



Published in final edited form as:

J Biomed Inform. 2015 December ; 58(Suppl): S1–S5. doi:10.1016/j.jbi.2015.10.007.

Practical Applications for Natural Language Processing in Clinical Research: The 2014 i2b2/UTHealth Shared Tasks

Özlem Uzuner and

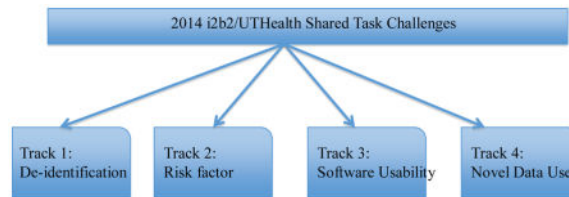
Department of Information Studies, State University of New York at Albany, Albany, New York USA

Amber Stubbs

School of Library and Information Science, Simmons College, Boston, Massachusetts USA

Özlem Uzuner: ouzuner@albany.edu; Amber Stubbs: stubbs@simmons.edu

Graphical Abstract



i2b2 (Informatics for Integrating Biology and the Bedside) is a National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 center was funded by the National Institutes of Health, from 2004 to 2014, to build a “scalable framework” and resources for enabling researchers to use existing data for investigation of diseases with genetic origins. The framework and resources built by i2b2 investigators have “enjoyed wide international adoption by the CTSA network, academic health centers, and industry”¹.

Secondary use of electronic health records (EHRs) constituted one of the cornerstones of i2b2. Over the years, i2b2 investigators built various corpora in support of Natural Language Processing (NLP) research that aimed to extract meaningful information from narrative health records in order to enable their use for disease discovery and investigation. These corpora were developed with two goals in mind:

1. To enable progress in the field of clinical NLP by advancing the state of the art.
2. To enable clinical applications of NLP.

Conflict of interest

The authors declare that they have no conflict of interest that can interfere with this work. OU is on the advisory board of NarrativeDx, a startup specialized in patient feedback on hospital experiences.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

¹<https://www.i2b2.org>

The development of NLP systems for EHRs is contingent upon access to EHRs. However, patient privacy and confidentiality concerns limit this access for research purposes. Furthermore, the relevant research is often conducted by privileged parties and on EHRs of a specific institution. EHRs differ from one institution to another and even from one department to another within the same institution. As a result, systems developed on different EHRs cannot be fairly compared, and learning from these systems and building on them (and advancing the state of the art) becomes difficult.

Given its goal to advance the state of the art in clinical NLP, i2b2 investigators focused on:

- Improving access to EHRs for NLP research by generating annotated corpora and organizing shared tasks, and
- Conducting systematic head-to-head evaluation of NLP systems for drawing conclusions from the state of the art.

The diseases under investigation at i2b2 guided the development of the annotated corpora. Over the years, i2b2 investigators have studied airways diseases [1], type 2 diabetes [2], and obesity [3], among others. They built a large body of knowledge and resources that improved our understanding of these diseases, and they built annotated corpora that spurred advancements in clinical NLP in support of clinical studies in (a subset of) these diseases [4, 5].

i2b2 investigators (in particular, the authors of this editorial with support from the rest of the i2b2 investigators) shared their annotated corpora with the research community within the scope of annual shared-task challenges that enabled evaluation of the state of the art. After issuing a call for participation, we received significant global interest in the annotated corpora, from both academia and industry. In response, we distributed each of the corpora to the interested researchers (with a data use agreement) in two parts: a training set that enabled the development of systems on the annotated corpora, and a held-out test set that was only available to the researchers at the end of the development period (typically a duration of two to three months) and on which the researchers generated outputs (within a two to three day time frame) from their systems for submission to us. We evaluated the system outputs, identified the state of the art, and organized annual workshops, co-sponsored by the American Medical Informatics Association, to disseminate information about the identified state of the art. Findings from these workshops have been published in top journals such as the Journal of Biomedical Informatics and the Journal of the American Medical Informatics Association [4,5, 6, 7, 8, 9, 10].

