TEXT2TABLE:
Medical Text Summarization System based on Named Entity Recognition and Modality Identification

Eiji ARAMAKI

The university of Tokyo
eiji.aramaki@gmail.com

Yasuhide MIURA

Fuji Xerox
Yasuhide.Miura@fujixerox.co.jp

Masatsugu TONOIKE

Fuji Xerox
masatsugu.tonoike@fujixerox.co.jp

Tomoko OHKUMA
Fuji Xerox
ohkuma.tomoko@fujixerox.co.jp

Hiroshi MASHUICHI
Fuji Xerox
hiroshi.masuichi@fujixerox.co.jp

Kazuhiko OHE
The university of Tokyo Hospital
kohe@hcc.h.u-tokyo.ac.jp

Abstract

With the rapidly growing use of electronic health records, the possibility of large-scale clinical information extraction has drawn much attention. It is not, however, easy to extract information because these reports are written in natural language. To address this problem, this paper presents a system that converts a medical text into a table structure. This system's core technologies are (1) medical event recognition modules and (2) a negative event identification module that judges whether an event actually occurred or not. Regarding the latter module, this paper also proposes an SVM-based classifier using syntactic information. Experimental results demonstrate empirically that syntactic information can contribute to the method's accuracy.

1 Introduction

The use of electronic texts in hospitals is increasing rapidly everywhere. This study specifically examines discharge summaries, which are reports generated by medical personnel at the end of a patient's hospital stay. They include massive clinical information about a patient's health, such as the frequency of drug usage, related side-effects, and correlation between a disease and a patient's actions (e.g., smoking, drinking), which enables unprecedented large-scale research, engendering promising findings.

Nevertheless, it is not easy to extract clinical information from the reports because these reports are written in natural language. An example of a discharge summary is presented in Table 1. The table shows records that are full of medical jargon, acronyms, shorthand notation, misspellings, and sentence fragments (Tawanda et al., 2006).

To address this problem, this paper presents a proposal of a system that extracts medical events and date times from a text. It then converts them into a table structure. We designate this system *TEXT2TABLE*, which is available from a web site [1]. The extraction method, which achieves a high accuracy extraction, is based on Conditional Random Fields (CRFs) (Lafferty et al., 2001).

Another problem is posed by events that do not actually occur, i.e., future scheduled events, events that are merely intended to take place, or hypothetical events. As described herein, we call such non-actual events **negative events**. Negative events are frequently mentioned in medical records; actually, in our corpus, 12% of medical events are negative. Several examples of negative events (in italic letters) are presented below:

(1) *no* **headache**
(2) keep *appointment* of **radiotherapy**
(3) .. *will* have **intravenous fluids**

---

[1] http://lab0.com/

(4) .. came *for* **radiotherapy**

(4') .. came *for* **headache**

(5) Every week **radiation therapy** and chemical therapy *are scheduled*

(6) *Please* call Dr. Smith with worsening **headache** or back pain, or any other concern.

Negative events have two characteristics. First, various words and phrases indicate that an event is negative. For this study, such a word or phrase that makes an event negative is called a *negative trigger*. For instance, a negation word "*no*" is a negative trigger in (1). A noun "*appointment*" in (2) is a negative trigger. Similarly, the auxiliary "*will*" in (3) signals negation. More complex phenomena are presented in (4) and (4'). For instance, "**radiotherapy**" in (4) is a negative event because the therapy will be held in the future. In contrast, "**headache**" in (4') is not negative because a patient actually has a "**headache**". These indicate that a simple rule-based approach (such as a list of triggers) can only imply classification of whether an event is negative or not, and that information of the event category (e.g., a therapy or symptom) is required.

Another characteristic is a long scope of a negative trigger. Although negative triggers are near the descriptive words of events in (1)–(4), there could alternatively be a great distance of separation, as portrayed in (5) and (6). In (5), a noun coordination separates a negative trigger from the event. In (6), the trigger "*please*" renders all events in that sentence negative. These indicate that neighboring words are insufficient to determine whether an event is negative or not. To deal with (5), syntactic information is helpful because the trigger and the event are neighboring in the dependency structure, as portrayed in Fig. 2. To deal with (6), bag-of-word (BOW) information is desired.

Because of the observation described above, this paper presents a proposal of a classifier: whether an event is negative or not. The proposed classifier uses various information, the event category, neighboring words, BOW, and dependent phrases.

