermore, the University of Sofia in Bulgaria has performed a study for the identification of negations in the Bulgarian language, which is by now a merely researched field in this language.

In order to fully understand the explanation of used Negation Detection techniques, the knowledge of some important resources supporting Natural Language Processing in the realm of medicine will be necessary. The following section gives an insight into used resources relevant for the comprehension of the performed studies. Section 5 describes statistical measures necessary for the evaluation of the presented methods. In Sections 3.3 to 3.10 we describe related work for Negation Detection in the medical scope.

3.1 Background

For the automated processing of medical texts systematized terminology collections are applied to index biomedical concepts. A brief introduction into the used terminologies is given in the following subsections. First, we will give a brief summary of the basic vocabulary used under those circumstances, according to [Kaiser, 2007].

- Nomenclature: a collection of names for a certain field of knowledge
- Classification: the attempt to identify similarities in entities and to aggregate them into classes
- Thesaurus (such as the UMLS metathesaurus): simplifies the search in bibliographic databases by providing an inter-relation between concepts.

In the subsequent section we present popular terminology systems created for the medical scope.

3.1.1 SNOMED - Systematized Nomenclature of Medicine - Clinical Terms[®]

SNOMED[®] represents a systematized medical terminology. It is a "comprehensive and precise clinical reference terminology that provides unsurpassed clinical content and expressivity for clinical documentation and reporting." [SNOMED].

SNOMED[®] was developed by the College of American Pathologists (CAP). The CAP had by now worked with the ancestor of SNOMED[®], SNOP (Systematized Nomenclature of Pathology). In the year 1974 SNOP was extended with the help of Dr. Roger Coté in order to contain vocabulary beyond the scope of the pathologic special field and SNOMED[®] was born.

In 1977 SNOMED[®] was turned from its printed representation into an electronic tool, created to facilitate the use for physicians, researchers, and healthcare professionals.

With SNOMED-RT (Reference Terminology) a new generation of SNOMED[®] was introduced in January 1999. SNOMED-RT was created to better fulfill the needs in the fields of health states, disease states, pathophysiology, treatments and outcome.

In the same year in March a fusion of SNOMED-RT and the United Kingdom National Health Service (NHS) Clinical Terms Version 3 (former known as the Read Codes) was performed. SNOMED-CT[®] (Clinical Terms), the currently most comprehensive vocabulary for medical terms, evolved from this merging.

In SNOMED-CT[®] medical information such as diseases is defined by concepts. These concepts are refined by several hierarchies or axes, for example, "disease", "finding site" or "procedure". In order to identify concepts they are provided with a concept ID, which is a simple numerical value. Existing relations among concepts are stored too.

The information provided for a given concept is composed of a unique numerical code, a unique name (the so called "Fully Specified Name") and further descriptions such as synonyms. Moreover, information about the concept's location in the hierarchy and relations between other concepts are recorded.

Important facts of SNOMED-CT[®][SNOMED]:

- Its included Semantic Net contains over 300,000 medical concepts and the relationships among them
- More than 7 million relationships are available
- The concepts cover several fields such as diagnosis, drug definition, findings, procedures, anatomy etc.
- Multiple axes and hierarchies are provided
- It includes three main hierarchies (finding, disease, procedure) and 15 supporting hierarchies
- A description logic is used for concept representation

Figure 3.1 on page 10 shows a graphical representation of the underlying architecture:

As of April 2007, SNOMED-CT[®] was taken over by the International Health Terminology Standards Development Organisation (IHTSDO), which is a non-profit association with place of residence in Denmark.

3.1.2 International Classification of Diseases and Related Health Problems (ICD)

The ICD is released by the World Health Organization (WHO) and emerged from a catalogue of causes of death which was introduced by the French physician Jaques Bertillon in the year 1893. The American Public Health Association (APHA) recommended the adoption of this catalogue for the improvement of medical treatment in the USA, Canada, and Mexico. In order to keep the register up to date the APHA recommended a revision every ten years, which was accomplished

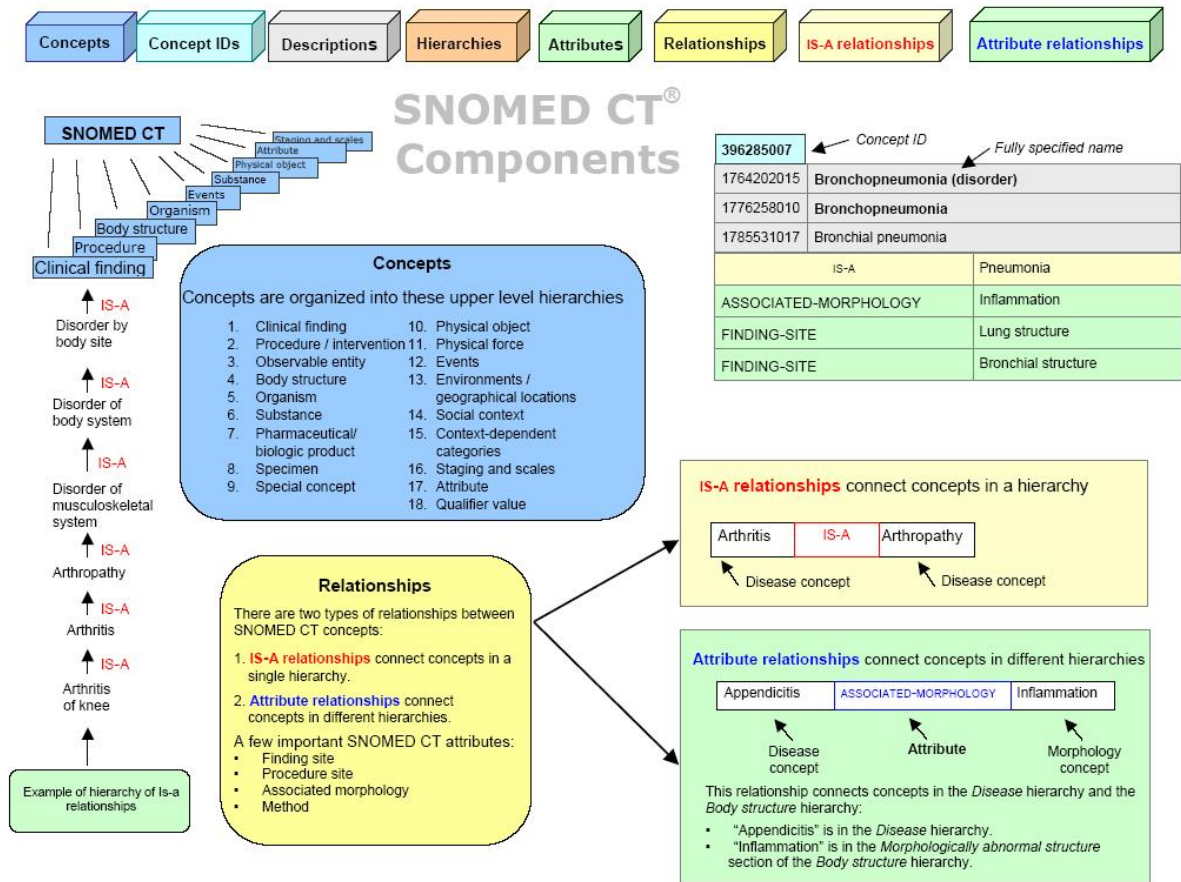


Figure 3.1: Illustration of the SNOMED-CT[®] components [SNOMED International, 2006]

by the Mixed Commissions, a group consisting of the International Statistical Institute and the Health Organization of the League of Nations until 1948.

At that time the WHO was committed the function of revising the ICD, which happened one year later with the 6th revision.

In 1977 the WHO published the 9th revision of the ICD (ICD-9) consisting of three volumes, the first two of them containing diagnosis codes, the third covering procedure codes. From the ICD-9 emerged the ICD-9-CM (Clinical Modifications), which is used for the calculation of hospital treatment costs in order to facilitate the account with concerned insurance companies.

The currently used versions are the 1999 published 10th revision (ICD-10) and the ICD-9-CM.

The ICD-10 provides alphanumeric codes for the identification of diseases or abnormal findings in humans. The ICD contains 22 different fields, each having a unique code. Diseases or findings within a special field are then identified by this field's code and an additional numeric code specific for the finding. For example, the occurrence of neoplasms is coded with the letters C and D, malign neoplasm have the range from C00 to C97 applied. C30 to C39 stands for malign neoplasms of the respiratory and intrathoracic organs. The value C34 is used for malign neoplasms of bronchus and lung. To conclude, the diagnosis "lung cancer in the upper lobe" would be codified by the alphanumeric value C34.1.

3.1.3 Unified Medical Language System (UMLS)

The UMLS is a controlled collection of vocabularies with its internal structure designed to map between them. The system was initially designed in 1986 by the National Library of Medicine, USA.

As UMLS is a collection of various vocabularies, it allows conversion of medical terms from one vocabulary into another. This fact is described in [Humphreys and Schuyler, 1993] as follows:

The UMLS approach assumes continuing diversity in the formats and vocabularies of different information sources and in the language employed by different elements of the biomedical community. It is not an attempt to build a single standard biomedical vocabulary.

The structure of the UMLS contains three main knowledge sources:

1. **The Metathesaurus[®]**: It represents the base of the UMLS, containing 5 million concept names identifying more than 1 million biomedical concepts. These names come from over 100 controlled medical vocabularies such as SNOMED CT or the ICD-9-CM. Metathesaurus is intended to enable the work with the different vocabularies and provides facilities for the information exchange between different medical databases. The scope of it is as wide as the underlying used source vocabularies. A software program can receive information from Metathesaurus, users can pose inquiries and they can obtain help for a conversion of a vocabulary to the existing uniform vocabularies.
2. **The Semantic Network**: Semantic networks are generally the representation of a given knowledge base [McCray, 1989]. The network contains nodes to represent semantic concepts and links for the representation of relations between the concepts. For example, there are nodes for the representation of organisms, biologic function or chemicals. For UMLS the network is established to facilitate the use of the system. It categorizes the concepts in the UMLS Metathesaurus and offers relations between concepts.
3. **The SPECIALIST Lexicon**: It lexicon provides the lexical information for the SPECIALIST Natural Language Processing System. It can be used as a general English lexicon expanded by a medical vocabulary. The tools were developed to deal with the high variability of words in the natural language. The identification of inflected words (i.e., treat: treats - treated - treating) can be accomplished with these.

The UMLS is an ontology — a system describing instances, concepts, attributes, and relations. But other than common ontologies it has a two-level structure:

1. The Metathesaurus, a unified collection of many different medical terminologies, a compilation of terms, concepts, relationships, and associated information;
2. The Semantic Network containing semantic types (one may think of semantic types as high-level concepts, i.e., broad categories), organized in a hierarchy of IS-A links.

Lee and Geller [Lee and Geller, 2006] call such a structure a Terminological Knowledge Base (TKB):

...any structure that consists of (1) a semantic network of semantic types; (2) a thesaurus of concepts; and (3) assignments of every concept to at least one semantic type

3.2 Evaluation

The accuracy of an implemented algorithm needs to be analysed with approved statistical parameters, in order to allow a comparison with different, already existing algorithms. The value of a statistical parameter is also a statement about the efficiency and precision of an algorithm. High values signify a high quality of the used procedure. Although this seems to be a simple decision strategy for the usage of an algorithm, one has to be careful with the outcome of statistical analyses. Ignorance or the intentional abuse of these statistical parameters can very easily lead to a wrong decision for the choice of an algorithm. In a strict medical context adulterations are used to enable the launching of a new medication which has not been tested carefully enough before.

For the evaluation mainly four statistical parameters are used, which are listed below.

3.2.1 Sensitivity/Recall

The sensitivity is a statistical parameter and very often used in medicine. It provides information about the ability of a test to identify individuals having a special disease (an individual having a given disease is called “positive”, e.g., “HIV-positive”, meaning that the person carries the HI-virus). In other words, the sensitivity evaluates the probability of detecting a statistical unit to be positive, if this unit is really positive. The better the test, the more of the individuals having a given disease are identified.

The formula for this statistical parameter is relatively simple. As sensitivity reflects the accuracy of finding positive test results (i.e., the presence of a disease, the presence of a negation), the number of the found true positive statistical units is divided by all existing positive units in a basic population (when talking of sensitivity in biological circumstances):

$$\text{Sensitivity} = \frac{\sum \text{TruePositives}}{\sum \text{TruePositives} + \sum \text{FalseNegatives}}$$

In the case of information retrieval systems, the sensitivity is replaced by the so-called recall. For the calculation of this statistical parameter, partially correct identifications of entities are used too. Partially correct means, that in a selection process an entity has been correctly identified, but only a part of it. Therefore, the partially correct results contribute half the value of a true positive result [Lehnert et al., 1994]. The formula looks as follows:

$$\text{Recall} = \frac{\sum \text{TruePositives} + \frac{1}{2} * \sum \text{PartiallyCorrects}}{\sum \text{TruePositives} + \sum \text{FalseNegatives} + \sum \text{PartiallyCorrects}}$$

3.2.2 Specificity

This is the counterpart of the sensitivity, measuring the accuracy of finding true negative entities. The specificity is the probability of a test resulting negative, if the examined statistical unit is negative in actual fact. It is calculated by dividing the number of true negative statistical units by all existing negative units in the basic population:

$$\text{Specificity} = \frac{\sum \text{TrueNegatives}}{\sum \text{TrueNegatives} + \sum \text{FalsePositives}}$$

The existence of an HI-virus is tested with the so called ELISA test (Enzyme-linked Immunosorbent Assay). This test comes with both sensitivity and specificity about 99.5 %, other sources speak of 99.9 %. Such high values for this two statistical parameters indicate, that the test is of an excellent quality.

3.2.3 Positive Predictive Value (PPV)/Precision

The positive predictive value provides information about the probability of having a given disease, when a test result is positive. A medical test should have a high PPV, otherwise a patient or his examining physician could never be sure, if the searched disease is present in fact.

