



Extracting medication information from clinical text

Özlem Uzuner,¹ Imre Solti,² Eithon Cadag²

► Additional figures and tables are published online only. To view these files please visit the journal online (<http://jamia.bmj.com>).

¹Department of Information Studies, University at Albany, State University of New York, Albany, New York, USA

²Division of Biomedical and Health Informatics, University of Washington, Seattle, Washington, USA

Correspondence to

Dr Özlem Uzuner, College of Computing and Information, University at Albany, SUNY, Draper 114A, 135 Western Ave, Albany, NY 12222, USA; ouzuner@albany.edu

Received 22 February 2010

Accepted 25 June 2010

ABSTRACT

The Third i2b2 Workshop on Natural Language Processing Challenges for Clinical Records focused on the identification of medications, their dosages, modes (routes) of administration, frequencies, durations, and reasons for administration in discharge summaries. This challenge is referred to as the medication challenge. For the medication challenge, i2b2 released detailed annotation guidelines along with a set of annotated discharge summaries. Twenty teams representing 23 organizations and nine countries participated in the medication challenge. The teams produced rule-based, machine learning, and hybrid systems targeted to the task. Although rule-based systems dominated the top 10, the best performing system was a hybrid. Of all medication-related fields, durations and reasons were the most difficult for all systems to detect. While medications themselves were identified with better than 0.75 F-measure by all of the top 10 systems, the best F-measure for durations and reasons were 0.525 and 0.459, respectively. State-of-the-art natural language processing systems go a long way toward extracting medication names, dosages, modes, and frequencies. However, they are limited in recognizing duration and reason fields and would benefit from future research.

INTRODUCTION

Clinical records contain information that can be invaluable, for example, for pharmacovigilance, for comparative effectiveness studies, and for detecting adverse events. The structured and narrative components of clinical records collectively provide a comprehensive account of the medications of patients.^{1 2} The Third i2b2 Workshop on Natural Language Processing Challenges for Clinical Records focused on the extraction of medications and medication-related information from discharge summaries for the purpose of making this information readily available for use. We refer to this challenge as the medication challenge. This article provides an overview of the medication challenge, describes the top 10 systems, evaluates and analyzes their outputs, and provides directions for the continued improvement of the state of the art.

The medication challenge was designed as an information extraction task.^{3 4} The goal, for each discharge summary, was to extract the following information (called 'fields'⁵) on medications experienced by the patient:

- Medications (m): including names, brand names, generics, and collective names of prescription substances, over the counter medications, and other biological substances for which the patient is the experimenter.
- Dosages (do): indicating the amount of a medication used in each administration.

- Modes (mo): indicating the route for administering the medication.
- Frequencies (f): indicating how often each dose of the medication should be taken.
- Durations (du): indicating how long the medication is to be administered.
- Reasons (r): stating the medical reason for which the medication is given.
- List/narrative (ln): indicating whether the medication information appears in a list structure or in narrative running text in the discharge summary.

The medication challenge asked that systems extract the text corresponding to each of the fields for each of the mentions of the medications that were experienced by the patients. The text corresponding to each field was specified by its line and token offsets in the discharge summary so that repeated mentions of a medication could be distinguished from each other. The values for the set of fields related to a medication mention, if presented within a two-line window of the mention, were linked in order to create what we defined as an 'entry'. If the value of a field for a mention were not specified within a two-line window, then the value 'nm' for 'not mentioned' was entered and the offsets were left unspecified. Medication and list/narrative fields could never be 'nm'. The list/narrative field required no offset as no text was extracted for it.

Figure 1 and figure 2 (available online at <http://jamia.bmj.com>) show two excerpts along with the 'gold standard' entries corresponding to the italicized portions of these excerpts. In both figures, there is one entry per mention of each medication. Each entry lists the fields that relate to the same medication mention. The fields are separated by double pipes (||).

i2b2 provided annotation guidelines for the medication challenge, released a small set of discharge summaries with their gold standard annotations, and invited the development of systems that could address the challenge. Twenty teams from 23 organizations and nine countries participated in the challenge. The findings of the medication challenge were presented in a workshop that i2b2 organized, in co-sponsorship with the American Medical Informatics Association (AMIA), at the Fall Symposium of AMIA in 2009.