Starting in 2006 with a shared task in de-identification (i.e., identifying and removing protected health information from EHRs) [6] and a task in smoking status classification (i.e., determining from narrative EHRs whether a patient is a smoker or not, in support of investigation of airway diseases) [4], we built annotated corpora and organized shared tasks for predicting obesity diagnosis [5], and creating structured lists of medications [7]. We complemented these shared tasks on clinical applications of NLP with building-block tasks, such as concept extraction, assertion classification, relation extraction (i.e., finding the key medical concepts such as diseases, whether they are stated to be present or absent in a patient, and how they relate to medications mentioned in the narratives) [8], coreference

resolution (i.e., whether two mentions refer to the same concept) [9], and temporal relations (i.e., creating a partial timeline of key events mentioned in a narrative) [10]. These shared tasks addressed progressively more complex, and progressively more useful cutting-edge clinical NLP tasks and applications, and each one stood on its own in identifying the state of the art in a key NLP task.

The 2014 i2b2 Shared Task and Workshop on Challenges in Natural Language Processing for Clinical Data was the last of the i2b2 shared-task series. As such, its goal was to revisit some key NLP tasks and products from past i2b2 challenges, to put the lessons learned from the past i2b2 challenges to use for a clinical application, and to provide a sense of the overall impact of seven years of i2b2 challenges. We organized this shared task in collaboration with University of Texas Health Science Center in Houston (UTHealth).

The 2014 shared task asked four fundamental questions:

1. Given a task such as de-identification, do NLP systems generalize and robustly respond to variations in the nature of the task and in the complexity of the data?
2. Given the past i2b2 shared tasks, can we put the product of years of i2b2 shared tasks to good use for a clinical application?
3. Are the systems developed for the previous i2b2 shared tasks usable and do they support a repository of tools that future generations can build on?
4. As has been the case with the past shared-task corpora, is the 2014 i2b2/UTHealth corpus useful for purposes beyond those for which it was explicitly designed and can it support future research on new questions?

The 2014 i2b2/UTHealth shared task answered these questions by setting up four separate tracks, with four separate research goals: the de-identification track (Track 1) focused on the ability of systems to robustly respond to variations in this task; the risk factor track (Track 2) pulled together elements from past i2b2 challenges to identify risk factors that would allow assessment of the progression towards heart disease in diabetic patients; the software usability assessment track (Track 3) studied systems from all past i2b2 challenges for adoptability and usability; and the novel data use track (Track 4) opened the 2014 i2b2/UTHealth data for novel uses that went beyond the original intent behind the design of this corpus.

The 2014 i2b2/UTHealth shared task was conducted on a brand new and novel corpus of longitudinal records [11] drawn from the Research Patient Data Repository of Partners Healthcare. This corpus was first de-identified for release to the research community. A significant by-product of this de-identification process is that, given that the annotation was performed systematically and in sufficient detail [12], the data not only satisfied the Health Insurance Portability and Accountability Act (HIPAA) but they also provided a gold standard for the de-identification task, posed as Track 1 and also referred to as the “de-identification track”, of the 2014 i2b2/UTHealth shared task [13].

Putting the 2014 i2b2/UTHealth corpus to use for a clinical application, Track 2, i.e., the “risk factor track” of the 2014 i2b2/UTHealth shared task, aimed to discover the information in EHR narratives about progression towards heart disease [14]. Adequately addressing this task required insights from concept extraction, assertion classification, medication extraction, diagnosis classification, smoking status classification, and family history extraction, most of which had been addressed in past i2b2 shared tasks.

The de-identification and risk factor tracks (Tracks 1 and 2) of the 2014 i2b2/UTHealth shared tasks utilized longitudinal data that represent 2–5 records from 2–5 separate hospital visits of three cohorts of patients: those who have a diagnosis of Coronary Artery Disease (CAD) starting from the first record included in the shared-task corpus; those who do not have a diagnosis of CAD in their first record included in the shared-task corpus but get a diagnosis of CAD during their documented encounters; and those who do not have a diagnosis of CAD in any of their records included in the corpus [11, 14, 15].