For the calculation of this value a simple formula can be used. It is the number of true positive results divided by all positive result (false positives are included):

$$\text{PPV} = \frac{\sum \text{TruePositives}}{\sum \text{TruePositives} + \sum \text{FalsePositives}}$$

The positive predictive value for the ELISA test is 91 %. This means, that of 100 persons with a positive test result, in fact nine of them do not carry the HI-virus, so they are mistakenly considered to be positive.

In the area of information retrieval the positive predictive value corresponds to the so-called precision. Here, partially correct results also get only half the value of a true positive [Lehnert et al., 1994]. The precision can be calculated as follows:

$$\text{Precision} = \frac{\sum \text{TruePositives} + \frac{1}{2} * \sum \text{PartiallyCorrects}}{\sum \text{TruePositives} + \sum \text{FalsePositives} + \sum \text{PartiallyCorrects}}$$

3.2.4 Negative Predictive Value (NPV)

This value is the counterpart of the positive predictive value. It gives information on the probability of the lack of a given feature, if a test result is negative.

The formula is similar to the formula of the PPV:

$$\text{NPV} = \frac{\sum \text{TrueNegatives}}{\sum \text{TrueNegatives} + \sum \text{FalseNegatives}}$$

The ELISA test has a negative predictive value of 99.999 %.

3.2.5 Correlation between Statistical Parameters

Sensitivity and specificity are strongly related: the knowledge of only one of these two statistical parameters is not a very meaningful measurement of the accuracy of a testing procedure. For example, the sensitivity can easily be pushed to 100 % by regarding each and every test result as positive. Only when combined with the specificity (which would in this case normally be surprisingly low) the sensitivity becomes a suitable measurement.

Both PPV and NPV are dependent on the so called prevalence, which is either the total number of occurring positive objects in a statistical universe or the ratio of the number of occurring positive objects at a specific time and the total number of objects in that universe at that time. When the prevalence is high it is easy to gain a good PPV, vice versa it is easy to obtain a good NPV if the prevalence is low.

3.3 A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries - NegEx

This study carried out by the working group around Wendy Chapman proceeded on the assumption that a relatively simple detection algorithm can produce usable results [Chapman et al., 2001]. They designed an algorithm called NegEx and applied it to discharge summaries to find

out, if findings and diseases were negated by the physicians. The found results were compared to the results of a former used baseline algorithm.

3.3.1 Preprocessing of Selected Documents

The input of both algorithms was a preprocessed text sequence, the output of the both is the decision if a phrase is negated in a sentence.

Preprocessing was done in three steps:

1. **Only one sentence per line:** this leads to an information loss when more then one sentence is negated by one and the same negation.
2. **Removal of punctuation:** syntactic information represented by punctuation was extracted.
3. **Indexing of findings and diseases:** in each sentence medical terms were replaced by their UMLS-representation. Because of the large extent of UMLS the indexed terms were delimited to the scope of “Finding”, “Disease or Syndrome” and “Mental or Behavioral Dysfunction”.

3.3.2 Description of the Algorithms

For the realization of the study two algorithms were used. A simple baseline algorithm was compared with the more sophisticated NegEx algorithm.

The baseline algorithm was an algorithm used by the so called IPS system to identify patient subgroups and was created by the University of Pittsburgh. It searches for six negation phrases and marks all UMLS terms from the negation phrase to the end of the sentence.

NegEx was designed on the basis of this simple algorithm but was extended by 29 negation phrases to a total amount of 35 detectable negations. The phrases were searched by manually reading 1050 reports. In comparison with the baseline algorithm NegEx uses a more complex way of extraction. While the baseline algorithm negates everything from the occurrence of the negation until the end of the sentence, NegEx differs between two basic negation types, where one group comprises pseudo-negations which only change the meaning of a term (e.g., “gram-negative”) or which are double negatives (e.g., “not ruled out”). The second group covers the real negations and these are extracted by NegEx. In this group a differentiation between the position of the negation phrase in relation to the negated phrase is carried out. In the first type the negation phrase is positioned before the UMLS term:

<negation phrase> * <UMLS term>

In the second type the negation phrase is positioned after the UMLS term:

<UMLS term> * <negation phrase>

The asterisk stands in this context for up to five tokens which can be either words or UMLS terms. With this method negated UMLS terms can be detected even if they do not stand next to the negation phrase. Moreover, a more distinct extraction than with the baseline algorithm is possible.

3.3.3 Performance Analysis

A performance analysis was carried out which came to the result, that both algorithms marked negation with practicable precision. The baseline algorithm had its strength in sensitivity and negative predictive value, NegEx performed better in specificity and negative predictive value. Table 3.1 shows the detailed results of each algorithm:

Table 3.1: Comparison of the performance of the baseline algorithm and NegEx.

	Baseline algorithm (%)	NegEx (%)
Sensitivity	88.27	77.84
Specificity	85.27	94.51
PPV	68.42	84.49
NPV	93.01	91.73

These results follow from the different designs of the two algorithms. The baseline algorithm, extracting everything from a found negation to the end of the sentence were the negation was found, results in a high sensitivity, but this also leads to an extraction of words which are not really negated. This fact lowers the positive predictive value.

NegEx with its higher precision results in a very high specificity, however, the disadvantage of its design is, that words, which are actually negated, are not extracted because of their spatial distance to the negation phrase. This loss in sensitivity could be removed by a widening of the scope, within words are extracted, yet would this lead to a loss of specificity.

3.3.4 Applications of NegEx

In [Meystre and Haug, 2005] an implementation of NegEx serves for the identification of negations too. In this work a system was created to allow automated processing of problem lists in medical reports. As they are important components of medical records they need to be as accurate and up to date as possible. In reality, the hand written problem lists are often incomplete and therefore there is a need for simple processing routines, so that the problem lists gain in accuracy.

In order to identify medical concepts a mapping of the free text to the UMLS by MetaMap [Aronson, 2001], an element of the SPECIALIST lexicon, is carried out. As MetaMap does not support negation detection a following implementation of a detection algorithm was required and the authors chose the NegEx algorithm. The authors have not carried out an efficiency analysis of their implementation of NegEx but relied on the values given in the study by Chapman and colleagues [Chapman et al., 2001].

Another work [Mitchell et al., 2004] uses the NegEx-algorithm for the detection of negations in pathology reports. In a first step, a number of pathology was randomly selected. Four pathologists read through these reports and assigned negated concepts. Discrepancies between the pathologists were mostly eliminated by a comparison between the negated concepts. The reports assigned by the pathologists were used as a gold-standard for the evaluation of the automated system.

The system works in three steps: the pathology reports are at first split up into the document sections “Gross Description”, “Final Diagnosis”, “Microscopic Description”, “Comment” and “Other”. Afterwards, concepts are annotated using a set of UMLS semantic types. This set was chosen to fit the needs of pathology reports. In a last step, negated concepts are identified

using the algorithms of NegEx. For the evaluation the author used recall and precision, both in lenient and strict version.

Although NegEx was originally not designed for a usage in pathology reports it performed quite well. Table 3.2 shows the results of the evaluation:

Table 3.2: Recall and Precision of NegEx in Pathology Reports

Reports section	Precision			Recall		
	Lenient	Average	Strict	Lenient	Average	Strict
Gross Description	0.87	0.68	0.49	0.57	0.45	0.32
Final Diagnosis	0.88	0.84	0.80	0.84	0.80	0.76
Microscopic Description	0.29	0.19	0.10	0.55	0.37	0.19
Comment	0.63	0.50	0.37	0.42	0.34	0.25
Other	0.068	0.047	0.027	0.074	0.052	0.029

The table shows, that NegEx performed worst in the case of the Microscopic description section. This is due to the fact, that this section is generally written in a more narrative way, whereas the Final diagnosis is written in a more restricted way. This circumstance leads to the good values within this section.

3.4 Use of Methods for the Parsing Process of Formal Computer Languages - NegFinder

In [Mutalik et al., 2001] a different approach to negation detection was chosen. Here, a lexical scanner and a parser were applied, which are methods used to develop tools for the processing of (formal) computer languages. This procedure resulted in considerable values of sensitivity and specificity.

To achieve these results, a system of several components was used like a pipeline. Each component represents a step to the final result, which is the accentuation of negated phrases in the processed text.

3.4.1 The Components Pipeline

The interesting text file undergoes four steps of operation. The result is a modified version of the original with all negations marked up, so the found negations are easier to be verified by a human reader.

The pipeline consists of the following components:

1. **Concept finding:** This step is carried out to find UMLS concepts in the document. If such a concept is found, some information about it is stored in the output. This information consists of the phrase that matches the concept, its length, the position in the text, and the UMLS concept.
2. **Input transformation:** In the second step the information of the first step is combined with the original document. Each phrase that represents a UMLS concept is replaced by its UMLS concept ID.
3. **Lexing/Parsing:** The output of the former steps is passed to a lexical scanner (lexer). This lexical scanner contains knowledge of a large amount of possible negation signals.

Moreover, the lexer is able to classify found negation. It is able to decide, if a given negation signal normally precedes or succeeds its negated phrase, or if it generally negates more than one concept. After this, the parser receives the classified sentences as single tokens. It applies its grammar rules to connect the negation signals with the phrases they negate. When this step is finished, a document is available that contains the information about found negations.

4. **Verification:** The original document is marked up using different colours for different types of negation, the negation signal themselves are marked with a special colour. Negation signals get the color blue, a negated concept is marked in red, combined concepts can be seen in orange and text phrases not representing medical concepts are highlighted in magenta.

3.4.2 Concept Finding

The algorithm used for the identification of concepts tries to match phrases up to five words. If this is not possible, it splits the phrases into subsections and performs another matching attempt.

For the purpose of concept matching the authors use the IBM Intelligent Miner for Text, which allows linguistic analysis and is able to identify special terms like names or abbreviations¹. The medical vocabulary used comes from the UMLS Metathesaurus, which is stored on a MS SQL Server.

3.4.3 Input Transformation

In this processing step recognised medical concepts are replaced by their UMLS-IDs. The output of this step is a file containing numbers and textual sequences, very difficult to read for a human except he/she knows the concepts associated with the ID. This output file is then passed to the lexical scanner.

3.4.4 The Lexical Scanner

The lexer is used for the identification of occurring negation signals. Moreover, it is able to determine the scope of the affected terms. For the purpose of finding negation signals more than 60 words or patterns are recorded, which enable the lexer to mark up negations.

Thanks to its implementation the lexer is able to answer following questions considering a negation:

1. Is the negation signal a preceding or succeeding one? The word “no” normally negates terms following it, whereas in the case of “not present” the negated terms stands before the negation signal.
2. Is only one concept or more than one concept negated? “No”, for example, is able to negate more than one following term.
3. The lexer is also able to decide in negations of more than one concept, if the negation signal normally precedes or succeeds the negated phrases. For this purpose, it uses a difference between the conjunction operators “or” and “and”. “Or” normally indicates a preceding negation signal (“no murmurs, rubs or gallops”), whereas “and” signifies a succeeding signal (“murmurs, rubs and gallops are absent”).

¹<http://www.searchtools.com/index.html>, accessed 18.09.06

Above this, the lexer is also able to tell the parser about the scope of a negation. For this purpose it uses a vocabulary of terminating terms. These terms indicate, that the end of the negated phrases is reached and that a new part of the sentence begins. For example, many prepositions terminate a negated sequence as a general rule. In the sentence “no complications during surgery” only the concept of complications is negated. Another example for terminating terms are relative pronouns, such as “which” or “that”. In the note “There was absence of temperature sense which supports spinothalamic involvement” the negation of the concept “temperature sense” just indicates the occurrence of “spinothalamic involvement”.

When the lexer has finished its decision finding process, it passes the indexed phrases as single tokens to the parser.

3.4.5 The Parser

The parser used in this study works on the tokens it receives from the lexical scanner. Information about negation terminators or the scope in which a negation signal negates is provided by the lexical scanner. Because there is no need for an exhaustive analysis of the syntax, the grammar used by the parser is much simpler than normally used for the processing of natural language.

The actual job of the parser is to finally decide about the characteristics of a negation. It decides, where a negation starts or ends, how many words are concerned by the negation signal and if the negated phrases precedes or succeeds the negation operator. The accomplishment of this job is strongly dependent on the information it receives from the lexer. If there is no information about negation terminators, it considers the end of the negation three words after the negation operator. The number of exactly three words for the scope of a negation was given by an analysis of the sensitivity and specificity reached with a scope of more, or less, than three words.

3.4.6 Evaluation

The evaluation of the algorithm was splitted into two parts. In one part, the designers of NegFinder validated marked up documents, in the other part an independent observer was engaged to do this.

In the first of these two mentioned parts, the NegFinder was applied to a test set of 60 medical documents. The output of this procedure was then validated by the three authors. For this, they read through the documents, in which the negations were already marked up. They counted as well negations the NegFinder had missed (false-negatives) as negations the algorithms had marked by mistake (false-positives). Moreover, information about incorrectly marked negations was collected.

For the other part of the evaluation an independent human observer read through ten medical documents and tried to allocate negations exhaustively. After this, the NegFinder was applied to this set and the results were compared.

To sum it up it can be said, that in the first part already marked documents were controlled by human observers, in the second part the human observer was controlled by Negfinder. This led to following values for the sensitivity and specificity of the algorithm (see Table 3.3):

A slight difference in the specificity of the two evaluation procedures can be noticed. This difference, statistically not significant, shows, that reading through already marked up documents creates a bias distorting the true result. On the other hand the difference is so small, that the two evaluation procedures can be treated equivalent.

Table 3.3: Differences in sensitivity and specificity in the two different examining methods.

	Human observers read through documents marked up with NegFinder (%)	Human observer and NegFinder worked independently (%)
Sensitivity	95.3	95.7
Specificity	97.7	91.8

3.4.7 Applications of NegFinder

The NegFinder algorithm was applied in the work about the knowledge-based MediClass system, which supports the processing of both free text and coded data in electronic medical records (EMR) [Hazlehurst et al., 2005]. MediClass (for “medical classifier”) can operate on any EMR which can use the Clinical Document Architecture (CDA) for data representation. CDA is a document structure based on XML to store and transfer clinical documents, which was developed by the Health Layer Seven health care standards organization (the seven stands for the seventh layer of the ISO/OSI Basic Reference Model).