RELATED WORK

The medication challenge continued i2b2's efforts to release clinical records to the natural language processing (NLP) community for the advancement of state of the art. This challenge built on past i2b2 challenges,^{6–8} as well as past shared task efforts^{9–11} outside of the clinical domain.

The Genomics Track of Hersch *et al*¹¹ in the Text Retrieval Evaluation Conferences encouraged the

Line no.	text
63	well. Although left transmetatarsal amputation being considered .
64	it was felt that she had a good chance of healing the wound
65	appropriately. She had a single temperature spike , although all
66	cultures remained negative. <i>She had continuation of her Heparin</i>
67	<i>while she was started on a course of Coumadin to reserve patency of</i>
68	<i>her graft. ...</i>
Gold standard	
m="heparin" 66:8 66:8 do="nm" mo="nm" f="nm" du="nm" r="nm" ln="narrative"	
m "coumadin" 67:8 67:8 do "nm" mo "nm" f "nm" du "nm" r "her graft." 68:0	
68:1 ln="narrative"	

Figure 1 Sample discharge summary narrative and the output for the italicized portion. m, do, mo, f, du, r, and ln stand for medication, dosage, mode, frequency, duration, reason, and list/narrative, respectively.

development of systems that can retrieve the documents relevant to a query from a set. The BioCreAtIvE of Hirschman *et al*¹⁰ evaluated systems that can extract biomedical entities and their relations from bioscience text. The ICD-9-CM coding challenge of Pestian *et al*⁹ ranked practical systems that can automatically classify radiology reports.

Past i2b2 challenges complemented these shared-task efforts. The first i2b2 challenge included an information extraction task focused on deidentification⁶ and a document classification task focused on the smoking status of patients.⁷ The second i2b2 challenge was a multi-document classification task on obesity and 15 co-morbidities; this challenge required information extraction for disease-specific details, negation and uncertainty extraction on diseases, and classification of patient records.⁸

The third i2b2 challenge—that is, the medication challenge—extends information extraction to relation extraction; it requires extraction of medications and medication-related information followed by determination of which medication belongs to which medication-related details. For an overview of studies focused on medication extraction, please see online supplements at <http://jamia.bmj.com>.^{12–16}

DATA

The medication challenge was run on deidentified discharge summaries from Partners Healthcare. The summaries were released under a data use agreement that allows their use for research beyond the challenge. All relevant institutional review boards approved the challenge and the use of the discharge summaries.

A total of 1243 deidentified discharge summaries were used for the medication challenge; 696 of these summaries were released during the development period. The i2b2 team generated ‘gold standard’ annotations for 17 of the 696. A total of 547 discharge summaries were held out for testing. Challenge participants (after turning in their system submissions) collectively developed the gold standard annotations for 251 summaries out of the 547. Details of this annotation process can be found in Uzuner *et al*.⁵ Table 1 (available online at <http://jamia.bmj.com>) shows the number of examples of each field that were found in the gold standard annotations and the average per discharge summary.

METHODS

We defined two sets of evaluation metrics, referred to as the horizontal and vertical metrics, for evaluating the system outputs submitted to the challenge. We used horizontal metrics for measuring system performance on ‘entries’; we used vertical metrics for measuring performance on individual ‘fields’. We computed both metrics at phrase and token level, using

precision, recall, and F-measure. For our purposes, phrases refer to the complete text of field values; tokens are delimited by spaces and punctuation. We defined phrase-level and token-level precision, recall, and F-measure on the basis of the definitions provided for evaluating the ‘list questions’ of the Question Answering track of the Text Retrieval Evaluation Conferences (see equations 1–6)¹⁷:

$$\text{Phrase-level precision (PP)} = \frac{\# \text{Correctly returned phrases by system}}{\# \text{Phrases returned by the system}} \quad \text{Equation 1}$$

$$\text{Phrase-level recall (PR)} = \frac{\# \text{Correctly returned phrases by system}}{\# \text{Phrases in gold standard}} \quad \text{Equation 2}$$

$$\text{Phrase-level F-measure} = \frac{(\beta^2 + 1) \times \text{PP} \times \text{PR}}{(\beta^2 \times \text{PP}) + \text{PR}} \quad \text{where } \beta=1 \quad \text{Equation 3}$$

where β marks the relative weights of precision and recall.