In contrast to the de-identification and risk factor tracks, Tracks 3 and 4, also called the “software usability track” and the “novel data use track”, respectively, were not data and task specific. The software usability track aimed to assess the usability of systems developed for any of the past i2b2 shared tasks since 2006 [16]. The novel data use track, on the other hand, built on the observation that past i2b2 corpora have often been successfully put to use for purposes outside of their original goals and opened the 2014 shared-task corpus to any research project that fit the participants’ existing goals.

Collectively, the four tracks of the 2014 i2b2/UTHealth shared task assessed the impact of seven years of i2b2 challenges in terms of generalizing NLP methods to variations of a task, providing tools that can support clinical applications, and generating software and data that make long term contributions to the field of clinical NLP.

Track 1: De-identification Track

The task of de-identification was originally tackled in the 2006 i2b2 shared task and focused on the extraction of seven types of Protected Health Information (PHI) that frequently occurred in discharge summaries: patient names, doctor names, hospital names, identification (ID) numbers, dates, locations, and phone numbers [6]. The de-identification track of the 2014 shared task posed a more advanced and more complex version of this problem: the 2014 de-identification track included again seven high level PHI, but this time each of the PHI had to be further classified into more granular categories. For example, ID numbers needed to be broken into social security numbers, medical record numbers, health plan numbers, account numbers, license numbers, vehicle IDs, device IDs, biometric IDs, or other generic ID numbers [12]. This granularity was necessitated by the desire to assess NLP systems on all HIPAA categories (not just those prevalent in one data set) while paying extra attention to the longitudinal nature of the 2014 shared-task records, which while relatively benign when de-identified individually, could enable the deduction of sensitive information about patients when used collectively [12, 13].

In addition to being longitudinal, the 2014 data were raw in their formatting, with line breaks and tables that broke the flow of the text and did not always maintain individual tokens. This raw nature of the 2014 data reflected the realities of what the de-identification systems would have to handle in real life. In contrast, the 2006 de-identification challenge data had been manually sentence broken (i.e., broken into individual sentences so that each sentence occupies a single line of text) and tokenized (i.e., so that the punctuation is separated from words and each word stands on its own).

Despite the more complex nature of the 2014 shared task, NLP systems performed extremely well in 2014, as they did in 2006. For comparison, the F-measure of the top system in 2014 was 0.976 versus an F-measure of 0.981 in 2006. In general, the methods employed for the 2006 version of the de-identification task could respond to the more complex 2014 version of the same task: again Conditional Random Fields (CRFs), a machine learning method which has a proven track record in discovering phrases, characterized the most successful approaches [17, 18, 19, 20] and were usually accompanied with hand-built rules and/or dictionaries. Very few systems deviated from CRFs and still obtained good results [21].

The robustness of these methods across data sets is encouraging for NLP. At a time when systems are only developed on individual data sets and the generalizability of the approaches still remains in question, the results of the de-identification task indicate that the state-of-the-art approaches survive and succeed, even when the complexity of the data and the complexity of the task increase.

Track 2: Risk Factor Track

Real-life, practical clinical NLP tasks tend to be complex and to draw from multiple distinct NLP tasks. The risk factor track of the 2014 i2b2/UTHealth shared task is one example of real-life, practical clinical NLP tasks. It draws from concept extraction (it requires individual risk factors such as blood pressure measurements to be identified), assertion classification (it requires that risk factors that are only asserted to be present in the patient be marked), diagnosis extraction (it requires existing diseases of the patient to be found, e.g., diabetes), medication extraction (it needs the list of medications to be identified, especially if they are used for certain conditions that are risk factors), smoking history extraction (so that life style choices that compound the risk of heart disease are found), family history extraction (so that the hereditary nature of heart disease can be taken into consideration), and it needs all of these to be put into context with respect to their timing in relation to the current hospital visit. By addressing these tasks on longitudinal records, the risk factor track enables the construction of a timeline that demonstrates the progression towards CAD, in a patient [14].