In order to recognize the change of the semantic meaning of concepts in EMR modifiers are searched for which can influence the meaning of a concept regarding it’s severity, classification and negation. For negation detection, a small set of negation terms is used for the processing with NegFinder’s lexer and parser, since the authors believe that even a little number of detected negation signals lead to reasonable results.

Unfortunately, the authors do not provide statistical values for the reached accuracy of their implementation of NegFinder.

3.5 A Controlled Trial of Automated Classification of Negation from Clinical Notes

The aim of this work was to draw a comparison between the precision of an automated negation detection procedure and a human observer [Elkin et al., 2005]. In the study 41 clinical documents were parsed with the Mayo Vocabulary Server Parsing Engine, which uses the SNOMED CT[®] terminology to discover medical concepts.

For the analysis of this procedure an independent human observer was employed, who read through the resulting records. Incorrect negations were detected to find out possibilities of improvement of the algorithm.

3.5.1 Method

A set of 41 medical reports was used to evaluate the accuracy of an automated negation detection system. These reports were provided to the Mayo Vocabulary Server as ASCII files containing data in free text. With the Mayo Health Record Parser these text files were splitted into subsections typical for the reports, such as “History”, “Physical Examination”, “Diagnostic Testing”.

This output was then preprocessed to differ between textual fragments and operators like “And”, “Or”, “Maybe” and so on. Medical concepts in the text were then indexed using SNOMED CT[®]. After this procedure, the actual negation assignment takes place. For this

purpose, a so called “automated negation assignment grammar” is used. This grammar splits free text phrases into four semantic components:

1. **Kernel concepts:** these contain the semantic basis of the phrase
2. **Modifiers:** Modifiers change the meaning of kernel concepts in a medical way, e.g. ”“severity” expresses the level or stage of a disease.
3. **Qualifiers:** Here a qualifier changes the meaning of a term regarding its temporal or administrative content like “recurrent”.
4. **Negation qualifiers:** these are signals for the occurrence of a negation.

Moreover, a collection of negation stopping phrases is used. These phrases signal the end of a negation’s scope. For example, in the phrase “The patient denied a history of previous cardiac disease other than palpitations which he experienced while giving representation resulting in syncope”, other then limits the scope of the negation signal “denied” to “a history of previous cardiac disease”.

3.5.2 Statistical Analysis

To evaluate the quality of this automated negation detection system, a human reviewer had to read through the reports. By the evaluation of the human reader the indexing of SNOMED-CT[®] turned out to be error-prone. Out of actually 2028 negative concepts, SNOMED-CT[®] had missed 205 concepts, so these concepts were not passed to the negation assignment module.

All in all, the system provides very good values in accuracy. Table 3.4 shows the outcome of the statistical analysis:

Table 3.4: The procedure proposed by [Elkin et al., 2005] reaches high values in all statistical parameters.

	Value of the statistical parameter (%)
Sensitivity	97.2
Specificity	98.8
PPV	91.2
NPV	99.6

3.6 Negation Processing in Electronic Health Records in Bulgarian

This study comes from the university of Sofia and deals with the problem of processing a language different from English, namely Bulgarian [Boycheva et al., 2005]. For the English language the strongest effort in Natural Language Processing and negation detection has been made, whereas other languages have been treated more like stepchildren. The University of Sofia worked on a system called MEHR (Maintaining Electronically Health Records) for the indexing of medical terms and facts like negations in electronic health records, written in Bulgarian.

In the process of indexing medical terms and tagging negations, an electronic health record (EHR) undergoes five separate steps which are worked off in a serial pipeline. The output of

this procedure is a patient's chronicle containing position and scope of found negations and the medical information of this patient.

3.6.1 System Architecture

As mentioned earlier, the system contains of five separate modules, each adding more information to the processed file. Before an EHR is passed to MEHR, it is splitted in the eleven topics typical for bulgarian health records, which are: personal data, anamnesis, status, examinations, consultations, debate, treatment, treatment results, recommendations, working abilities, diagnosis. After this procedure MEHR begins with the analysis of the received file in five steps:

1. **Annotation and chunking:** The A&C module begins with tagging syntactical structures in the received file. It is programmed in Perl and works like a part-of-speech tagger. This module contains a lexicon with about 50 000 words and contains medical vocabulary and terms typical for EHRs.
2. **Post processing module:** The post processing module uses a lexicon with verbs specific for medicine. It also works with verb frames and grammar rules to identify verbal chunks.
3. **Negation treatment module:** In this step the negation detection begins and found negations are marked. A detailed description of the working method of this module is given in the following subsection.
4. **Extractor:** This module uses a database with medicine-specific terms to describe the patient's symptoms and diagnosis.
5. **Filling scenario templates module:** To obtain a correct patient's chronicle this module fills in obligatory and optional fields of a form mask. It ultimately creates the patient's chronicle.

3.6.2 Details of Negation Identification

The authors of the study differ between two kinds of negation, which are common in the bulgarian language: general negations, where a whole statement is negated, and partial negations, in which only one or more parts of a statement are negated.

A general negation in Bulgarian is often indicated by the word "не" (not). With this negation signal an action signalled by a verb is negated, and general negation means, that a verb is negated. Added negation signals like "ни" (nor) or "нищо" (neither) are used to repeat the negation or to increase its strength. However, the use of these two negation signals turns the general negation into a partial type. For example, in the sentence "He does not drink." the whole concept of drinking is negated, whereas in the sentence "He does not drink neither wine nor whiskey" the two drinks are negated seperately. Another signal for a negation a verb action, i.e. a general negation, is in Bulgarian the preposition "без" (without) followed by the concerned verb succeeded by the particle "да". MEHR treats both variants in the same way.

In a partial negation a phrase or even an attribute of a verb, but not the verb itself is negated. The preposition "без" is also used in this context, but this time there is no succeeding "да" provided. A single attribute is negated, e.g. "дишане без хрипове" (breathing without crepitating). Another case are the so called inherent negatives. These are verbs having a negative meaning but are presented in a positive form. Examples for them are:

- “няма” (there is not, not exist) as the negation of “има” (there is, exist). The positive verb differs from the negating verb just by the letter “я”, so they are different words, whereas the English only uses “not” in this case to negate an existence.
- “липсвам” (absent) as the negation of “присъствам” (be available).
- “отричам” (deny) as the negation of “признавам” (confess).

The authors also direct their attention to sentences with a more complex structure. Combined elements which then are negated by only one negation signal need a special treatment, i.e. “без хирзутизм аменореа” (without hirsutism, amenorea). The same applies for sentences, in which more verbs are negated. In Bulgarian for this purpose a repetition of “не” (not) is used, so that the number of “не” and the number of the negated verbs are the same.

3.6.3 Evaluation

With the proposed procedure an identification rate of negations of about 57 % could be reached. 15 % of the negations could not be identified and 28 % were registered incorrectly. An incorrect match means, that the range of the by a negation signal affected concepts was misunderstood.

3.6.4 Conclusio

As one can see from the description of the five modules, MEHR is more than only a negation recognition system. Although one module is only responsible for the treatment of negations in an EHR, the main goal of the system is more ambitious. The aim is to create a system that automatically is able to produce completely filled chronicles out of electronic health records. Negation detection is just the first step towards this ultimate goal.

3.7 Automatic Mapping Clinical Notes to Medical Terminologies

The study of [Patrick et al., 2006] was carried out to measure the accuracy of medical concept matching in free text medical reports. Additionally, the used algorithm was extended in order to identify negated medical concepts too. The following description of this study focuses primarily on the aspects of negation processing.

The procedure of negation assignment consists of four correlative steps. The input to the systems were clinical notes, which in practice are stored in written free text.

3.7.1 Method

As mentioned before, the processing routine contains four steps, beginning with the preprocessing of the input record:

1. **Preprocessing the input:** The input record undergoes a normalisation process.
2. **Concept matching with SNOMED-CT®:** Medical concepts are replaced by their concept ID.
3. **Detection of negations:** The identification process is explained in detail in subsection 3.7.3.

4. **Qualification and term composition:** This step is used to assign medical concepts that are combined with qualifiers, which in these circumstances are terms that contain extra information about a concept but which do not alter its sense.

3.7.2 Preprocessing

This step has two main tasks to achieve, which are interesting since they help to simplify the input records for a later negation assignment. At first, the input is converted into a basic format. This means, that the file is split into its separate sentences, all letters are converted to lower case, and differently written phrases are unified (haemocyte vs. hemocyte). The authors of the study call this the normalisation. After this, the record is POS tagged by the GENIA tagger. This part-of-speech tagger is especially designed for the usage in a biomedical scope.

Another important point in this phase of processing is the identification of so called “administrative entities”. These entities include information about the dosages of medication, duration of an inhabitation in hospital etc. Quantitative units like “kilogram” are identified as well.

3.7.3 Negation Detection

In this work the same simple algorithm is used as in [Chapman et al., 2001]. For the detection of the negation signal, the same signal list as used in this study is employed, but unlike [Chapman et al., 2001] here a general distinction between two superior types of negation is carried out. On the one hand there are so called pre-coordinated negated phrases, on the other hand there are phrases explicitly negated by a negation signal. Pre-coordinated are negated phrases which are already stored as concepts in SNOMED-CT[®], like “no headache”, whereas the other type is the classic negation, where a negation term or phrase negates a concept. This second type of negations is treated by an algorithm similar to NegEx, differentiating between phrases that are preceded or succeeded by their negation flag. The scope of a negation signal is also set to five words before or after the signal.

3.7.4 Evaluation

As the algorithm used in this study is the same as in [Chapman et al., 2001], the same statistical values can be applied. Unfortunately, the authors do not deliver any information about an improvement in relation to NegEx thanks to the usage of pre-coordinated phrases.

3.8 Concept Negation in Free Text Components of Vaccine Safety Reports

In the work of [Tolentino et al., 2006] methods of concept matching and integrated negation detection were applied to surveillance systems for the monitoring of adverse events following immunization (AEFI).

AEFI reports contain similar free text components as other clinical narrative reports like radiology reports, physical examinations or discharge summaries. Considering this fact the study was carried out to evaluate results of the application of UMLS mapping and negation detection algorithms in this medical special field.

After the tagging of the AEFI free text records with the UMLS Metathesaurus a rule-based, finite state machine algorithm was applied for the detection of negated concepts. The algorithms negation vocabulary list contained only five negation signal, which were *no*, *neither/nor*, *ruled out*, *denies* and *without*.

Despite the only very little number of used negation signals the algorithm results in rather high values in recall and precision. Recall and precision are parameter used in the special field of information retrieval. They are applied to evaluate the results of a searching process. The recall measures, how many of the relevant documents of a statistical universe could be found by the searching strategy. The precision describes, how many of the found documents were relevant documents and is therefore a measurement for the accuracy of a searching strategy. The detection algorithm used in this study reached a recall of 89 % and a precision of 94 %.

3.9 Ad-Hoc Classification of Electronic Clinical Documents – NegExpander

This work presents an early work on the topic of Negation Detection. The authors mainly occupied themselves with the topic of “Ad-Hoc Classification”, where “a user needs to sort a large number of documents into non-standard categories.” [Aronow and Feng, 1997]

For their work they also conducted research to identify negated findings in radiology reports. They state, that the challenges of an automated processing of medical writings are constituted of the often poor quality of medical data. Medical data is in most cases not explicitly intended for a processing with computers.

For their part of Negation Detection they implemented a system called NegExpander. This system expands the scope of a negation trigger across conjunctive phrases or colons. The sentence

No suspicious masses, suspicious calcifications or secondary signs of malignancy are seen.

is expanded to

No_suspicious_masses, no_suspicious_calcifications or no_secondary_signs_of_malignancy are seen.

Thanks to this expansion, all three phrases affected by the negation trigger “no” can be identified. For the identification of noun phrases a part-of-speech tagger is used.

Due to the fact, that the work was conducted on the topic of ad-hoc classification, the authors do not give explicit numbers of their outcome on Negation Detection. In spite of this, the expansion to determine the scope of a negation trigger is interesting.

3.10 A Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports

In this hybrid approach a regular expression matching is combined with a grammatic parsing [Huang and Lowe, 2007]. The grammatic parsing is carried out by the Stanford parser [Klein and Manning, 2003], which is capable of generating a full parse tree of sentences. The information derived from this parsing tree is used to initially identify possible negated phrases in the sentence, identifying negation triggers by a regular expression matcher. Afterwards, the actual negated phrase is determined by using grammar rules typical for the specific negation trigger.

In order to achieve the grammar rules, a set of 30 radiology reports was used. The quality of the grammar was then tested with another set of 470 reports. With this information, the negation detection module (NDM) was implemented. The performance of the NDM was again tested on a set of 132 reports.

The authors did not constrain to UMLS concepts as negated phrases. They marked up biomedical noun phrases, regardless of whether they were UMLS concepts or not.

The evaluation of the hybrid approach was carried out with the help of four physicians and the test set of 132 reports. The physicians were put into (changing) groups of two and obtained at a time 30 reports, which were already processed by the NDM. Twelve of these reports were used to evaluate a bias due to the usage of pre-tagged reports. The left 120 reports were used to establish the gold-standard, against which the NDM was tested.

As statistical parameters the authors used slightly different version of recall and precision as presented in Sections 3.2.1 and 3.2.3, a strict and a lenient version. In the strict version, partially correct findings are not added to the true positives, whereas in the case of the lenient version they are added. In the following the used formulas are shown:

$$\text{Strict Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives} + \frac{1}{2} * \text{PartiallyCorrects}}$$

$$\text{Strict Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives} + \frac{1}{2} * \text{PartiallyCorrects}}$$

$$\text{Lenient Precision} = \frac{\text{TruePositives} + \frac{1}{2} * \text{PartiallyCorrects}}{\text{TruePositives} + \text{FalsePositives} + \frac{1}{2} * \text{PartiallyCorrects}}$$

$$\text{Lenient Recall} = \frac{\text{TruePositives} + \frac{1}{2} * \text{PartiallyCorrects}}{\text{TruePositives} + \text{FalseNegatives} + \frac{1}{2} * \text{PartiallyCorrects}}$$

Moreover, the average value of each parameter in its lenient and strict version was calculated as follows:

$$\text{Average Precision} = \frac{\text{Strict Precision} + \text{Lenient Precision}}{2}$$

$$\text{Average Recall} = \frac{\text{Strict Recall} + \text{Lenient Recall}}{2}$$

Table 3.5 shows the results of the evaluation. As it can be clearly seen, the hybrid approach performed very well in the scope of radiology reports.