The point of this paper is two-fold: (1) We propose a new type of text-summarizing system (TEXT2TABLE) that requires a technique for a negative event identification. (2) We investigate what kind of information is helpful for negative event identification.

The experiment results revealed that, in spite of the risk of parsing error, syntactic information can contribute to performance, demonstrating the feasibility of the proposed approach.

Although experiments described in this paper are related to Japanese medical reports, the proposed method does not depend on specific languages or domains.

Table 1: A Health Record Sample.

```
BRIEF RESUME OF HOSPITAL COURSE : 57 yo with
NSCLCa with back pain and headache . Trans-
ferred from neurosurgery for additional mgmt
with palliative XRT to head . Pt initially
presented with cough and hemoptysis to his
primary MD . On CXR he was found to have a
upper left lobe mass . He subsequently un-
derwent bronchoscopy and bx revealed non-
small cell adeno CA. STaging revealed multi-
ple bony mets including skull, spine with
MRI revealing mild compression of vertebral
bodies at T9, T11, T12 . T9 with encroach-
ment of spinal cord underwent urgent XRT
with no response so he was referred to neu-
rosurgery for intervention . MRI-rt. fron-
tal, left temporal, rt cerebellar
hemorrhagic enhancing lesions- most likely
extensive intracranial mets- T-spine surgery
considered second priority and plan to radi-
ate cranially immediately with steroid and
anticonvulsant . He underwent simulation on
3/28 to whole brain and T3-T7 fields with
plan for rx to both sites over 2.5 weeks.
Over the past 2 weeks he has noted frontal
and occipital HA with left eyelid swelling,
ptosis, and denies CP, SOB, no sig. BM in
past 5 days, small amt of stool after sup-
pository. Neuro-He was Dilantin loaded and a
level should be checked on 3/31 . He is to
continue Decadron . Onc-He is to receive XRT
on 3/31 and daily during that week . Pain
control-Currently under control with MS con-
tin and MSIR prn. regimen . Follow HA, LBP.
ENDO-Glucose control monitored while on de-
cadron with SSRI coverage . Will check
HgbA1C prior to discharge . GI-Aggressive
bowel regimen to continue at home . Pt is
Full Code . ADDITIONAL COMMENTS: Please call
Dr. Xellcaugh with worsening headache or
back pain, or any other concern . Keep ap-
pointment as scheduled with XRT . Please
check fingerstick once a day, and record,
call MD if greater than 200 .
```

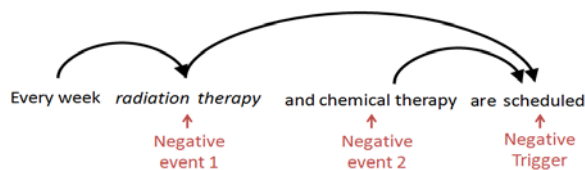Figure 1: Visualization result (Left), magnified (Right).

Figure 2: Negative Triggers and Events on a Dependency Structure.

Table 2: Corpora and Modalities

CORPUS	MODALITY
ACE	asserted, or other
TIMEML	must, may, should, would, or could
Prasad et al., 2006	assertion, belief, facts or eventualities
Saurí et al., 2007	certain, probable, possible, or other
Inui et al., 2008	affirm, infer, doubt, hear, intend, ask, recommend, hypothesize, or other
THIS STUDY	S/O, necessity, hope, possible, recommend, intend

Table 3: Markup Scheme (Tags and Definitions)

Tag	Definition (Examples)
R	Remedy, Medical operation (e.g. radiotherapy)
T	Medical test, Medical examination (e.g., CT, MRI)
D	Deasese, Symptom (e.g., Endometrial cancer, headache)
M	Medication, administration of a drug (e.g., Levofloxacin, Flexeril)
A	patient action (e.g., admitted to a hospital)
V	Other verb (e.g., cancer spread to ...)

2 Related Works

2.1 Previous Markup Schemes

In the NLP field, fact identification has not been studied well to date. Nevertheless, similar analyses can be found in studies of sentence modality.

The Automatic Content Extraction (ACE)[2] information extraction program deals with event extraction, by which each event is annotated with temporal and modal markers.

A similar effort is made in the TimeML project (Pustejovsky et al., 2003). This project specifically examines temporal expressions, but several modal expressions are also covered.

Prasad et al. (2006) propose four factuality classifications (certain, probable...etc.) for the Penn Discourse TreeBank (PDTB)[3].

Saurí et al. (2007) propose three modal categories for text entailment tasks.