Capturing such rich information about patients from longitudinal records can result in a very complex annotation task. The richness of the annotation task makes it also error prone and costly, both in terms of time and labor. The 2014 i2b2/UTHealth shared-task corpus addresses these challenges by employing “light annotations” that enable the experts to make correct judgments on individual records and balances the annotator workload with the

quality of the annotations, demonstrating that light annotations are both effective for NLP purposes and more easily attainable in terms of time and cost [15].

When a task is this rich, even the efforts to solve a small portion of the problem require significant effort and tuning [22, 23] and some approaches to such a task require that the light annotations be supplemented with fine-grained information [24]. As a result, there is not a single approach that can dominate the overall solution and the diversity of the solutions is limited only by the experience and creativity of their authors. While existing tools enable the rapid development of good quality solutions [25, 26], improvement on these solutions is possible by carefully evaluating each step of the solution in terms of their contribution to their own purpose and the overall task [27, 28], and by designing more task-specific solutions.

In their specific solutions, Chang et al. [29] handled different record types separately, through different classifiers and feature sets. Chen et al. [30], on the other hand, took a more fine-grained approach to modularizing the task: they tackled directly-asserted risk factors separately from those that required some inference. They also divided the directly-asserted risk factors into two, based on the syntactic complexity of the phrasing: token-based and phrase-based. Torii et al. [31] partitioned the text based on “hot spots” that would support the extraction of a specific risk factor, and they built classifiers for each type of risk factor. Yang and Garibaldi [32] designed separate systems, some rule-based and some based on machine learning, for each component of the task, without splitting the data based on its characteristics. Karystianis et al. [33] built a solution solely using rules.

Track 3: Software Usability Assessment Track

Clinical NLP systems are often designed with the intention of applying them in real-life settings; however, their assessment primarily focuses on the quality of the NLP output. Given the intent to utilize and implement these systems widely, usability of the systems should also be assessed. Track 3 of the 2014 i2b2/UTHealth shared task was organized for this purpose and was open to all systems that were developed on any i2b2 data since 2006. This track also exemplified years of i2b2 investigator’s commitment to outreach and education: The evaluators of the systems were graduate students from Master’s of Health Informatics Program in the University of Michigan. They each had a projected career path that would lead to the IT departments of healthcare organizations. They performed these evaluations as a part of their course requirements in Human Computer Interaction.

Eight systems were submitted to this track, five of which were thoroughly evaluated.² Two of these systems were on concept extraction and understanding, two were on medication extraction, and one was on de-identification. Zheng et al. [16] describes these systems and their evaluation in detail, with one major take away that affects all NLP systems in the clinical domain: the long pipeline of preprocessing components, from tokenizers to metathesauri, that are essential to most NLP goals reduce the adoptability and portability of

²The remaining three were dropped because “one was withdrawn before the evaluations started, a second was not an NLP system, and the third was a software library that did not have a user interface.” [16]

systems, especially if the systems are to be used by novices. While these preprocessing components cannot be excluded from NLP systems, they can be standardized in their input and output formats to allow some degree of interchangeability so that each new system does not come with a completely new set of preprocessing components.

Track 4: Novel Data Use Track

Over the years, i2b2 data have been put to use for purposes outside of their original goals. For example, Mishra et al. [34] utilized the i2b2 obesity challenge data for diabetes case detection and ABCs (A1c, blood pressure, cholesterol) protocol compliance assessment, Fan et al. [35] extended the annotations on the i2b2/VA relation extraction challenge to syntactic parsing, Bejan et al. [36] utilized the concepts and assertions annotations from the relations extraction challenge to specifically study the role of assertion classification in phenotype extraction, Galescu and Blaylock [37] annotated the relation extraction challenge data for temporal relations before i2b2 investigators organized their own temporal relations challenge, they also used the same data to evaluate a general domain semantic parser [38].