$$\begin{aligned} \text{Token-level precision (TP)} \\ = \frac{\# \text{Correctly returned tokens from each phrase in system output}}{\# \text{tokens in system output}} \end{aligned} \quad \text{Equation 4}$$

$$\begin{aligned} \text{Token-level recall (TR)} \\ = \frac{\# \text{Correctly returned tokens from each phrase in system output}}{\# \text{tokens in gold standard}} \end{aligned} \quad \text{Equation 5}$$

$$\text{Token-level F-measure} = \frac{(\beta^2 + 1) \times \text{TP} \times \text{TR}}{(\beta^2 \times \text{TP}) + \text{TR}} \quad \text{where } \beta=1 \quad \text{Equation 6}$$

Entries that are ‘nm’ in the gold standard are omitted from the evaluation. We align entries in the system output with entries in the gold standard using the medication name and offset. In the case of multiple entries for a medication at a given offset, for each entry of the gold standard, the system entry that gives the best F-measure is selected as the alignment match; each gold standard entry can match only one entry in the system output. The list/narrative field is omitted from this process as it requires no text to be extracted from the discharge summaries. Pseudocode for this alignment is presented in Figure 3 (available online at <http://jamia.bmj.com>).

Horizontal metrics

The primary evaluation metric for the challenge computed phrase-level and token-level F-measures on entries. We refer to this metric as the horizontal metric.

Vertical metrics

Vertical metrics measure phrase-level and token-level performance on individual fields. For example, vertical metrics for medications measure performance only on medication names without any consideration of the entries of these medications.

Significance tests

We used approximate randomization for testing significance.¹⁸ For a pair of outputs *A* and *B* from two different systems, we

computed the horizontal phrase-level F-measure of each output and noted the difference in performance ($f=A-B$). Let j be the number of entries in A , and let k be the number of entries in B ; A and B are combined to a superset of entries, C , of size $j+k$. For each iteration i up to n iterations, we draw j entries randomly and without resampling from C and create the pseudoset of entries A_i . The remainder of C then forms the pseudoset of entries B_i . We compute the phrase-level horizontal F-measures for A_i and B_i , and note the difference between them ($f_i=A_i-B_i$). We then count the number of times that $f_i-f\geq 0$, for all i , and divide this count by n to calculate the p value between A and B .¹⁹ We ran significance tests on the outputs of the top 10 teams with $n=1000$.

Systems

Twenty teams participated in the medication challenge (see table 2 available online at <http://jamia.bmj.com>). The teams received the annotated discharge summaries and the annotation guidelines in June 2009. They were then given approximately 2.5 months to develop their systems and to (optionally) annotate their own training data.^{20–22} At the end of the development period, i2b2 released test discharge summaries and collected system outputs on the test summaries 3 days after their release. Each team was permitted to submit up to three system outputs for evaluation. At the time of submission, the teams classified their system outputs along three dimensions.

1. External resources (marked as ‘Yes’ or ‘No’): system outputs that used proprietary systems, data, and resources that were not available to other teams were considered to have been generated with the aid of external resources and were marked as ‘Yes’.
2. Medical experts (marked as ‘Yes’ or ‘No’): system outputs that were the product of a team that involved medical experts—for example, medical doctors, nurses, etc—were marked as ‘Yes’.
3. Methods (marked as ‘Supervised’, ‘Rule-based’, or ‘Hybrid’): system outputs were marked on the basis of the approaches that the teams utilized.

Each team was evaluated on the basis of their best performing system output. Of the 20 evaluated outputs, four were declared to have utilized external resources, five were declared to have benefitted from medical experts, 10 were described by their authors as rule-based, four as supervised, and six as hybrids.