Table 3.5: Performance of the Hybrid Approach on 120 Radiology Reports

	Specificity (%)	Sensitivity (%)	PPV (%)
CT (computed tomography)	100	91.9	100
Mammogramm	99.7	97.9	97.9
MR (magnetic resonance imaging)	100	92.7	100
PROC (radiology procedure)	99.9	90.5	97.4
RAD (radiograph)	99.5	77.3	94.4
US (ultrasound)	99.7	96.3	98.1
All	99.8	92.6	98.6

3.11 Discussion

In the previous sections we presented approaches related to the topic of Negation Detection in the medical scope. All of them used different algorithms, which performed quite well in practice. They had in common, that they all used a rather simple type of clinical writing for which the algorithms were applied. Discharge summaries (see Section 3.3), electronic health records (Section 3.6), or pathology reports (Section 3.10) are all written in a restricted medical language. This fact simplifies the problem of a Negation Detection.

In clinical practice guidelines, the used language is more complex. They are written in a prosaic style, with numerous nested sentences and a more sophisticated, speech-like language. Due to this, a Negation Detection needs to take care of this problems.

Although CPGs are more complicated than medical reports, they are not as complicated as real free text, as it is used in novels or poems. They will not use stylistic elements like metaphors, alliterations, or double negations. Such elements would complicate the identification of negations even more, since they do not require the adherence to strict linguistic patterns. On the contrary, they are often used to intentionally complicate sentences to provide additional possibilities of expression. The management of such would require algorithms, which really can understand the human language.

In the following section we describe our method of Negation Detection in CPGs, called *NegHunter*. It uses grammatical characteristics of the English language to search for negations.

Chapter 4

NegHunter

In the following we describe the way of developing our Negation Detection algorithms. We took a set of four CPGs (cp. Table 1.1) in order to get an overview about different kinds of negations. This analysis showed that the occurring trigger phrases strongly vary from the triggers used in related work. Here, triggers are composed of several terms, which represent the trigger as a whole. For example, in the work of [Chapman et al., 2001] triggers such as “sufficient to rule the patient out against” are used. Triggers can either affect a phrase which precedes or succeeds them.

In CPGs the situation is different. Here, a combination of terms to triggers is not useful. Due to a more complex language style in CPGs, such a way of proceeding would lead to a vast amount of trigger phrases. A classification of negations leading to a smaller number of triggers is required. Thus, we came to the decision to use the English syntax for a classification. The syntax provides possibilities to decide, which phrase in a sentence is affected by a trigger. This happens by the usage of tenses and prepositions. Thus we need a method to automatically analyze sentences on a syntactical level. We use the MetaMap Transfer (MMTx) in order to acquire a part-of-speech tagging of the sentences in a guideline.

Our method *NegHunter* uses the part-of-speech information to identify negated phrases. In the following, we give a short overview of MMTx. Afterwards, we present the algorithm we developed for Negation Detection in CPGs (see Section 4.2).

4.1 MetaMap Transfer (MMTx)

MMTx was designed to provide Natural Language Processing in combination with a mapping to UMLS terminology. It is built upon the original MetaMap [Aronson, 2001], which was implemented in Prolog, and uses the same algorithms as the original. In order to make the program available for a larger public MMTx uses to the Java language.

The software takes a natural language text as input and allows a hierarchical splitting into the following levels of text parts (see also Figure 4.1):

1. **Section:** Each text is cut into several sections, which are either determined by their textual mark-up (such as abstract, method summary, etc.) or by the text structure, where paragraphs are used to identify the sections.
2. **Sentence:** A sentence is, as one would assume, an arrangement of words with a semantical content.
3. **Phrase:** This is a meaningfully arranged combination of words within a sentence. Phrases also include syntactical elements like articles and adverbs.

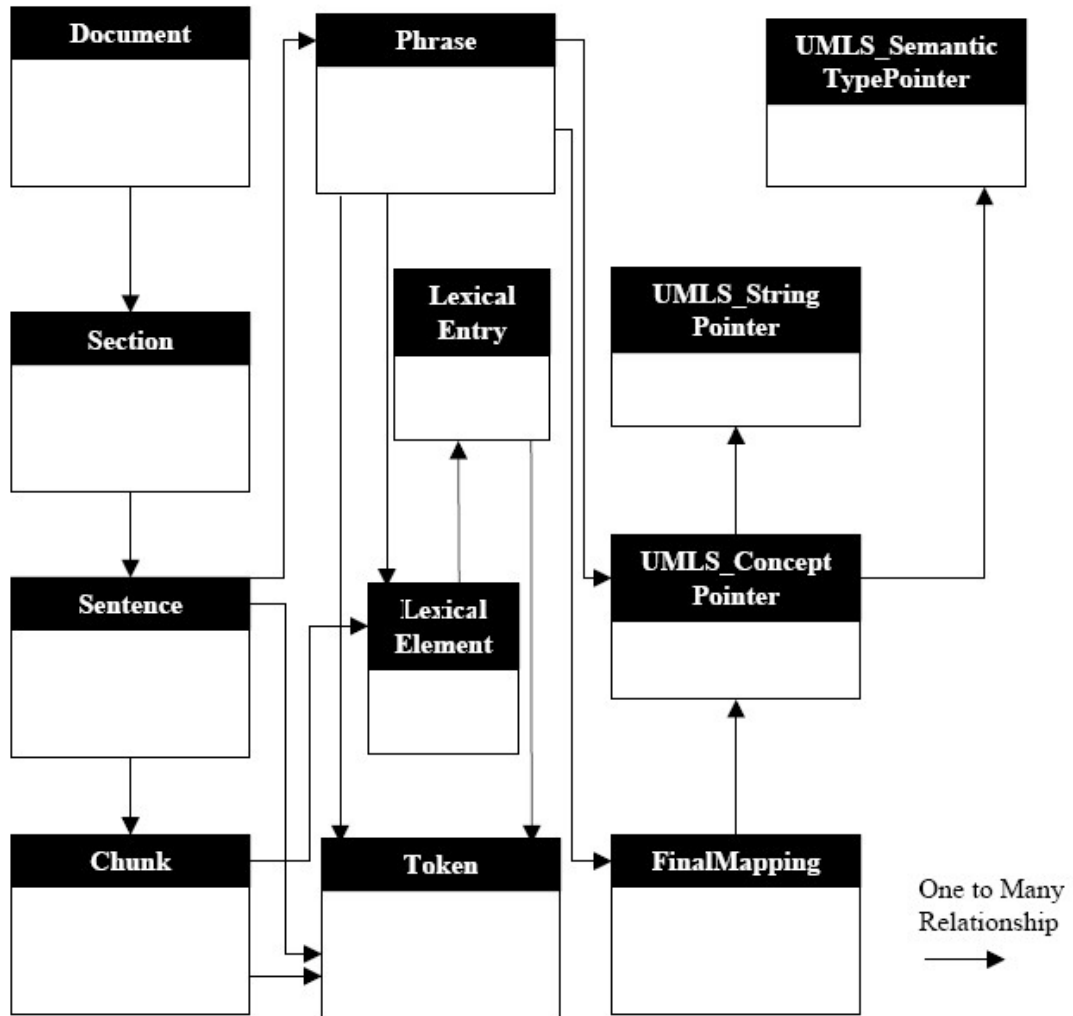


Figure 4.1: Entity-Relationship-diagram of the textfeature package of MMTx [Divita, 2005]

4. **Lexical Element:** Lexical elements can consist of one or more words. “Cancer” is a lexical element just as “breast cancer” is.
5. **Token:** The token is the atomic element of each sentence, determined by a preceding and succeeding whitespace.

For the implementation of *NegHunter*, the classification down to the “phrase”-level was used (see Figure 4.2). The used negation detection algorithms work with the information provided by this level of sentence classification.

The MMTx “phrase” is separated into several types of phrases. These types, differing concerning their tense, part-of-speech type (e.g., a noun phrase or adverb phrase) are used to decide the scope of the trigger, in which negated phrases are expected.

The most relevant “phrase”-classifications available in MMTx are:

- Noun phrase,
- Verb phrase,

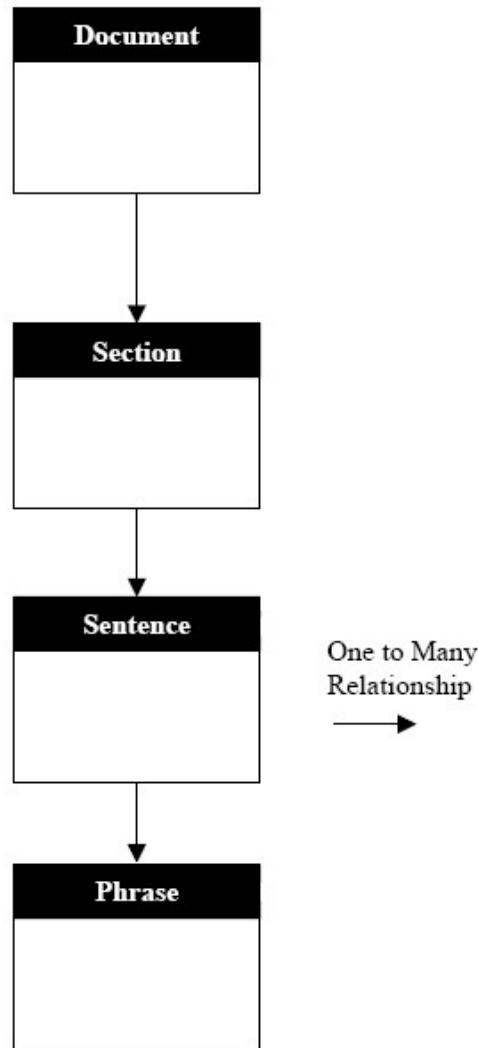


Figure 4.2: Entities of MMTx’s textfeature package used within *NegHunter*

- Adverb phrase,
- Prep phrase, and
- Unknown phrase: A phrase is categorized as an unknown phrase if no other assignment is sufficient.

All of these phrase types are syntactical elements, of which the English language is built upon. A brief summary on these elements will allow to have their real meaning in mind:

Noun phrase. A noun is the attempt to subsume the massive diversification of physical artifacts of the sensually ascertainable environment and of the abstract concepts of the intellectual world with solitary intellectual entities. They are the semantic objects, with which the content of speech is modeled. A physical artifact is everything that one is (in the broadest sense) to touch (like “sunglass” or “ice”. The “sun” is also a physical artifact, although no one would ever be able to touch it). Abstract concepts, however, are everything

that is ascertainable with the mind, like ideas, thoughts or theories (“subconsciousness”, “peace” or “freedom” would be examples for this category). The borders between them can not be strictly drawn. For example, a “person” can either be physically present (e.g., an unknown human in a crowded place) or an abstract representation for the variety of all persons (e.g., the empty database storage for the clients of a company). The true meaning can only be determined within the context of the used noun. In a noun phrase, the noun itself is the so called “head” of the phrase. Although a noun phrase can consist of only the noun, this is a rather infrequent case. Normally, they occur together with articles and specifying adjectives.

Verb phrase. The linguistic element of verbs puts all objects into a specific relationship to their environment. They occur in different tenses, are affected by adverbs and can put a sentence in the active or passive voice. They express the state of objects (e.g., “the man **is** old”), their actions with other objects (e.g., “the player **hits** the ball) or processes on themselves (e.g., “the tree **grows**”).

Adverb phrase. Adverbs are normally used to change or intensify the meaning of verbs, adjectives or other adverbs. MMTx regards the term “not” as an adverb phrase.

Prepositional phrases. Prepositional phrases consist of noun phrases which include a preposition. Prepositions add auxiliary information like place, time, or direction declarations to the original noun phrase. Initializing triggers, which indicate such a phrase can be:

- **Place:** “in” or “on”, for example:

*The guideline developers met **in** Cincinnati.*

- **Time:** “in” or “on”, for example:

*The guideline was reviewed **in** the year 2005.*

- **Direction:** “to” or “into”, for example:

*The patient was pushed **into** the CT.*

- **Other:** “by” or “of”, for example:

*The treatment is recommended **by** leading oncologists.*

4.2 The Negation Detection Algorithms

Negation Detection in CPGs is a great challenge due to the complicated language structure of this type of medical document. Contrary to medical reports, such as discharge summaries, radiology reports etc., they use a sophisticated, speech-like language. This leads to the occurrence of negations that cannot be determined by a certain number of negation triggers. In reports these triggers, which are several terms combined to one trigger phrase (“sufficient to rule the patient out against”, [Chapman et al., 2001]), are repeated continuously. A search for such constructed triggers in CPGs is not as easy. One would have to create a large collection of trigger phrases with only little possibility of classifying into manageable types. Moreover, every new guideline would contain new trigger phrases, making a permanent enlargement of the collection of trigger phrases obligatory. Such a method would be very cumbersome and time-consuming. Thus, a different approach was used. As language itself provides information on which phrases are negated in a sentence via the syntax (or the grammar of the language), we decided to use this information for our Negation Detection. This approach is also similar to the way, the human

brain processes negations. Humans are able to decode the semantic content of a phrase and can decide, if this specific phrase is a negation trigger. If it is, the brain decides using the syntax which phrase in the sentence is negated.

Our approach is undoubtedly not as efficient as the brain's methods. Yet it uses methods of language processing which help us to identify negated phrases in natural language writings, where a simple subsumption of terms to triggers is not possible due to the complexity of the underlying text.

Speaking of the usage of the syntax of the language, we mean that grammatical elements like the tenses influence the direction, in which a trigger term works. It can either affect phrases that have been mentioned earlier in a sentence (a succeeding trigger respectively a preceding negated phrase) or phrases that follow after the trigger (a preceding trigger respectively a succeeding negated phrase). We have come up to five classes in which the negations are split up:

The Adverbial Negation: This type of negation works via a verb, which is affected by a negation term. The tense helps to decide, where in the sentence the concerning phrase is allocated.