Given the use of past i2b2 challenge data for purposes beyond their design, Track 4, the novel data use track of the 2014 i2b2/UTHealth shared task, opened the 2014 challenge data to all research projects. The participants were given only two months, the usual development time given to all of the other tracks of the shared task, to develop systems (and annotations, as appropriate) that built on these data.

Even in such a limited time, we received five responses to this track. Solomon and Nielsen [39] utilized the heart disease risk factors data to predict changes in systolic blood pressure, Jonnagaddala et al. [40] used these data to assess the risk of CAD as predicted by the Framingham risk scores, Shivade et al. [41] studied the records for mapping patients to eligibility criteria for clinical trials, Ling et al. [42] built methods to visualize the data, and Grouin [43] developed annotations and methods for identifying medication side effects. Naturally, as has been the case for the earlier i2b2 data, we expect that the longer the data remain available for research, the more versatile the uses and the higher the number of publications they facilitate.

Summary

The 2014 i2b2/UTHealth shared tasks featured four tracks that collectively demonstrate that NLP systems can be adapted and generalized from the simpler to more complex versions of the same problem, and can be combined creatively to respond to new problems. These systems, while cutting edge in terms of their NLP performance, often suffer from structural and computational complexity, which affects their adoption and acceptance by new users. This presents a challenge for the whole field of clinical NLP to address as it moves forward.

The various i2b2 NLP corpora, developed and distributed for research since 2006, have over the years supported the development of many systems, some of which have creatively used the data for purposes outside of their original intent. The novel data use track of the 2014 shared task, along with evidence from past i2b2 challenges, shows that these data sets are valuable and will continue to find uses in the research community, even in the absence of

new annotations from i2b2 investigators. The data sets can be accessed through <https://i2b2.org/NLP/DataSets> with a data use agreement that enables their continued use for research.

Acknowledgments

We would like to thank the participants of the 2014 i2b2/UTHealth NLP Shared Task. Special thanks go to program committee members: Hua Xu, co-chair, University of Texas, Houston; John Aberdeen, MITRE; Susanne Churchill, Partners Healthcare; Cheryl Clark, MITRE; Dina Demner Fushman, NIH/NLM; Joshua Denny, Vanderbilt University; Bill Hersh, Oregon Health and Science University; Lynette Hirschman, MITRE; Issac Kohane, Partners Healthcare; Vishesh Kumar, Massachusetts General Hospital; Anna Rumshisky, UMass Lowell; Stanley Shaw, Massachusetts General Hospital; Peter Szolovits, MIT; Meliha Yetisgen, University of Washington; and Kai Zheng, University of Michigan.

Funding for this project was provided by: NIH NLM 2U54LM008748, PI: Isaac Kohane; NIH NLM 5R13LM011411, PI: Ozlem Uzuner; and NIH NIGMS 5R01GM102282, PI: Hua Xu.