Among various markup schemes, the most recent one is *Experience Mining* (Inui et al., 2008), which collects personal experiences from the web. They also distinguish whether an experience is an actual one or not, which is a similar problem to that confronting us.

Table 2 portrays a markup scheme adopted by each project. Our purpose is similar to that of Experience Mining. Consequently, we fundamentally adopt its markup scheme. However, we modify the label to suit medical mannerisms. For example, "doubt" is modified into "(S/O) suspicion of". Rare modalities such as "hear" are removed.

2.2 Previous Algorithms

Negation is a traditional topic in medical fields. Therefore, we can find many previous studies of the topic in the relevant literature.

An algorithm, *NegEx*[4] was proposed by Chapman et al. (Chapman et al., 2001a; Chapman et al., 2001b). It outputs an inference of whether a term is positive or negative. The original algorithm is based on a list of negation expressions. Goldin et al. (2003) incorporate machine learning techniques (Naïve Bayes and decision trees) into the algorithm. The extended version (ConText) was also proposed (Chapman et al., 2007).

Elkin et al. (2005) use a list of negation words and a list of negation scope-ending words to iden-

---

[2] http://projects.ldc.upenn.edu/ace/
[3] http://www.seas.upenn.edu/~pdtb/
[4] http://www.dbmi.pitt.edu/chapman/NegEx.html

tify negated statements and their scope. Their technique was used in The MAYO Clinic Vocabulary Server (MCVS)[5], which encodes clinical expressions into medical ontology (SNOMED-CT) and identifies whether the event is positive or negative.

Mutalik et al. (2001) earlier developed *Negfinder* to recognize negated patterns in medical texts. Their system uses regular expressions to identify words indicating negation. Then it passes them as special tokens to the parser, which makes use of the single-token look-ahead strategy.

Huang and Lowe (2007) implemented a hybrid approach to automated negation detection. They combined regular expression matching with grammatical parsing: negations are classified based on syntactic categories. In fact, they are located in parse trees. Their hybrid approach can identify negated concepts in radiology reports even when they are located distantly from the negative term.

The Medical Language Extraction and Encoding (MedLEE) system was developed as a general natural language processor to encode clinical documents in a structured form (Friedman et al., 1994). Negated concepts and certainty modifiers are also encoded within the system.

Veronika et al. (2008) published a negation scope corpus[6] in which both negation and uncertainty are addressed.

Although their motivations are identical to ours, two important differences are apparent. (1) Previous (except for Veronika et al., 2008) methods deal with the two-way problem (positive or negative), whereas the analyses proposed herein tackle more fine-grained modalities. (2) Previous studies (except for Huang et al., 2007) are based on BOW approaches, whereas we use syntactic information.

3 Medical Text Summarization System: TEXT2TABLE

Because the core problem of this paper is to identify negative events, this section briefly presents a description of the entire system, which consists of four steps. The detailed algorithm of negative identification is explained in Section 4.

STEP 1: Event Identification
First, we define the event discussed in this paper. We deal with events of six types, as presented in

Table 3. Two of the four are Verb Phrases (base VPs); the others are noun phrases (base-NPs). Because this task is similar to Named Entity Recognition (NER), we use the state-of-the art NER method, which is based on the IOB2 representation and Conditional Random Fields (CRFs). In learning, we use standard features, as shown in Table 4.

Table 4: Features for Event Identification

Lexicon and Stem	Current target word (and its stem) and its surrounding words (and stem). The window size is five words (-2, -1, 0, 1, 2).
POS	Part of speech of current target word and its surrounding words (-2, -1, 0, 1, 2). The part of speech is analyzed using a POS tagger[7].
DIC	A fragment for the target word appears in the medical dictionary (Ito et al., 2003).

STEP 2: Normalization
As described in Section 1, a term in a record is sometimes an acronym: shorthand notation. Such abbreviations are converted into standard notation through (1) date time normalization or (2) event normalization.

(1) Date Time Normalization
As for date time expressions, relative date expressions are converted into YYYY/MM/DD as follows.

On Dec Last year → 2007/12/XX
10 Dec 2008 → 2008/12/10

These conversions are based on heuristic rules.

(2) Event Normalization
Medical terms are converted into standard notation (dictionary entry terms) using orthographic disambiguation (Aramaki et al., 2008).

STEP 3: TIME–EVENT Relation Identification
Then, each event is tied with a date time. The current system relies on a simple rule (i.e., an event is tied with the latest date time).