Seven out of the top 10 teams’ best performing submissions were rule-based.^{23–29} In general, these teams applied text filtering to eliminate the content that was not related to the medications of the patient.^{23–26} They built vocabularies from

publicly available knowledge sources such as the Unified Medical Language System (UMLS),^{23–25} enriched these vocabularies with examples from the training data and the annotation guidelines,²³ and bootstrapped examples from unlabeled i2b2 discharge summaries²³ as well as the web.²⁸

Given their vocabularies, the teams employed pattern matching and regular expressions to determine most of the medication information.^{23–28} However, they differed in their core tools: BME-Humboldt used GNU software for regular expressions and the Unstructured Information Management Architecture (UIMA) as their base²³; NLM used MetaMap for marking reasons²⁵; Open University (OpenU) employed the Genia Tagger for part of speech tagging²⁶; University of Paris (UParis) used their Ogmios platform for word and sentence segmentation, part-of-speech tagging, and lemmatization²⁷; Vanderbilt University used the MedEx system for tagging medication names, frequencies, etc and a context free grammar for converting the text into a structured format.²⁹

The University of Utah (UofUtah),³⁰ the University of Wisconsin-Milwaukee (UWisconsinM),²¹ and the University of Sydney (USyd)²⁰ submitted hybrid systems. These teams also obtained, compiled, and used vocabularies; however, they utilized these resources with various machine learning approaches. University of Utah³⁰ compiled a knowledge base for medical reasons and extended it with software components and databases that included OpenNLP, MetaMap, and UMLS. University of Wisconsin-Milwaukee trained conditional random fields (CRFs) and combined its output with output from a rule-based system for determining medication names. They used Adaboost trained on annotated data for pairing medication names with their doses, modes, frequencies, durations, etc. They used support vector machines (SVMs) to distinguish narrative and list entries.²¹ University of Sydney combined CRFs with SVMs and rules. They used a CRF to identify entities. They relied on SVMs to determine whether two entities were related. They used pattern matching rules to supplement the detection of fields, to assign the scope of negation, and to declare entries as list or narrative.²⁰

RESULTS AND DISCUSSION

We ranked the system outputs on the basis of horizontal metrics.

Horizontal results and discussion

Tables 3 and 4 (and tables 5 and 6 available online at <http://jamia.bmj.com>) show phrase-level and token-level horizontal performances of complete system outputs, outputs that represent only

Table 3 Phrase-level horizontal evaluation: overall, narrative, and list

Rank	Group (external resources, medical experts)	Phrase-level horizontal evaluation								
		Overall			Narrative			List		
		Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
1	USyd (N,Y)	0.896	0.820	0.857	0.685	0.63	0.656	0.914	0.835	0.873
2	Vanderbilt (Y,Y)	0.840	0.803	0.821	0.571	0.606	0.588	0.901	0.814	0.855
3	Manchester (N,N)	0.864	0.766	0.812	0.692	0.542	0.608	0.858	0.805	0.831
4	NLM (N,N)	0.784	0.823	0.803	0.54	0.623	0.579	0.861	0.86	0.861
5	BME-Humboldt (N,N)	0.841	0.758	0.797	0.505	0.576	0.538	0.894	0.701	0.786
6	OpenU (N,N)	0.850	0.748	0.796	0.4	0.585	0.475	0.903	0.536	0.673
7	UParis (N,N)	0.799	0.761	0.780	0.336	0.482	0.396	0.765	0.552	0.641
8	LIMSI (N,N)	0.827	0.725	0.773	0.491	0.535	0.512	0.887	0.685	0.773
9	UofUtah (N,Y)	0.832	0.715	0.769	0.504	0.531	0.517	0.859	0.657	0.744
10	UWisconsinM (N,N)	0.904	0.661	0.764	0.366	0.405	0.384	0.931	0.51	0.659

N and Y in parentheses indicate whether the system was declared to have used external resources or benefitted from medical experts. N, No; Y, Yes.

Table 4 Statistical significance of the overall phrase-level horizontal F-measure differences of the top 10 systems from each other

	Vanderbilt	Manchester	NLM	BME-Humboldt	OpenU	UParis	LIMSI	UofUtah	UWiscnsinM
USyd	*	*	*	*	*	*	*	*	*
Vanderbilt		NS	*	NS	NS	*	*	*	NS
Manchester			*	*	*	*	*	*	*
NLM				NS	NS	*	*	*	NS
BME-Humboldt					NS	*	*	*	NS
OpenU						*	*	*	NS
UParis							*	*	NS
LIMSI								NS	NS
UofUtah									NS

*Indicates that the system along the row is significantly different from the system along the column at $p=0.05$.