References

- Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc.* 2009; 16(3):371–9. [PubMed: 19261943]
- Liu M, Liberzon A, Kong SW, Weil RL, Park PJ, Kohane IS, Kasif S. Network-based analysis of affected biological processes in Type 2 diabetes models. *PLoS Genetics.* 2007; 3:0001–0015. DOI: 10.1371/journal.pgen.0030096
- Pihlajamaki J, Itkonen P, Crunkhorn S, Vanttinen M, Dearie F, Boes T, Jimenez-Chillaron J, Lappalainen T, Miettinen P, Park P, Nasser I, Goldfine AB, Laakso M, Patti ME. Expression of splicing factor genes is reduced in human obesity: links to altered Lipin 1 splicing and enhanced lipogenesis. *Cell Metabolism.* 2011 Aug 3; 14(2):208–218. [PubMed: 21803291]
- Uzuner Ö, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc.* Jan; 2008 15(1):14–24. [PubMed: 17947624]
- Uzuner Ö. Recognizing obesity and co-morbidities in sparse data. *J Am Med Inform Assoc.* Jul; 2009 16(4):561–570. [PubMed: 19390096]
- Uzuner Ö, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc.* 2007; 14(5):550–63. [PubMed: 17600094]
- Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. *J Am Med Inform Assoc.* 2010; 17:514–8. [PubMed: 20819854]
- Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions and relations in clinical text. *J Am Med Inform Assoc.* 2011; 18:552–6. [PubMed: 21685143]
- Uzuner Ö, Bodnari A, Shen S, Forbush T, Pestian J, South BR. Evaluating the state of the art in co-reference resolution for electronic medical records. *J Am Med Inform Assoc.* 2012; 19(5):786–91. [PubMed: 22366294]
- Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *J Am Med Inform Assoc.* 2013; 20:806–13. [PubMed: 23564629]
- Kumar V, Stubbs A, Shaw S, Uzuner Ö. Creation of a new longitudinal corpus of clinical narratives. *J Biomed Inform.* this issue.
- Stubbs A, Uzuner Ö. Annotating longitudinal clinical narratives for de-identification: the 2014 i2b2/UTHealth Corpus. *J Biomed Inform.* this issue.
- Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. *J Biomed Inform.* this issue.
- Stubbs A, Kotfila C, Xu H, Uzuner Ö. Identifying risk factors for heart disease over time: Overview of 2014 i2b2/UTHealth shared task Track 2. *J Biomed Inform.* this issue.
- Stubbs A, Uzuner Ö. Annotating risk factors for heart disease in clinical narratives for diabetic patients. *J Biomed Inform.* this issue.