STEP 4: Negative Identification
The proposed SVM classifier distinguishes negative events from other events. The detailed algorithm is described in the next section.

4 Modality Identification Algorithm

First, we define the negative. We classify modality events into eight types (Table 5). These classifications are motivated by those used in previous stud-

ies (Inui et al., 2008). However, we simplify their scheme because several categories are rare in this domain.

These classes are not exclusive. For that reason, they sometimes lead to multiple class events. For example, given "No **chemotherapy** is planned", an event "**chemotherapy**" belongs to two classes, which are "NEGATION" and "FUTURE".

Training Phase

Using a corpus with modality annotation, we train a SVM classifier for each category. The training features come from four parts:

(1) Current phrases: words included in a current event. We also regard their STEMs, POSs, and the current event category as features.

(2) Surrounding phrases: words included in the current event phrase and its surrounding two phrases (p1, p2, n1, n2, as depicted in Fig. 3). The unit of the phrase is base-NP/VP, which is produced by the Japanese parser (Kurohashi et al., 1994). Its window size is two in the neighboring phrase (p1, p2, c, n1, n2). We also deal with their STEMs and POSs.

(3) Dependent phrases: words included in the parent phrase of the current phrase (d1 in Fig. 3), and grandparent phrases (d2 in Fig. 3). We also deal with their STEMs and POSs.

(4) Previous Event: words (with STEMs and POSs) included in the previous (left side) events. Additionally, we deal with the previous event category and the modality class.

(5) Bag-of-words: all words (with STEMs and POSs) in the sentence.

TEST Phrase

During the test, each SVM classifier runs.

Although this task is multiclass labeling, several class combinations are unnatural, such as FUTURE and S/O. We list up possible label combinations (that have at least one occurrence in the corpora); if such a combination appears in a text, we adapt a high confidence label (using a marginal distance).

5 Experiments

We investigate what kind of information contributes to the performance in various machine learning algorithms.

Table 5: Classification of Modalities

NEGATION	An event with negation words such as "not" or "no".
FUTURE	An event that is scheduled for execution in the future.
PURPOSE	An event that is planed by a doctor, but its time schedule is ambiguous (just a hope/intention).
S/O	An event (usually a disease) that is suspected. For example, given "suspected microscopic tumor in ...", "microscopic tumor" is an S/O event."
NECESSITY	An event (usually a remedy or medical test) that is required.
INTEND	An event that is hoped for by a patient. Note that if the event is hoped by a doctor, we regard is a PURPOSE or FUTURE. For example, given "He hoped for chemical therapy", "chemical therapy" is INTEND.
POSSIBLE	An event (usually remedy) that is possible under the current situation.
RECOMMEND	An event (usually remedy) that is recommended by other doctor(s).

5.1 Corpus and Setting

We collected 435 Japanese discharge summaries in which events and the modality are annotated. For training, we used the CRF toolkit[8] with standard parameters. In this experiment setting, the input is an event with its contexts. The output is an event modality class (positive of negative in two-way) (or more detailed modality class in nine-way).

The core problem addressed in this paper is modality classification. Therefore, this task setting assumes that all events are identified correctly. Table 6 presents the event identification accuracy. Except for the rare class V (the other verb), we got more than 80% *F*-scores. It is true that the accuracy is not perfect. Nevertheless, most of the remaining problems in this step will be solved using a larger corpus.

5.2 Comparable Methods

We conducted experiments in the 10-fold cross validation manner. We investigated the perform-

---

[8] http://crfpp.sourceforge.net/

ance in various feature combinations and the following machine learning methods.
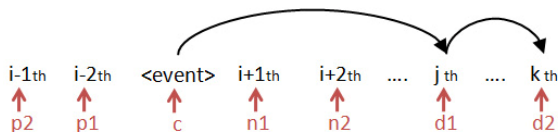
Figure 3: Features

Table 6: Event Identification Result. Tag precision recall *F*-score.

	#	P	R	F
A (ACTION)	1,556	94.63	91.04	92.80
V (VERB)	1,047	84.64	74.89	79.47
D (DISEASE)	3,601	85.56	80.24	82.82
M (MEDICINE)	1,045	86.99	81.34	84.07
R (REMEDY)	1,699	84.50	76.36	80.22
T (TEST)	2,077	84.74	76.68	80.51
ALL	11,025	84.74	76.68	80.51

Table 7: Various Machine Learning Method

SVM	Support Vector Machine (Vapnik, 1999). We used TinySVM[9] with a polynomial kernel (degree=2).
AP	Averaged Perceptron (Collins, 2002)
PA1	Passive Aggressive I (Crammer et al., 2006)*
PA2	Passive Aggressive II (Crammer et al., 2006)*
CW	Confidence Weighted (Dredze et al., 2008)*

* The online learning library[10] is used for AP PA1,2 CW .