The Intra-Phrase Triggered Negation: Here, the negation trigger is integrated in the phrase itself. The phrase can then be regarded as negated.

The Prepositional Negation: The trigger is a phrase which requires a succeeding prepositional phrase. The negated phrase is this prepositional phrase.

The Adjective Negation: Triggers are adjectives with capability of negating phrases in a sentence.

The Verb Negation: This negation class is triggered by verbs.

In the following sections we describe, how our approach works and which pieces of information of the grammar are used to support the decision process.

4.2.1 The Adverbial Negation

This first type of negation is a very common one. An adverb is the actual negation trigger. Its occurrence means, that the verb affected by this adverb negates a noun phrase, which is related to the verb. Verbs work on subjects or objects in a sentence. The task is to correctly identify this subject or object. That is, an adverb negates a noun phrase via a verb. The verb itself is not the negated term, since a negation of a verb alone would not make sense. For example, "not existing" says, that something does not exist. The important question is to identify the thing that is not existent. Considering verbs as negated would on the other hand be a simple and trivial task, since verbs always stand closely to the adverb working as trigger phrase. So, a proper method of identifying related noun phrases is crucial. In CPGs, we have the triggers "not" and "never" as adverbial triggers.

Negations using the trigger "not" are the most frequent negations in medical guidelines. Nearly 50 % of all negations occurring in the observed twenty guidelines are triggered by this term. Moreover, it is not only a very frequent trigger, it is also a very complex one. They do not only underlie inflections due to varying tense and number but also to the active and passive voice. This circumstances require a sophisticated treatment of this type of negation.

NegHunter uses the specific appearance of verb structures regarding their tense to decide, whether a sentence contains an active or passive voice context. On this basis it decides, which term is meant to be negated.

In order to clarify the considerations behind this way of proceeding, we provide a brief summary of the active and passive voice.

The active voice: Here, the subject operates via the verb with the object in a specific way. In the example

The radiogram identifies multiple calcifications in the left breast.

the subject “radiogram” comes via “identifies” with “multiple calcifications” into action.

The passive voice: Here, the subject itself is operated by the verb. When turning an originally active form sentence into passive, the former object becomes the subject. For example:

Multiple calcifications in the left breast are identified by the radiogram.

Here, “multiple calcifications” are targeted by the verb structure “are identified” and specified by the prepositional phrase “by the radiogram”.

These considerations allow the modelling of this kind of negations. When in the case of the active voice a subject operates on an object, then, in the negated form, this object is excluded from the operation. The negated object is the noun phrase following the concerning verb phrase. *NegHunter* uses this information to identify the object.

In order to find out, if a verb structure is in active or passive voice, *NegHunter* uses information of the verb combination to decide, which tense the verb structure has. Based on this information we can see, whether the phrase is in active or passive voice.

The following two sentences explain the same matter of fact. Whereas, the first sentence is written in active voice, the second is in passive voice. Both of them use the same tense (Past Tense):

Guideline developers did not perform a formal cost analysis.

A formal cost analysis was not performed by the guideline developers.

In the first case, the structure of the verb form is recognized by MMTx as follows (see Table 4.1:

Table 4.1: A Simple Past Active Sentence and the splitting into its phrase components

Phrase Classification	Verb Phrase	Adverb Phrase	Verb Phrase
Term	<i>did</i>	<i>not</i>	<i>perform</i>

The composition of the verb structure “did not perform” as “verb phrase + not + verb phrase” indicates that the sentence is in active voice. It is unique for the tenses

- Simple Present Active
- Simple Past Active
- Present Perfect Active
- Future I

- Past Perfect
- Conditional I.

In an analog manner, the indicators for the passive voice in the second sentence are defined (see Table 4.2).

Table 4.2: The classification of the phrase in a Simple Past Passive sentence

Phrase Classification	Be Verb	Adverb Phrase	Verb Phrase
Term	<i>was</i>	<i>not</i>	<i>performed</i>

An order like “be verb + not + verb phrase” is typical for following tenses:

- Simple Present Passive
- Simple Past Passive

On the basis of this information, we apply separate rules for the identification of negated phrases. In the active voice, the rule works as follows:

$\langle \text{Adverbial Negation Active Voice} \rangle ::= \langle \text{Active Negation Trigger} \rangle \langle \text{Arbitrary Phrase Thread} \rangle \langle \text{Negated Phrases} \rangle$

For the rule the following definitions are applied:

$\langle \text{Active Negation Trigger} \rangle ::= (\langle \text{Verb Phrase} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Verb Phrase} \rangle)$
 $| (\langle \text{Be Verb} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Verb Phrase -ing} \rangle)$
 $| (\langle \text{Verb Phrase} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Verb Phrase} \rangle \langle \text{Verb Phrase} \rangle)$
 $\langle \text{Adverb Trigger} \rangle ::= \text{“not”} | \text{“never”}$
 $\langle \text{Verb Phrase -ing} \rangle ::= \langle \text{Verb Phrase} \rangle \text{“ing”}$
 $\langle \text{Negated Phrases} \rangle ::= [\langle \text{Noun Phrase} \rangle] [\langle \text{Noun Phrase} \rangle] [\langle \text{Noun Phrase} \rangle]$
 $\langle \text{Arbitrary Phrase Thread} \rangle ::= \langle \text{Arbitrary Phrase except Noun Phrase} \rangle | (\langle \text{Arbitrary Phrase except Noun Phrase} \rangle \langle \text{Arbitrary Phrase Thread} \rangle)$

For examples on these rules see Tables 4.4 to 4.6.

In the case of the passive voice, we apply the following rule:

$\langle \text{Adverbial Negation Passive Voice} \rangle ::= \langle \text{Negated Phrases} \rangle \langle \text{Arbitrary Phrase Thread} \rangle \langle \text{Passive Negation Trigger} \rangle$

The definitions for this rule are the same as for the former rule, except for the negation trigger. Its definition is as follows (see examples in Tables 4.7 to 4.10):

$\langle \text{Passive Negation Trigger} \rangle ::= (\langle \text{Be Verb} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Verb Phrase} \rangle)$
 $| (\langle \text{Verb Phrase} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Be Verb} \rangle \langle \text{Verb Phrase} \rangle)$
 $| (\langle \text{Be Verb} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Be Verb} \rangle \langle \text{Verb Phrase} \rangle)$
 $| (\langle \text{Verb Phrase} \rangle \langle \text{Adverb Trigger} \rangle \langle \text{Verb Phrase} \rangle \langle \text{Be Verb} \rangle \langle \text{Verb Phrase} \rangle)$

The following example shows the result of the tagging process of a sentence in the active voice with *NegHunter*:

Guideline developers do **not** recommend **chemotherapy**.¹

The tagging process for the sentence expressed in passive voice leads to the following result:

Chemotherapy is **not** recommended by guideline developers.

Other examples for a common sentence structure would be the following tenses, each with example sentences:

- Present Perfect Passive: Chemotherapy *has not been recommended*
- Future I Passive: Chemotherapy *will not be recommended*
- Past Perfect Passive: Chemotherapy *had not been recommended*
- Conditional I Passive: Chemotherapy *would not be recommended*

In these cases, the sentence structure is built in the order represented in Table 4.3:

Table 4.3: The classification of a phrase in Present Perfect Passive

Phrase Classification	Verb Phrase	Adverb Phrase	Be Verb	Verb Phrase
Term	<i>has</i>	<i>not</i>	<i>been</i>	<i>recommended</i>

The structure is shown exemplarily for Present Perfect Passive in Table 4.3, but it applies to all tenses formerly mentioned.

The following tables show the sentence structure of all tenses, starting with the common structure of Simple Present, Simple Past, Present Perfect, Future I, Past Perfect and Conditional I, all of them in the active voice. For each tense a short example sentence is provided showing only the verb structure, no noun phrases or other sentence components are included.

Table 4.4: In the active voice the phrase classification is similar for seven tenses.

Phrase Classification	Verb Phrase	Adverb Phrase	Verb Phrase
Simple Present	<i>do</i>	<i>not</i>	<i>recommend</i>
Simple Past	<i>did</i>	<i>not</i>	<i>recommend</i>
Present Perfect	<i>have</i>	<i>not</i>	<i>recommended</i>
Future I	<i>will</i>	<i>not</i>	<i>recommend</i>
Past Perfect	<i>had</i>	<i>not</i>	<i>recommended</i>
Conditional I	<i>would</i>	<i>not</i>	<i>recommend</i>

Table 4.4 shows the collection of the tenses “Simple Present”, “Simple Past”, “Present Perfect”, “Future II”, “Past Perfect” and “Conditional I”. Table 4.5 shows the tenses Present Progressive Active and Past Progressive Active:

Table 4.6 shows the classification of Future II and Conditional II in active voice. These three tables cover all tenses in the active voice. Tables 4.7 to 4.10 show the specific sentence structure for all tenses but this time in the passive voice. Present Perfect Passive, Future I Passive, Past Perfect Passive and Conditional I Passive are shown in Table 4.8. The passive voice structure of Present Progressive and Past Progressive is listed in Table 4.9. Table 4.10 shows the tenses Future II and Conditional II in passive voice.

¹The trigger term is highlighted in yellow, the negated phrase in springgreen.

Table 4.5: The structure of Present and Past Progressive Active.

Phrase Classification	Be Verb	Adverb Phrase	Verb Phrase
Present Progressive	<i>are</i>	<i>not</i>	<i>recommending</i>
Past Progressive	<i>were</i>	<i>not</i>	<i>recommending</i>

Table 4.6: Future II and Conditional II in Active Voice spanning three Verb Phrases with their usage.

Phrase Classification	Verb Phrase	Adverb Phrase	Verb Phrase	Verb Phrase
Future II	<i>will</i>	<i>not</i>	<i>have</i>	<i>recommended</i>
Conditional II	<i>would</i>	<i>not</i>	<i>have</i>	<i>recommended</i>

Table 4.7: Simple Present and Simple Past in the Passive Voice

Phrase Classification	Be Verb	Adverb Phrase	Verb Phrase
Simple Present	<i>is</i>	<i>not</i>	<i>recommended</i>
Simple Past	<i>was</i>	<i>not</i>	<i>recommended</i>

Table 4.8: Present Perfect, Past Perfect, Future I and Conditional I Passive.

Phrase Classification	Verb Phrase	Adverb Phrase	Be Verb	Verb Phrase
Present Perfect	<i>has</i>	<i>not</i>	<i>been</i>	<i>recommended</i>
Past Perfect	<i>had</i>	<i>not</i>	<i>been</i>	<i>recommended</i>
Future I	<i>will</i>	<i>not</i>	<i>be</i>	<i>recommended</i>
Conditional I	<i>would</i>	<i>not</i>	<i>be</i>	<i>recommended</i>

Table 4.9: Present and Past Progressive Passive.

Phrase Classification	Be Verb	Adverb Phrase	Be Verb	Verb Phrase
Present Progressive	<i>is</i>	<i>not</i>	<i>being</i>	<i>recommended</i>
Past Progressive	<i>was</i>	<i>not</i>	<i>being</i>	<i>recommended</i>

Table 4.10: The passive form of Future II and Conditional II.

Phrase Classification	Verb Phrase	Adverb Phrase	Verb Phrase	Be Verb	Verb Phrase
Future II	<i>will</i>	<i>not</i>	<i>have</i>	<i>been</i>	<i>recommended</i>
Conditional II	<i>would</i>	<i>not</i>	<i>have</i>	<i>been</i>	<i>recommended</i>

As one can see in the tables above there are several decisions to be made in order to receive the right tense. After this decision is made the negated concept represented by a noun is tagged. This noun can be placed before or after the negation trigger depending on whether the verb phrase is in active or passive form. In order to fully cover negated concepts we use a range of three noun phrases tagged before or after the verb phrase. We have chosen a number of three noun phrases in order to receive a good compromise between false-positive and false-negative phrases. We believe that a negation covers only a small number of phrases and not a thread of numerous phrases.

The next section deals with negations, where the triggers are already included in noun

phrases.

4.2.2 The Intra-Phrase Triggered Negation

In this kind of negation the negation trigger is included in a phrase. The trigger therefore negates exactly this phrase. For example, the sentence

Evidence obtained from at least one well-designed controlled study without randomisation.

contains the negated concept “randomisation”. In this case, the negation is triggered by the term “without”. MMTx classifies “without randomisation” as a prepositional phrase.

NegHunter has classified four triggers for the Intra-Phrase Triggered Negation:

1. *without*: occurring within a prepositional phrase
2. *no*: indicates negated noun phrases
3. *inappropriate*: within noun phrases
4. *ineffective*: also within noun phrases

The following rules are applied for the detection negated phrases of this type:

<Adjective Negation> ::= <Trigger Phrase> <Noun Phrase>
 <Trigger Phrase> ::= “without” | “no” | “inappropriate” | “ineffective”

For example, in the sentence

There is no clinical evidence that Hormone Replacement Therapy should be withheld from patients with a history of cervical cancer

the concept “clinical evidence” is negated by the term “no”.

NegHunter identifies the negation triggers and afterwards it splits up the phrase allowing a differentiation in negation trigger and negated term. The term which is contained in the phrase with the trigger is then regarded as the negated phrase. In the subsequent sentence the effect of the trigger “no” is shown:

*There is no clinical evidence that Hormone Replacement Therapy should be withheld from patients with a history of cervical cancer.*²

In case of the negation with “without” the result would look similar:

Evidence obtained from at least one well-designed controlled study without randomisation.

All triggers are searched for with and without capital letters in order to also identify negations placed at the beginning of a sentence.

The two triggers “ineffective” and “inappropriate” can also occur singularly in a sentence. Related to the Intra-Phrase Triggered Negation they both are combined with other terms resulting in a noun phrase respectively a prepositional phrase. When this precondition is not given, these two triggers are handled by the rules of the class “Adjective Negation”, explained in 4.2.4.

In the following section we provide information about a negation type, where prepositional phrases are affected by negation triggers.

²The trigger is again in yellow, the negated terms are in green. This kind of highlighting phrases will be continued in the whole subsequent work.