16. Zheng K, Vydiswaran VGV, Liu Y, Wang Y, Stubbs A, Uzuner O, Gururaj AE, Bayer S, Aberdeen J, Rumshisky A, Pakhomov S, Liu H, Xu H. Ease of adoption of clinical natural language processing software: An evaluation of five systems. *J Biomed Inform.* this issue.
17. Yang H, Garibaldi JM. Automatic detection of protected health information from clinic narratives. *J Biomed Inform.* this issue.
18. He B, Guan Y, Cheng J, Cen K, Hua W. CRFs based de-identification of medical records. *J Biomed Inform.* this issue.
19. Liu Z, Chen Y, Tang B, Wang X, Chen Q, Li H, Wang J, Deng Q, Zhu S. Automatic de-identification of electronic medical records using token-level and character-level conditional random fields. *J Biomed Inform.* this issue.
20. Dehghan A, Kovacevic A, Karystianis G, Keane JA, Nenadic G. Combining knowledge- and data-driven methods for de-identification of clinical narratives. *J Biomed Inform.* this issue.
21. Chen T, Cullen R, Godwin M. Hidden Markov Model using Dirichlet Process for de-identification. *J Biomed Inform.* this issue.
22. Kotfila C, Uzuner Ö. A systematic comparison of feature space effects on disease classifier performance for phenotype identification of five diseases. *J Biomed Inform.* this issue.
23. Shivade C, Malewadkar P, Fosler-Lussier E, Lai AM. Comparison of UMLS terminologies to identify risk of heart disease using clinical notes. *J Biomed Inform.* this issue.
24. Roberts K, Shooshan SE, Rodriguez L, Abhyankar S, Kilicoglu H, Demner-Fushman D. The role of fine-grained annotations in supervised recognition of risk factors for heart disease from EHRs. *J Biomed Inform.* this issue.
25. Cormack J, Nath C, Milward D, Raja K, Jonnalagadda SR. Agile text mining for the 2014 i2b2/UTHealth Cardiac risk factors challenge. *J Biomed Inform.* this issue.
26. Khalifa A, Meystre S. Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes. *J Biomed Inform.* this issue.
27. Grouin C, Moriceau V, Zweigenbaum P. Combining glass box and black box evaluations in the identification of heart disease risk factors and their temporal relations from clinical records. *J Biomed Inform.* this issue.
28. Urbain J. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. *J Biomed Inform.* this issue.
29. Chang N-W, Dai H-J, Jonnagaddala J, Chen C-W, Tsai RT-H, Hsu W-L. A Context-aware approach for progression tracking of medical concepts in electronic medical records. *J Biomed Inform.* this issue.
30. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, Liu S, Wang W, Deng W, Zhu S, Chen Y, Wang J. An automatic system to identify heart disease risk factors in clinical texts over time. *J Biomed Inform.* this issue.
31. Torii M, Fan J-W, Yang W-L, Lee T, Wiley MT, Zisook DS, Huang Y. Risk factor detection for heart disease by applying text analytics in electronic medical records. *J Biomed Inform.* this issue.
32. Yang H, Garibaldi JM. A hybrid model for automatic identification of risk factors for heart disease. *J Biomed Inform.* this issue.
33. Karystianis G, Dehghan A, Kovacevic A, Keane JA, Nenadic G. Using local lexicalized rules to identify heart disease risk factors in clinical notes. *J Biomed Inform.* this issue.
34. Mishra NK, Son RY, Arnzen JJ. Towards automatic diabetes case detection and ABCS protocol compliance assessment. *Clin Med Res.* 2012 Aug; 10(3):106–21. Epub 2012 May 25. DOI: 10.3121/cm.2012.1047 [PubMed: 22634542]
35. Fan JW, Yang EW, Jiang M, Prasad R, Loomis RM, Zisook DS, Denny JC, Xu H, Huang Y. Syntactic parsing of clinical text: guideline and corpus development with handling ill-formed sentences. *J Am Med Inform Assoc.* 2013 Nov-Dec;20(6):1168–77. Epub 2013 Aug 1. DOI: 10.1136/amiajnl-2013-001810 [PubMed: 23907286]
36. Bejan CA, Vanderwende L, Xia F, Yetisgen-Yildiz M. Assertion modeling and its role in clinical phenotype identification. *J Biomed Inform.* 2013 Feb; 46(1):68–74. Epub 2012 Sep 21. DOI: 10.1016/j.jbi.2012.09.001 [PubMed: 23000479]

37. Galescu, L.; Blaylock, N. A. corpus of clinical narratives annotated with temporal information. Proceedings of the 2nd ACM SIGHIT international Health Informatics Symposium; New York, NY, USA: ACM; 2012. p. 715-720.
38. Swift, M.; Blaylock, N.; Allen, J.; de Beaumont, W.; Galescu, L.; Jung, H. Augmenting a deep natural language processing system with UMLS. Proceedings of the Fourth International Symposium on Semantic Mining in Biomedicine (SMBM) Poster Session; October 25–26; Hinxton, Cambridgeshire, UK: European Bioinformatics Institute;
39. Solomon JW, Nielsen RD. Predicting changes in systolic blood pressure using longitudinal patient records. *J Biomed Inform.* this issue.
40. Jonnagaddala J, Liaw S-T, Ray P, Kumar M, Chang N-W, Dai H-J. Coronary artery disease risk assessment from unstructured electronic health records using text mining. *J Biomed Inform.* this issue.
41. Shivade C, Hebert C, Lopetegui M, de Marneffe M-C, Fosler-Lussier E, Lai AM. Textual inference for eligibility criteria resolution in clinical trials. *J Biomed Inform.* this issue.
42. Ling, Y.; Xingpeng, J.; An, Y.; Hu, X. Data exploration and visualization of risk factors for heart disease from medical documents using non-negative matrix factorization (NMF). Seventh i2b2 Shared Task and Workshop; Washington DC. November14, 2014;
43. Grouin, C. Identification of medication side effects in clinical records: an experiment based on the 2014 i2b2/UTHealth corpus. Seventh i2b2 Shared Task and Workshop; Washington DC. November14, 2014;