5.3 Evaluation Metrics
We adopt evaluation of two types:
(1) Two-way: positive or negative:
(2) Nine-way: positive or one of eight modality categories.
Recall and *F*-measure are investigated in both for evaluation precision.

5.4 Results
The results are shown in Table 8 (Two-Way) and in Table 9 (Nine-Way).
Current Event Category
The results in ID0–ID1 indicate that the current event category (CAT) is useful. However, events are sometimes misestimated in real settings. We

[9] http://chasen.org/ taku/software/TinySVM/
[10] http://code.google.com/p/oll

must check more practical performance in the future.

Bag-of-words (BOW) Information
Results in ID1–ID2 indicate that BOW is important.

Surrounding Phrase Contribution
The results appearing in ID2–ID9 represent the contribution of each feature position. From ID3, ID4, and ID7 results, next phrases (n1, n2) and parent phrases (d1) were able to boost the accuracy. Despite the risk of parsing errors, parent phrases (d1) are helpful, which is an insight of this study.

In contrast, we can say that the following features had little contribution: previous phrases (p1, p2 from ID5 and ID6), grandparent phrases (d2 from ID8), and previous events (e from ID9).

Regarding p1 and p2, these modalities are rarely expressed in the previous parts in Japanese.

As for d2, the grandparent phrases might be too removed from the target events.

As for e, because texts in health records are fragmented, each event might have little relation.

However, the above features are also helpful in cases with a stronger learning algorithm.

In fact, among ID10–ID14, the SVM-based classifier achieved the best accuracy with all features (ID14).

Table 8: Two-way Results

ID	Learning Algorithm	CAT	BOW	P2	P1	c	n1	n2	d1	d2	e	Precision	Recall	F-measure
0	CW					●						61.25	50.10	54.25
1	CW	●				●						63.73	51.18	56.76
2	CW	●	●			●						78.55	70.21	74.14
3	CW	●	●			●	●					84.40	78.08	81.11
4	CW	●	●			●	●	●				**85.71**	77.72	81.51
5	CW	●	●		●	●	●	●				84.61	75.53	79.81
6	CW	●	●	●	●	●	●	●				85.34	71.85	78.01
7	CW	●	●			●	●	●	●			85.48	**78.22**	**81.68**
8	CW	●	●			●	●	●	●	●		83.94	78.08	80.80
9	CW	●	●			●	●	●	●		●	83.20	78.04	80.53
10	SVM	●	●			●	●	●				88.57	80.84	84.52
11	SVM	●	●			●	●	●	●			89.24	81.05	84.94
12	SVM	●	●			●	●	●	●	●		89.29	81.54	85.23
13	SVM	●	●			●	●	●	●		●	89.30	81.54	85.24
14	SVM	●	●	●	●	●	●	●	●	●	●	**89.40**	**82.50**	**85.81**
15	AP	●	●			●	●	●	●			92.28	64.56	75.97
16	PA1	●	●			●	●	●	●			93.08	68.03	78.60
17	PA2	●	●			●	●	●	●			92.99	67.67	78.33

● indicates the used feature. c are features from the current phrase. p1, p2, n1, n2 are features from surrounding phrases. e are features from a previous event. **BOW** is a bag-of-words using features from an entire sentence. **CAT** is the category of the current event.

Learning Methods

Regarding the learning algorithms, all online learning methods (ID7 and ID15–17) showed lower accuracies than SVM (ID11), indicating that this task requires heavy learning.

Nine-way Results

Table 9 presents the accuracies of each class. Fundamentally, we can obtain high performance in the frequent classes (such as NEGATION, PURPOSE, and S/O). In contrast, the classifier suffers from low frequent classes (such as FUTURE). How to handle such examples is a subject of future study.

Table 9: Two-way Results

	#	Precision	Recall	*F*-measure
NEGATION	441	84.19	77.36	80.63
PURPOSE	346	91.35	63.87	75.17
S/O	242	90.74	72.39	80.53
FUTURE	97	23.31	55.96	32.91
POSSIBLE	36	83.33	40.55	54.55
INTEND	32	76.66	29.35	42.44
RECOMMEND	21	95.71	38.57	54.98
NECESSITY	4	100	0	0

4.5 Future Works

In this section, we will discuss several remaining problems. First, as described, the classifier suffers from low frequent modality classes. To give more examples for such classes is an important problem.
 Our final goal is to realize precise information extraction from health records. Our IE systems are already available at the web site (http://lab0.com). Comprehensive evaluation of those systems is required.