4.2.3 The Prepositional Negation

The prepositional negation is initialized by certain phrases which require a following prepositional phrase. Indeed, this prepositional phrase is the negated concept in the particular sentence. Trigger phrases can either be nouns or adjectives.

For CPGs we use the subsequent terms initializing a prepositional negation:

- *absence*
- *none*
- *lack*
- *free*
- *freedom*

Each of these phrases is followed by a prepositional phrase which is the negated phrase. The following schema represents the rule for this type of negation:

<Prepositional Negation> ::= <Negation Trigger> <Negated Phrase>

We apply the subsequent definitions:

<Negation Trigger> ::= <Noun Trigger> | <Adjective Trigger>

<Noun Trigger> ::= “absence” | “lack” | “freedom”

<Adjective Trigger> ::= “free” | “none”

<Negated Phrase> ::= <Prepositional Phrase>

The sentence

*Performing an anastomosis in this setting is dependent on the patient’s general condition at the time of resection and **the absence** **of other factors** that indicate the need for a stoma to be created.*

shows an example of the usage of the trigger “the absence”. Although it is a prepositional phrase it is considered as a negated phrase³. “The absence of other factor” equals the statement “no other factors”, where “other factors” are the negated phrase (one must keep in mind, that “no other factors” actually is an intra-phrase triggered negation, the comparison should only simplify the understanding). With other verbs, negated phrases which are prepositional phrases too are not treated as prepositional. They are negated phrases and therefore are also tagged as such.

In the case of the trigger “lack” the tagging process is accomplished similarly:

*The chemotherapy offered will depend on the duration of the response after receiving first-line chemotherapy or **the lack** **of response** to first-line chemotherapy.*

The last sentence allows a similar comparison as mentioned in the last example. The statement “the lack of response to first-line chemotherapy” expresses the same fact as the statement “there is no response to first-line chemotherapy”. Because of this, the prepositional phrase is actually a negated phrase. Furthermore, the following example shows the effect of the trigger “none”:

³in Subsection 4.2.6 we describe the handling of prepositions, which represent accessory information that can be added to negated phrases

*Women can be cared for in primary care if the family history shows only one first-degree or second-degree relative diagnosed with breast cancer at older than age 40 years, provided that **none** **of the following** are present in the family history.*

Here, the prepositional phrase “of the following” is effected by the trigger. “None” always expresses the fact, that the semantic concept of the following phrase is not valid. Thus, the phrase after “none” is considered as negated. In CPGs, this trigger term is often used in sentence, where a related negated phrase does not exist. This is then the case, when the actual negated phrase occur in a preceding sentence. For example,

AVAILABILITY OF COMPANION DOCUMENTS. None available

In this case, “AVAILABILITY OF COMPANION DOCUMENTS” is a headline, and in the following paragraph the statement is expressed, that these companion documents are not available. In this case, the trigger term is tagged, but the related phrase is not. We do not provide a cross-sentence tagging, since this would lead to unpredictable results. Note the following sentences:

There are many patients in the hospital. None has cancer.

In the hospital are many patients. None has cancer.

These two sentences are only simple ones, yet the negated phrases occur on different locations within the sentences. When the sentence structure becomes more complex, these differences grow larger. This makes a proper Negation Detection to a task of the second sight. Therefore, *NegHunter* does not provide such tagging strategies.

There are also two triggers, which would not be regarded as triggers by persons not familiar with the medical language. These triggers are “free” and “freedom”. In the normal spoken or written language, the two terms always have a rather positive connotation. To be free of something is always a good thing. This is also the case in medicine, but here, the freedom from something means, that a specific concept could not be observed. The subsequent sentence shows an example:

*The patient is **free** **of pain**.*

To be free of pain is a positive fact for the patient. In the context of medicine this sentence also has the very important semantic of a lack of pain, i.e., there is no pain. This circumstances make the terms “free” and “freedom” to important triggers. “Freedom” works similarly:

***Freedom** **from pain** is an important indicator for the recovery of a patient.*

Here, pain is also negated by “freedom”. The presence of pain would otherwise lead to the need of a treatment of the specific patient. This matter of fact is very important for CPGs.

In this class of negations prepositional information can also be added to a negated phrase. When this case occurs, the prepositional phrase containing the additional information is highlighted in cornflowerblue too, as shown in the following example:

*Indicates **absence** **of directly applicable clinical studies** of good quality.*

The subsequent section deals with negation types, where the negation is initialized by adjective triggers.

4.2.4 The Adjective Negation

English adjectives can also be used as negation triggers. For CPGs we identified the terms “ineffective” and “inappropriate” as such triggers. The precondition of the usage of these two adjectives as triggers is that they are not combined to a noun phrase with other terms (e.g., “Ineffective treatment of patients...”, see Subsection 4.2.2). This means, that the terms need to occur singularly in a sentence. The following decision schema is applied for the adjective negation:

$$\langle \text{Adjective Negation} \rangle ::= \langle \text{Negated Phrase} \rangle \langle \text{Arbitrary Phrase Thread} \rangle \langle \text{Be Verb} \rangle \langle \text{Adjective Trigger} \rangle$$

The elements of this rule are defined as follows:

$$\begin{aligned} \langle \text{Negated Phrase} \rangle &::= \langle \text{Noun Phrase} \rangle \\ \langle \text{Arbitrary Phrase Thread} \rangle &::= \langle \text{Arbitrary Phrase except Noun Phrase} \rangle | (\langle \text{Arbitrary Phrase except Noun Phrase} \rangle \langle \text{Arbitrary Phrase Thread} \rangle) \\ \langle \text{Adjective Trigger} \rangle &::= \text{“ineffective”} | \text{“inappropriate”} \end{aligned}$$

The adjective trigger is preceded by a verb phrase indicating that this combination negates the status of a phrase. Moreover, a number of arbitrary phrase can be located between this combination and the noun phrase, which is negated. These arbitrary phrases must not be noun phrases themselves. In

*Recommendation indicates at least fair evidence that **the service** is **ineffective** or that harm outweighs benefit.*

the term “ineffective” occurs singularly in the sentence, indicating that the preceding phrase is negated.

In the sentence

***A superficial shave biopsy** is **inappropriate** for suspicious pigmented lesions.*

the trigger also occurs alone in the sentence. So, the phrase “A superficial shave biopsy” is considered as a negated phrase.

The two triggers for the Adjective Negation can also occur as triggers of another class of negation types, as mentioned earlier. When one of these triggers appears within a noun phrase or a prepositional phrase they are handled as an Intra-Phrase Triggered Negation. In this case the rules described in Section 4.2.2 are applied.

4.2.5 The Verb Negation

Negations can also be triggered by verbs. If such a verb occurs in a sentence, the phrase related to this verb is considered negated. According to the suffix we decide, whether a verb represents an expression in the active or passive voice. For example, the suffix “ing” indicates the passive voice (e.g., “A treatment option is lacking”), whereas the suffix “s” indicates the active voice (e.g., “The patient lacks a good health status”). The decision rule for the verb negation is the following:

$$\langle \text{Verb Negation Active Voice} \rangle ::= \langle \text{Active Verb Trigger} \rangle \langle \text{Negated Phrase} \rangle$$

Here, the negation trigger is defined as:

<Active Verb Trigger> ::= “lack” | “lacks” | “lacked” | “deny” | “denies” | “denied”

For the passive voice, the rule is:

<Verb Negation Passive Voice> ::= <Negated Phrase> <Passive Verb Trigger>

The negation trigger is a verb in its passive voice:

<Passive Verb Trigger> ::= “lacking” | “denying”

In both cases, the negated phrase is a noun phrase:

<Negated Phrase> ::= <Noun Phrase>

We identified the two triggers “lack” and “deny” as such verb triggers. The following example sentence shows the occurrence of the trigger “lack” in active voice:

The patient lacks a good health status.

Here, as the sentence is written in active voice, the triggers negateds the succeeding phrase “a good health status”. In the passive voice, a preceding phrase is negated by the triggers:

Recommendation indicates that evidence that the service is effective is lacking.

4.2.6 Assigning Prepositional Information

With prepositional phrases auxiliary information is provided in many cases. From our point of view it is important to label this additional pieces of information, since they deliver a specification of a negated concept. We decided to tag all prepositional phrases which directly follow negated phrases. In the sentence

Women who do not meet the criteria for referral should be cared for in primary care by giving standard written information.⁴

we see a prepositional phrase tagged in blue. In this case, only one preposition followed the negated concept. In opposition to this case, in the following sentence more than one prepositional phrase is added to the concept:

The use of anthracyclines in first line chemotherapy treatment of epithelial ovarian cancer is not recommended outside randomised controlled trials (RCTs)

For such cases we decided to tag all prepositional phrases which directly follow a concept. Once the thread of prepositional phrases is split up by another type of phrase the affinity to the concept is ended up and the tagging process stops too.

The following example shows prepositional information applied to an intra-phrase triggered negation:

⁴Prepositional information is highlighted by a blue box.

*Requires availability of well conducted clinical studies but **no** **randomised clinical trials** on the topic **of recommendation**.*

In this case we have two pieces of additional information added to the negated concept which for that reason are also provided with metadata that signs them as belonging to the negated phrase. The trigger “ineffective” shows a similar behaviour:

***Ineffective** **treatment** **of patients** maximizes the costs.*

In the class of prepositional negations, prepositional information can also be added to the negated phrase. When this case occurs, we discriminate between the negated phrase and the added information. It is important that the prepositional negated phrase and the prepositional information are not confused, since they are different concepts in the sentence. The following sentence, already introduced in Section 4.2.3, shows such a case:

*The chemotherapy offered will depend on the duration of the response after receiving first-line chemotherapy or **the lack** **of response** **to first-line chemotherapy**.*

Here the negated phrase and its added information are treated differently, although both originate from the same lexical category.

Prepositional information represents an essential part of information that must also be declared as belonging to the negated phrase. This kind of information often provides explanations which must not be lost in the Negation Detection process.

The next section gives a summary over our developed approach and resumes our aggregated knowledge.

4.3 Summary

In the preceding section we described our Negation Detection method called *NegHunter*. *NegHunter* uses the MMTx software as a framework for document and sentence analysis. The information gained from the processing of CPGs with MMTx are used for the identification of the underlying negation class. Each identified negation is matched to one of five defined negation categories, which have been developed and named in regard to their syntactical features. This means, that part-of-speech characteristics of the negation triggers themselves as well as their behaviour on negated phrases are used for their naming. For example, the “adverbial negation” uses adverbs, which affect a phrase via a verb and thus negate this phrase. “Adjective negation” as well as “verb negation” can be understood straightforward, as in these types the triggers are either adjectives or verbs. The “intra-phrase triggered negation” owes it’s name to the circumstance that the triggers are contained in the phrase they negate. Yet, triggers can be of several lexical categories, for example, “without” is a preposition, whereas “ineffective” is an adjective. The “prepositional negation” is triggered by terms, which affecta subsequent prepositional phrase, although they again can arise from several lexical categories (e.g. “lack” as a noun triggers a prepositional negation as well as “free”, which is an adjective).

This negation classification allows a targeted detection of negated phrases. This approach is necessary since CPGs are medical documents written in free speech, not restricted to conventions that could be used for a simpler algorithm design. As each individual has an individual writing style, all CPGs vary from from each other. *NegHunter* uses the basis of language itsel, so on this way we are able to overcome the vast possibilities of variations.

We believe thanks to the general structure of our algorithms that a deployment of *NegHunter* on documents different from CPGs can also be considered. *NegHunter* uses dealings of the English language which may also be applicable for other writings.

In the following section we provide the evaluation of our system. In Chapter 6 we resume our research and give a future outline of possible enhancements to our existing system.

Chapter 5

Evaluation

The final part of our work on Negation Detection consists of the evaluation of our approach. With this part we provide information about the performance of the presented algorithms. Moreover, we show the distribution of occurring negated phrases with their underlying classification.

The evaluation was carried out as follows: we initially developed a gold standard of our test set of totally 16 CPGs (see Section 1.2). In a second step we compared this gold standard with the results of the processing with *NegHunter*. Afterwards, we calculated the statistical values of recall and precision, which are generally used to evaluate Natural Language Processing systems (see Sections 3.2.1 and 3.2.3). Finally, we tried to identify errors in the tagging process which emerged due to the MMTx framework and errors related to *NegHunter*.

In the subsequent section we give an overview over the distribution of the phrases we identified as negated in our test set of 16 CPGs.

5.1 Distribution of Identified Negated Phrases

In the 16 CPGs we used to test our negation detection algorithms we found 642 sentences containing totally 731 negated phrases. Of these 731 negated phrases, an overwhelmingly large number belonged to the class of the adverbial negation. This class of negation is not only the most frequent class but also the most complicated to process. It is difficult to identify negations of this type, since the concerning phrases can occur on the complete opposite end of the place of the trigger in the sentence. Other types of negations are easier to treat, since the phrase related to trigger is always expected subsequently to the trigger.

Figure 5.1 shows a pie chart of the distribution of negated phrases. As one can see the most dominant class is the adverbial negation with a total of 69 % of all counted negations. Another very frequent class is the intra-phrase triggered negation which holds one quarter of all phrases. Prepositional negation and verb negation occur in 3 % and 2 %, respectively and the adjective negation as the smallest class amounts to only 1 % of the detected negated phrases.

The vast amount of adverbial negations in CPGs has also a large effect on the outcome. The complexity leads to difficulties in the identification of the actual negated phrases. As these difficulties concern a negation class with a vast amount of occurrences, overall recall and precision are strongly influenced and consequently lowered. Good results of the other classes cannot equalise this effect, because their number is too small. Alternatively, an efficient treatment of this negation type has a positive impact on the overall parameters.

In the following section we give an impression of the performance of *NegHunter* in each negation class. We show, how efficient *NegHunter* identifies negations of the different types.