NS, 'not significant difference' at $p=0.05$.

Please see section on Systems for an explanation of the acronyms of the team names.

list entries, and outputs that represent only narrative entries. The narrative entries are obtained from the complete system outputs by filtering out the entries that are marked 'list' for list/narrative field. For both phrase-level and token-level evaluation, performance on list entries was higher than performance on narrative entries.

Table 3 (and table 5 available online at <http://jamia.bmj.com>) show that the University of Sydney's system ranked first in both phrase-level and token-level horizontal evaluation. Table 4 (and table 6 available online at <http://jamia.bmj.com>) shows that their system performed significantly differently from the other nine systems. Vanderbilt University's system ranked second in both phrase-level and token-level horizontal evaluation. However, it was not significantly different from the University of Manchester (Manchester), BME-Humboldt, the Open University, and the University of Wisconsin-Milwaukee systems.

Tables 3 and 4 show that systems whose performances vary greatly in terms of the horizontal phrase-level F-measures may not be significantly different from each other. For instance, despite having a lower phrase-level horizontal F-measure than all other top 10 systems, University of Wisconsin-Milwaukee's system is statistically indiscernible from all but two systems, including one of the top three. In terms of the phrase-level horizontal F-measures, the only systems to perform significantly differently from all systems that scored below them came from the University of Sydney and the University of Manchester.

Figures 4 and 5 (available online at <http://jamia.bmj.com>) plot the recall and precision of the top 10 systems. University of Wisconsin-Milwaukee's system shows the highest precision and NLM's system gives the highest recall in both phrase-level and token-level horizontal evaluation.

Vertical results and discussion

Tables 7–10 (available online at <http://jamia.bmj.com>) show the vertical evaluation of the top 10 systems. In both phrase-level and token-level vertical evaluation, University of Sydney's system had the best performance in medications, dosages, modes, and frequencies; University of Manchester's system had the best performance in durations and reasons. In general, the top 10 teams achieved more than 0.75 for F-measures in medications, dosages, modes, and frequencies. On the other hand, the teams performed poorly on durations and reasons, with F-measures ranging from 0.03 to 0.525 in these two fields.

Tables 7 and 9 (available online at <http://jamia.bmj.com>) show that duration and reason were the hardest fields to extract for all teams. Our analysis showed that durations and reasons showed much variation in their content, were greater in their

length (as measured by number of tokens) than the rest of the fields, and also showed greater variability in length (see expanded Section 5.2 in online supplements), which made them difficult to capture.³¹

CONCLUSION

The Third i2b2 Workshop on Natural Language Processing Challenges for Clinical Records attracted 20 international teams and tackled a complex set of information extraction problems. The results of this challenge show that the state-of-the-art NLP systems perform well in extracting medication names, dosages, modes, and frequencies. However, detecting duration and the reason for medication events remains a challenge that would benefit from future research.

Acknowledgments This work was supported in part by the NIH Roadmap for Medical Research, Grant U54LM008748 in addition to grants 2 T15 LM007442-06, 5 U54 LM008748, and 1K99LM010227-0110; and grants T15 LM07442, HHSN 272200700057 C, and N00244-09-1-0081. IRB approval was granted for the studies presented in this manuscript. We thank all participating teams for their contributions to the challenge, University of Washington and Harvard Medical School for their technical support, and AMIA for co-sponsoring the workshop.

Funding NSF, NIH. The project described was supported in part by the i2b2 initiative, Award Number U54LM008748 from the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