6   Conclusions

This paper presented a classifier that identified whether an event has actually occurred or not. The proposed SVM-based classifier uses both BOW information and dependency parsing results. The experimental results demonstrated 85.8 *F*-measure% accuracy and revealed that syntactic information can contribute to the method's accuracy. In the future, a method of handling low-frequency events is strongly desired.

References

Wendy Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce Buchanan. 2001a. Evaluation of negation phrases in narrative clinical reports. In Proceedings of AMIA Symp, pages 105-109.

Wendy Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce Buchanan. 2001b. A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics, 5:301-310.

Wendy Chapman, John Dowling and David Chu. 2007. ConText: An algorithm for identifying contextual features from clinical text. Biological, translational, and clinical language processing (BioNLP2007), pp. 81–88.

Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe: Orthographic Disambiguation Incorporating Transliterated Probability International Joint Conference on Natural Language Processing (IJCNLP2008), pp.48-55, 2008.

Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom, and Dietlind L. Wahner Roedler. A controlled trial of automated classification of negation from clinical notes. BMC Medical Informatics and Decision Making 5:13.

C. Friedman, P.O. Alderson, J.H. Austin, J.J. Cimino, and S.B. Johnson. 1994. A general natural language text processor for clinical radiology. Journal of the American Medical Informatics Association, 1(2):161-174.

L. Gillick and S.J. Cox. 1989. Some statistical issues in the comparison of speech recognition algorithms. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 532-535.

Ilya M. Goldin and Wendy Chapman. 2003. Learning to detect negation with not in medical texts. In Workshop at the 26th ACM SIGIR Conference.

Yang Huang and Henry J. Lowe. 2007. A novel hybrid approach to automated negation detection in clinical radiology reports. Journal of the American Medical Informatics Association, 14(3):304-311.

Kentaro Inui, Shuya Abe, Hiraku Morita, Megumi Eguchi, Asuka Sumida, Chitose Sao, Kazuo Hara, Koji Murakami, and Suguru Matsuyoshi. 2008. Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence, pages 314-321.

M. Ito, H. Imura, and H. Takahisa. 2003. Igaku- Shoin's Medical Dictionary. Igakusyoin.

Sadao Kurohashi and Makoto Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. Computational Linguistics, 20(4).

Pradeep G. Mutalik, Aniruddha Deshpande, and Prakash M. Nadkarni. 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, 8(6):598-609.

J. Lafferty, A. McCallum, and F. Pereira: Conditional random fields: Probabilistic models for segmenting and labeling sequence data, In Proceedings of the International Conference on Machine Learning (ICML2001), pp.282-289, 2001.

R. Prasad, N. Dinesh, A. Lee, A. Joshi and B. Webber: Annotating Attribution in the Penn Discourse TreeBank, In Proceedings of the International Conference on Computational Linguistics and the Annual Conference of the Association for Computational Linguistics (COLING/ACL2006) Workshop on Sentiment and Subjectivity in Text, pp.31-38 (2006).

R. Saurí, and J. Pustejovsky: Determining Modality and Factuality for Text Entailment, Proceedings of ICSC2007, pp. 509-516 (2007).

Gaizauskas, A. Setzer, G. Katz, and D.R. Radev. 2003. New Directions in Question Answering: Timeml: Robust specification of event and temporal expressions in text. AAAI Press.

SNOMED-CT. 2002. SNOMED Clinical Terms Guide. College of American Pathologists.

Sibanda Tawanda, Tian He, Peter Szolovits, and Uzuner Ozlem. 2006. Syntactically informed semantic category recognizer for discharge summaries. In Proceedings of the Fall Symposium of the American Medical Informatics Association (AMIA 2006), pages 11-15.

Sibanda Tawanda and Uzuner Ozlem. 2006. Role of local context in automatic deidentification of ungrammatical, fragmented text. In Proceedings of the Human Language Technology conference and the North American chapter of the Association for Computational Linguistics (HLT-NAACL2006), pages 65-73.

Veronika Vincze, Gyorgy Szarvas, Richard Farkas, Gyorgy Mora, and Janos Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics, 9(11).