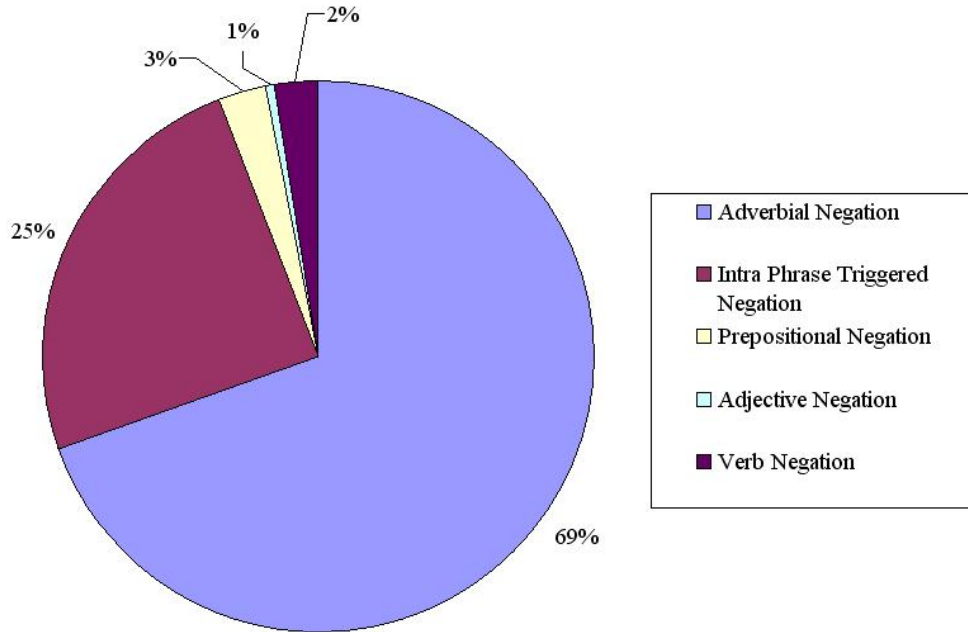


Figure 5.1: Distribution of negated phrases

5.2 Analysing the Performance of NegHunter

The analysis of the performance of *NegHunter* was carried out using the statistical parameters *recall* and *precision* presented in Sections 3.2.1 and 3.2.3. The recall is a measurement for the quality of a method to identify interesting targets. The following formula simplifies this matter of fact:

$$\text{Recall} = \frac{\sum \text{Identified Relevant Targets}}{\sum \text{All Existing Relevant Targets}}$$

On the other hand, the precision measures, how good a system discards elements of no interest. It's value gives information about how many irrelevant targets were (incorrectly) identified. See the subsequent formula for a simplified representation:

$$\text{Precision} = \frac{\sum \text{Identified Relevant Targets}}{\sum \text{All Identified Targets}}$$

A high value in recall means that a detection method has nearly covered all interesting entities, a high value in precision means that no redundant entities have been identified. An illegal way of maximizing recall would be to consider each and every entity in a writing as interesting. Although in this case recall would become 100 %, a high amount of needless information is detected too. Consequently, the value of precision would be very low (this depends on the amount of available entities: if there are 100 entities in a document of which 99 are of interest, then the consideration of all 100 entities as interesting would have high values in both recall and precision). The same can be applied to precision: minimizing the false positive detection to

gain a high precision means that a lot of interesting pieces of information may have been lost during the detection process.

To calculate the two parameters we also have to consider partially correct results. A result is partially correct, if only a fraction of the whole entity could be correctly identified. For example, if the result of a tagging process would be “surgery”, but the actual interesting target had been “minimal invasive surgery”, this lack of a part of the original target would lead to one only partially correct result. In this case, the results counts only half the value of a true positive. The partially correctness is therefore a crucial element of the calculation for the analysis of the performance of a method.

In order to evaluate the vast amount of sentences in CPGs containing negations, we implemented a program to accomplish this task. This program has as input a goldstandard CPG. Within this CPG it searches for negation triggers. When a negation trigger is found, the whole sentence is analysed for the negated phrases (which are already tagged since it is a goldstandard) and the analysis results are stored. Afterwards, the taggings are removed. What remains is the original sentence of the CPG before it was turned into the goldstandard. The original sentence is now sent to *NegHunter*, which processes it. The processed sentence is then again returned to the evaluation program which analyzes it again and stores these results. Then, both goldstandard results and *NegHunter* results are compared to each other. The consequence of this process is, that we afterwards have a number of true positive, false positive, false negative and partially correct results. These numbers are stored and the analysis results of all sentences in the guideline are added together, so we get overall true positives, false positives, etc.. The results of each guideline is stored in a spreadsheet. After the processing of all test-CPGs we had the true positives, false positives, etc. and calculated recall and precision with following two formulas:

$$\text{Recall} = \frac{\sum \text{TruePositives} + \frac{1}{2} * \sum \text{PartiallyCorrects}}{\sum \text{TruePositives} + \sum \text{FalseNegatives} + \sum \text{PartiallyCorrects}}$$

$$\text{Precision} = \frac{\sum \text{TruePositives} + \frac{1}{2} * \sum \text{PartiallyCorrects}}{\sum \text{TruePositives} + \sum \text{FalsePositives} + \sum \text{PartiallyCorrects}}$$

In the part now following we give a detailed analysis of the performance of *NegHunter* in each negation class.

The Adverbial Negation: This most frequent class of negation totally amounts to 507 negated phrases, these are 69 % of all negated phrases. Of these phrases 75.54 % were correctly identified, 19.92 % were missed by the detection algorithm and 4.54 % were detected only partially correct.

Table 5.1 shows a detailed listing of detected phrases.

Table 5.1: Evaluation results of the adverbial negation

Adverbial Negation	Absolute number	Percentage
True Positive	383	75.54 %
False Positive	248	
False Negative	101	19.92 %
Partially Correct	23	4.54 %
Overall	755	100 %

The cell of the percentage of the false positive is empty, because of the fact that false positives are actually not part of the collectivity of negated phrases. They are phrases incorrectly considered as negated by the algorithm.

Our method achieves a recall of 77.81 % and a precision of 60.32 % in this negation type.

The Intra-Phrase Triggered Negation: 25 % of all negated phrases originate from the intra-phrase triggered negation. 90.11 % of these were correctly identified, whereas 7.14 % were missed by the algorithm. 2.75 % could only be identified partially. In Table 5.2 all values are listed.

Table 5.2: Evaluation results of the intra-phrase triggered negation

Intra-Phrase Triggered Negation	Absolute number	Percentage
True Positive	164	90.11 %
False Positive	1	
False Negative	13	7.14 %
Partially Correct	5	2.75 %
Overall	183	100 %

In this class we measured a recall of 91.48 % and a precision of 97.94 %.

The Prepositional Negation: The third most frequent class had an occurrence of 19 negated phrases. Nearly 95 % of these were identified accurately. The exact values are listed in Table 5.3.

Table 5.3: Evaluation results of the prepositional negation

Prepositional Negation	Absolute number	Percentage
True Positive	18	94.74 %
False Positive	5	
False Negative	1	5.26 %
Partially Correct	0	0 %
Overall	24	100 %

Recall of this class is 94.74 %, precision measures to 78.26 %.

The Verb Negation: This negation class had an occurrence of 18 phrases with a remarkable high number of false positive results of 11. The values are shown in Table 5.4.

Table 5.4: Evaluation results of the verb negation

Verb Negation	Absolute number	Percentage
True Positive	14	77.78 %
False Positive	11	
False Negative	0	0 %
Partially Correct	4	22.22 %
Overall	29	100 %

We calculated a recall of 88.89 % and a precision of 55.17 %.

The Adjective Negation: In this smallest class of all negations (five negated phrases) we achieved results as shown in Table 5.5.

Table 5.5: Evaluation results of the adjective negation

Adjective Negation	Absolute number	Percentage
True Positive	5	83.33 %
False Positive	0	
False Negative	1	16.67 %
Partially Correct	0	0 %
Overall	6	100 %

In this class we achieved a recall of 83.33 % and a precision of 100 %.

Overall values: The overall values give information on how good *NegHunter* performed without distinction between different negation classes. Due to the large number of adverbial negations this type strongly influence overall recall and precision. We achieved a recall of 82.08 % and a precision in the amount of 68.1 %. Table 5.6 lists all values.

Table 5.6: Overall evaluation results

Overall values	Absolute number	Relative number
True Positive	584	58.58 %
False Positive	265	26.58 %
False Negative	116	11.63 %
Partially Correct	32	3.21 %
Overall	996	100 %

In Table 5.7 the particular values of recall and precision are also summarized again. In the subsequent section we give an impression of errors not related to the negation detection algorithms but to the underlying software MMTx.

Table 5.7: Recall and Precision in each negation class

Negation Class	Recall	Precision
Adverbial Negation	77.81 %	60.32 %
Intra-Phrase Triggered Negation	91.48 %	97.94 %
Prepositional Negation	94.74 %	78.26 %
Adjective Negation	83.33 %	100 %
Verb Negation	88.89 %	55.17 %
Overall	83.92 %	69.36 %

5.2.1 Errors Related to MMTx

In some cases, errors occurred due to an incorrect part-of-speech tagging by the MMTx software. MMTx is a very sophisticated software which leads to good results in the Natural Language Processing.

In the adverbial negation, we could find the most errors (this class is the most frequent one). We had 20 false positive, 21 false negative, and 15 partially correct results in that negation type. The following sentence gives an example:

While the presence of some indicators such as perineural involvement or vascular tumor emboli alone may not necessarily indicate a high risk of recurrence, ...

Here, the phrase “a high risk” is not properly identified by MMTx. While “a risk” is correctly detected as a noun phrase, the combination with “high” leads to an incorrect result (MMTx cannot identify “a high risk” and therefore regards it as an unknown phrase), which leads to a false negative result in the evaluation. Another good example to show the limitations of MMTx is this sentence:

Trials powered to detect specified differences in main outcomes may not have sufficient power to detect adverse events that are less frequent.

In this sentence, MMTx detects the phrase “sufficient power to” as belonging together and classifies it as a noun phrase. This wrong classification brings one more only partially correct result for our evaluation.

In the intra-phrase triggered negation we had only five partially correct results. For example, in

There was no systematic attempt to search for all the “grey literature” (conferences, abstracts, theses and unpublished literature).

the actual negated phrase would be the “systematic attempt”. MMTx unfortunately appends the “to” to this phrase, leading to a partially correct result.

In the verb negation we had a similar problem:

The panel lacked a consensus to recommend it.

In this example the term “to” is also appended to the actual negated phrase “a consensus”, which generates an only partially correct result. In the class of verb negation, four partially correct results were caused by MMTx. Table 5.8 shows the numbers of registered MMTx errors related to their particular negation class, whereas Table 5.9 contains the relative number of MMTx errors compared to the overall errors. In this table one can see that on overwhelmingly large number of partially correct results is generated by MMTx itself (75 %), whereas only a relative small number of false positive (7.55 %) and false negative (18.26 %) comes from MMTx.

Table 5.8: Errors of MMTx related to their negation classes

Negation Class	False Positive	False Negative	Partially Correct
Adverbial Negation	20	21	15
Intra-Phrase Triggered Negation	0	0	5
Verb Negation	0	0	4
Overall	20	21	24

Table 5.9: Percentage of MMTx errors compared to overall errors

Negation class	Percentage
False Positives	7.55 %
False Negatives	18.1 %
Partially Correct	75 %

5.2.2 Discussion

The program shows its strength in the class of the intra-phrase triggered negation, the prepositional negation as well as the adjective negation. In the case of the intra-phrase triggered negation and the prepositional negation this matter of fact emerges through the easy structure of these classes. The concerned negated phrases occurs immediately after the trigger. Thus, a proper handling of this type is not a problem.

The adverbial negation and the verb negation need more sophisticated strategies. In these types verbs give information about where a phrase affected by a trigger can be. For the purpose of a proper Negation Detection, one has to keep the effect of different tenses in mind. The knowledge of the tenses determines, whether a sentence is written in active or passive voice. On this basis, a decision can be made, where the negated phrase occurs in a sentence. The problem of finding negated phrases is also the fact that they can occur on positions far away from the specific trigger. Due to this problem we have not only the failure of not detecting the phrase, but also the failure of spuriously detecting phrases in the space between trigger and actual negated phrase. Although a tagging scope of three phrases away from an adverbial trigger is a good compromise between recall and precision, there nevertheless occur errors. We also had to place value on the optimization of the recall of our method, as it is easier to remove mistakenly tagged phrases afterwards as it is to find them again. Thus, we had to accept an increasing number of false positive tagging, leading to a decreasing value in precision.

We believe that a Negation Detection using syntactical methods can improve the handling of negations in medical writings. As we use methods which are also applicable in writings different from medical documents, our approach can also be applied to documents in other domains. It is imaginable, that an approach using full-parsing, which also regards the sentence position of phrases can increase the performance of our algorithms. Moreover, the deployment of our method of Negation Detection is also possible in a non-medical area, since our algorithms are universally applicable for the English language. A restriction to the medical topic is therefore not implicitly necessary.

A refinement for the search of negated phrases for a specific topic using semantic methods is also possible. One could search for phrases which explicitly only concern the field of negative recommendations, or otherwise, one could accomplish a search for diseases or conditions, in which a certain CPG, respectively a part of it, is not applicable.

Chapter 6

Conclusion

In our presented approach we use syntactical information of the English language to decide, whether phrases in a sentence are affected by a negation trigger or not. The English grammar provides with its tenses and parts-of-speech enough information to allocate potential negated phrases. We use this information and created a classification system which contains five global classes of negation, which are distinguished by their different grammatical type. Negations working on neutral (in the sense that they are not negation triggers themselves) verbs using adverbs are therefore classified as “adverbial negation”, while negations using adjectives are referred to as “adjective negations”, and negations triggered by verbs are “verb negations”. Furthermore, we have two more types called “intra-phrase triggered negation” and the “prepositional negation”. In the first case, the triggers are contained in the negated phrase itself, and can be of several lexical categories (e.g., a preposition such as “without” or an adjective such as “ineffective”). In the second case, the triggers concern subsequent prepositional phrases and are not assignable to one specific lexical category too (e.g. the adjective “free” or the noun “lack”). Currently, our system is strongly limited to syntactical features. As a future outline, the involvement of methods using semantical aspects is conceivable. On the current base of processing, a subsequent deployment of semantical methods can augment the quality of Negation Detection and refine search parameters for interested persons. Compare the following example of [Chapman et al., 2001], which demonstrates negations with different contents that may be useful under different circumstances:

We did not treat the infection.