Ethics approval This study was conducted with the approval of the Harvard Partners Health Care, MIT, SUNY, University of Washington.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Breydo EM, Chu JT, Turchin A. Identification of inactive medications in narrative medical text. *AMIA Annu Symp Proc* 2008;66–70.
2. Turchin A, Shubina M, Breydo E, et al. Comparison of information content of structured and narrative text data sources on the example of medication intensification. *J Am Med Inform Assoc* 2009;16:362–70.
3. Grishman R, Sundheim B. *Message understanding conference-6: A brief history. 16th Conference on Computational linguistics (COLING)*. 1996:466–71.
4. Sparck Jones K. Reflections on TREC. *Information Processing Management* 1995;31:291–314.
5. Uzun O, Solti I, Xia F, et al. Community Annotation Experiment for Ground Truth Generation for the i2b2 Medication Challenge *J Am Med Inform Assoc*. In current issue.
6. Uzun O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc* 2007;14:550–63.
7. Uzun O, Goldstein I, Luo Y, et al. Identifying patient smoking status from medical discharge summaries. *J Am Med Inform Assoc* 2008;15:14–24.
8. Uzun O. Recognizing obesity and co-morbidities in sparse data. *J Am Med Inform Assoc* 2009;16:561–70.
9. Pestian JP, Brew C, Matykiewicz P, et al. Shared task involving multi-label classification of clinical free text. *Proceedings of ACL: BioNLP* 2007:97–104.
10. Hirschman L, Yeh A, Blaschke C, et al. Overview of BioCreAtive: Critical assessment of information extraction for biology. *BMC Bioinformatics* 2005;6(Suppl 1):S1. Epub 2005 May 24.

11. **Hersh W**, Bhupatiraju RT, Corley S. Enhancing access to the bibliome: The TREC genomics track. *MEDINFO 2004*; **11**(Pt 2):773–7.
12. **Evans DA**, Brownlow ND, Hersh WR, *et al.* Automating concept identification in the electronic medical record: An experiment in extracting dosage information. *Proc AMIA Annu Fall Symp* 1996:388–92.
13. **Gold S**, Elhadad N, Zhu X, *et al.* Extracting structured medication event information from discharge summaries. *Proc AMIA Annu Fall Symp* 2008:237–41.
14. **Sirohi E**, Peissig P. Study of effect of drug lexicons on medication extraction from electronic medical records. *Pac Symp Biocomput* 2005:308–18.
15. **Levin MA**, Krol M, Doshi AM, *et al.* Extraction and mapping of drug names from free text to a standardized nomenclature. *Proc AMIA Annu Fall Symp* 2007:438–42.
16. **Jagannathan V**, Mullett CJ, Arbogast JG, *et al.* Assessment of commercial NLP engines for medication information extraction from dictated clinical notes. *Int J Med Inform* 2009;**78**:284–91.
17. **Vorhees EM**. Overview of the TREC 2003 question answering track. *The Twelfth Text REtrieval Conference (TREC 2003)*, National Institute of Standards and Technology (NIST) 2003:1–13.
18. **Chinchor N**, Hirschman L, Lewis DD. Evaluating message understanding systems: an analysis of the third message understanding conference (MUC-3). *Comput Ling* 1993;**19**:409–49.
19. **Noreen EW**. *Computer intensive methods for testing hypotheses: an introduction*. John Wiley & Sons New York, 1989.
20. **Patrick J**, Min L. A cascade approach to extract medication event (i2b2 challenge 2009). *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
21. **Li Z**, Cao Y, Antieau L, *et al.* Extracting medication information from patient discharge summaries. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
22. **Teichert AR**, Tabet JS, DuVall SL. System description. VA Salt Lake City Health Care System. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
23. **Solt I**, Tikk D. Yet another rule-based approach for extracting medication information from discharge summaries. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
24. **Grouin C**, Deleger L, Zweigenbaum P. A simple rule-based medication extraction system. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
25. **Mork JG**, Bodenreider O, Demner-Fushman D, *et al.* NLM's i2b2 tool system description. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
26. **Yang HA**. Linguistic Approach for medication extraction from medical discharge. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
27. **Hamon T**, Grabar N. Concurrent linguistic annotations for identifying medication names and the related information in discharge summaries. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
28. **Spasic I**, Sarafraz F, Keane JA, *et al.* Medication information extraction with linguistic pattern matching and semantic rules. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
29. **Doan S**, Bastarache L, Klimkowski S, *et al.* Vanderbilt's system for medication extraction. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
30. **Meystre SM**, Thibault J, Shen S, *et al.* Description of the Textractor system for medications and reason for their prescription extraction from clinical narrative text documents. *Proceedings of the Third i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, 2009.
31. **Halgrim SR**. *A pipeline machine learning approach to biomedical information extraction (Master's thesis, University of Washington Special Collections, Suzz/Allen & Auxiliary Locs P25 Th60212)*. Seattle: University of Washington, 2009.