We did not detect an infection.

The first sentence indicates the absence of the treatment of an infection. Nevertheless, this infection is existent, whereas in the second sentence the existence of the infection itself is negated. This slight difference may seem dispensable, but in the medical context it gets an importance not negligible. With *NegHunter’s* syntactical methods in combination with a semantic level, medical data can be obtained specifically. Whereas one may be interested in negated findings in a certain CPG, another person maybe wants to find treatment options not advisable for special circumstances. Therefore, *NegHunter* provides a broad basis for Negation Detection which can be armed with special semantic concepts for any medical topic of interest.

Another option of including semantic into the Negation Detection is the observance of terms which are capable of modifying the strength of a negation. In our approach, the detection is a binary decision, i.e., a phrase can either be negated or not. In some cases, this binary method may not be sufficient. For example, in the sentence

Chemotherapy is not always recommended as a treatment.

the term “Chemotherapy” is negated, but the term “always” softens this absolute negation. In other words, chemotherapy is not recommended in the most cases (for example, when the physiological status of a person is extremely bad), anyhow, it can be appropriate under certain circumstances. As practitioners are strongly advised to critically adopt CPGs, the presentation of this type of information can therefore also be necessary.

Furthermore, we decided to chose a flexible implementation of our method in order to make it portable to arbitrary documents. Thanks to *NegHunter*'s classification of negations, an augmentation of already available negation triggers is always possible, thus making the system applicable for the usage in other types of documents. The new trigger only has to be assigned to a negation class and *NegHunter* then applies it's rulebase for this trigger. This makes *NegHunter* portable on other documents types as well as extensible and maintainable.

Our system provides methods for a targeted detection of negations, thus making it capable of processing even CPGs, which under different circumstances are difficult to handle, since they are written in free speech. Former approaches lack these features, so *NegHunter* makes a great contribution for the topic of automated processing of CPGs. Thus, it aids the handling of CPGs and alleviates their deployment in the practical field.

Bibliography

- [Aronow and Feng, 1997] Aronow, D. B. and Feng, F. (1997). Ad-hoc classification of electronic clinical documents. *D-Lib Magazine*.
- [Aronson, 2001] Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proc. of the AMIA Symposium*, pages 17–21.
- [Atkins et al., 2004a] Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., Guyatt, G. H., Harbour, R. T., Haugh, M. C., Henry, D., Hill, S., Jaeschke, R., Leng, G., Liberati, A., Magrini, N., Mason, J., Middleton, P., Mrukowicz, J., O’Connell, D., Oxman, A. D., Phillips, B., Schünemann, H., Edejer, T., Varonen, H., Vist, G. E., Williams Jr., J. W., Zaza, S., and The GRADE Working Group (2004a). Grading quality of evidence and strength of recommendations. *328(7454):1490–1498*.
- [Atkins et al., 2005] Atkins, D., Briss, P. A., Eccles, M., Flottorp, S., Guyatt, G. H., Harbour, R. T., Hill, S., Jaeschke, R., Liberati, A., Magrini, N., Mason, J., O’Connell, D., Oxman, A. D., Phillips, B., Schünemann, H., Edejer, T., Vist, G. E., Williams Jr., J. W., and The GRADE Working Group (2005). Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system. *BMC Health Services Research*, 5(25).
- [Atkins et al., 2004b] Atkins, D., Eccles, M., Flottorp, S., Guyatt, G. H., Henry, D., Hill, S., Liberati, A., O’Connell, D., Oxman, A. D., Phillips, B., Schünemann, H., Edejer, T., Vist, G. E., Williams Jr., J. W., and The GRADE Working Group (2004b). Systems for grading the quality of evidence and the strength of recommendations i: Critical appraisal of existing approaches. *BMC Health Services Research*, 4(38).
- [Boycheva et al., 2005] Boycheva, S., Strupchanska, A., Paskaleva, E., and Tcharaktchiev, D. (2005). Some aspects of negation processing in electronic health records. *Proc. of International Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries*, pages 1–8.
- [Bruix et al., 2005] Bruix, J., Sherman, M., and Practice Guidelines Committee, American Association for the Study of Liver Diseases (2005). Management of hepatocellular carcinoma. *Hepatology*, 42(5):1208–1236.
- [Chapman et al., 2001] Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301–310.
- [Divita, 2005] Divita, G. (2005). MMTX-API Documentation. <http://mmtx.nlm.nih.gov/>.

- [Elkin et al., 2005] Elkin, P. L., Brown, S. H., Bauer, B. A., Husser, C. S., Carruth, W., Bergstrom, L. R., and Wahner-Roedler, D. L. (2005). A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(13):1–7.
- [Feder et al., 1999] Feder, G., Eccles, M., Grol, R., Griffiths, C., and Grimshaw, J. (1999). Clinical guidelines: Using clinical guidelines. *BMJ*, 318:728–730.
- [Field and Lohr, 1990] Field, M. J. and Lohr, K. N. (1990). *Clinical Practice Guidelines: Directions for a New Program*. National Academies Press, Institute of Medicine, Washington DC.
- [Hahn et al., 2003] Hahn, T., Wingard, J., Anderson, K., Bensinger, W., Berenson, J., Brozeit, G., Carver, J., Kyle, R., and PL, P. M. (2003). The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of multiple myeloma: an evidence-based review. *Biol Blood Marrow Transplant*, 9(1):4–37.
- [Hazlehurst et al., 2005] Hazlehurst, B., Frost, H. R., Sittig, D. F., and Stevens, V. J. (2005). Mediclass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *Journal of the American Medical Informatics Association (JAMIA)*, 12(5):517–529.
- [Horn, 1989] Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago, Illinois.
- [Hripcsak et al., 1995] Hripcsak, G., Friedman, C., Alderson, P. O., DuMouchel, W., Johnson, S. B., and Clayton, P. D. (1995). Unlocking clinical data from narrative reports: A study of natural language processing. *Annals of Internal Medicine*, 122(9):681–688.
- [Huang and Lowe, 2007] Huang, Y. and Lowe, H. J. (2007). A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association (JAMIA)*, 14:304–311.
- [Humphreys and Schuyler, 1993] Humphreys, B. L. and Schuyler, P. (1993). The Unified Medical Language System: Moving beyond the vocabulary of bibliographic retrieval. pages 31–44. Meckler, Westport, CT.
- [Kaiser, 2007] Kaiser, K. (2007). Medical terminology systems. Technical Report Asgaard-TR-2007-1, Vienna University of Technology, Institute of Software Technology and Interactive Systems.
- [Kaiser et al., 2007] Kaiser, K., Martini, P., Miksch, S., and Öztürk, A. (2007). A meta schema for evidence information in clinical practice guidelines as a basis for decision-making. volume 129 of *Studies in Health Technology and Informatics*, pages 925–929, Brisbane, Australia. AMIA, IOS Press.
- [Klein and Manning, 2003] Klein, D. and Manning, C. (2003). Accurate Unlexicalized Parsing. *In Proc of the 41st Meeting of the Association for Computational Linguistics*.
- [Kvale, 2006] Kvale, P. (2006). Chronic cough due to lung tumors: ACCP evidence-based clinical practice guidelines. *Chest*, 129(1):147–153.
- [Kvale et al., 2003] Kvale, P., Simoff, M., and Prakash, U. (2003). Lung cancer. palliative care. *Chest*, 123(1):284–311.

- [Lee and Geller, 2006] Lee, Y. and Geller, J. (2006). Semantic enrichment for medical ontologies. *J. of Biomedical Informatics*, 39(2):209–226.
- [Lehnert et al., 1994] Lehnert, W., Cardie, C., Fisher, D., McCarthy, J., Riloff, E., and Soderland, S. (1994). Evaluating an information extraction system. *Journal of Integrated Computer-Aided Engineering*, 1(6):1–29.
- [McCray, 1989] McCray, A. T. (1989). UMLS Semantic Network. *The 13th annual SCAMC*, pages 503–507.
- [Meystre and Haug, 2005] Meystre, S. and Haug, P. J. (2005). Automation of a problem list using natural language processing. *BMC Medical Informatics and Decision Making*, 5(30).
- [Miaskowski et al., 2005] Miaskowski, C., Cleary, J., Burney, R., Coyne, P. J., Finley, R., Foster, R., Grossman, S., Janjan, N. A., Ray, J., Syrjala, K., Weisman, S. J., Pettit, J. B., and Zahrbock, C. (2005). Guideline for the management of cancer pain in adults and children. Clinical practice guideline 3, American Pain Society (APS), Glenview (IL).
- [Mitchell et al., 2004] Mitchell, K. J., Becich, M. J., Berman, J. J., Chapman, W. W., Gilbertson, J., Gupta, D., Harrison, J., Legowski, E., and Crowley, R. S. (2004). Implementation and evaluation of a negation tagger in a pipeline-based system for information extraction from pathology reports. pages 663–667.
- [Mutalik et al., 2001] Mutalik, P. G., Deshpande, A., and Nadkarni, P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents: A quantitative study using the UMLS. *Journal of the American Medical Informatics Association (JAMIA)*, 8(6):598–609.
- [National Collaborating Centre for Acute Care, 2005] National Collaborating Centre for Acute Care (2005). The diagnosis and treatment of lung cancer. Clinical practice guideline, National Institute for Clinical Excellence (NICE), London (UK).
- [Otchy et al., 2004] Otchy, D., Hyman, N., Simmang, C., Anthony, T., Buie, W., Cataldo, P., Church, J., Cohen, J., Dentsman, F., Ellis, C., 3rd, J. K., Ko, C., Moore, R., Orsay, C., Place, R., Rafferty, J., Rakinic, J., Savoca, P., Tjandra, J., and Whiteford, M. (2004). Practice parameters for colon cancer. *Dis Colon Rectum*, 47(8):1269–1284.
- [Patrick et al., 2006] Patrick, J., Wang, Y., and Budd, P. (2006). Automatic mapping clinical notes to medical terminologies. *Proceedings of the 2006 Australasian Language Technology Workshop (ALTW2006)*.
- [Rizzo et al., 2002] Rizzo, J., Lichtin, A., Woolf, S., Seidenfeld, J., Bennett, C., Cella, D., Djulbegovic, B., Goode, M., Jakubowski, A., Lee, S., Miller, C., Rarick, M., Regan, D., Browman, G., and Gordon, M. (2002). Use of epoetin in patients with cancer: evidence-based clinical practice guidelines of the american society of clinical oncology and the american society of hematology. *Blood*, 100(7):2003–2020.
- [Scottish Intercollegiate Guidelines Network (SIGN), 2003a] Scottish Intercollegiate Guidelines Network (SIGN) (2003a). Cutaneous melanoma. A national clinical guideline. SIGN publication 72, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh (Scotland).
- [Scottish Intercollegiate Guidelines Network (SIGN), 2003b] Scottish Intercollegiate Guidelines Network (SIGN) (2003b). Epithelial ovarian cancer. A national clinical guideline. SIGN publication 75, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh (Scotland).

- [Scottish Intercollegiate Guidelines Network (SIGN), 2004] Scottish Intercollegiate Guidelines Network (SIGN) (2004). Long term follow up of survivors of childhood cancer. A national clinical guideline. SIGN publication 76, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh (Scotland).
- [Scottish Intercollegiate Guidelines Network (SIGN), 2005a] Scottish Intercollegiate Guidelines Network (SIGN) (2005a). Management of patients with lung cancer. A national clinical guideline. SIGN publication 80, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh (Scotland).
- [Scottish Intercollegiate Guidelines Network (SIGN), 2005b] Scottish Intercollegiate Guidelines Network (SIGN) (2005b). Management of transitional cell carcinoma of the bladder. A national clinical guideline. SIGN publication 85, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh (Scotland).
- [Shekelle et al., 1999] Shekelle, P. G., Woolf, S. H., Eccles, M., and Grimshaw, J. (1999). Clinical guidelines: Developing guidelines. *BMJ*, 318:593–596.
- [Simon and Wagner, 2003] Simon, G. and Wagner, H. (2003). Small cell lung cancer. *Chest*, 123(1):259–271.
- [Singapore Ministry of Health, 2004a] Singapore Ministry of Health (2004a). Cervical cancer. Clinical practice guideline, Singapore Ministry of Health, Singapore.
- [Singapore Ministry of Health, 2004b] Singapore Ministry of Health (2004b). Colorectal cancer. Clinical practice guideline, Singapore Ministry of Health, Singapore.
- [Singapore Ministry of Health, National Committee on Cancer Care, 2004] Singapore Ministry of Health, National Committee on Cancer Care (2004). Breast cancer. Clinical practice guideline, Singapore Ministry of Health, Singapore.
- [SNOMED International, 2006] SNOMED International (2006). SNOMED CT. <http://www.snomed.org/snomedct/index.html> (last assessed: September, 2006).
- [Socinsky et al., 2003] Socinsky, M. A., Morris, D. E., Masters, G. A., and Lilenbaum, R. (2003). Chemotherapeutic management of stage IV non-small cell lung cancer. *Chest*, 123(1):226–243.
- [Tolentino et al., 2006] Tolentino, H., Matters, M., Walop, W., Law, B., Tong, W., Liu, F., Fontelo, P., Kohl, K., and Payne, D. (2006). Concept negation in free text components of vaccine safety reports. *Proc. of the AMIA 2006 Symposium*, page 1122.
- [Winqvist et al., 2004] Winqvist, E., Oliver, T., and R. Gilbert, H. (2004). The role of post-operative chemoradiotherapy for advanced squamous cell carcinoma of the head and neck. Clinical practice guideline 5–10, Cancer Care Ontario (CCO), Toronto (ON).
- [Wolkov et al., 2005] Wolkov, H., Constine, L., Yahalom, J., Chauvenet, A., Hoppe, R., Abrams, R., Deming, R., Mendenhall, N., Morris, D., Ng, A., Hudson, M., Winter, J., and Mauch, P. (2005). Staging evaluation - hodgkin’s disease. Clinical practice guideline, American College of Radiology (ACR), Reston (VA).