

Making Sense of Big Textual Data for Health Care: Findings from the Section on Clinical Natural Language Processing

A. Névéol, P. Zweigenbaum, Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing

LIMSI, CNRS, Université Paris Saclay, Orsay, France

Summary

Objectives: To summarize recent research and present a selection of the best papers published in 2016 in the field of clinical Natural Language Processing (NLP).

Method: A survey of the literature was performed by the two section editors of the IMIA Yearbook NLP section. Bibliographic databases were searched for papers with a focus on NLP efforts applied to clinical texts or aimed at a clinical outcome. Papers were automatically ranked and then manually reviewed based on titles and abstracts. A shortlist of candidate best papers was first selected by the section editors before being peer-reviewed by independent external reviewers.

Results: The five clinical NLP best papers provide a contribution that ranges from emerging original foundational methods to transitioning solid established research results to a practical clinical setting. They offer a framework for abbreviation disambiguation and coreference resolution, a classification method to identify clinically useful sentences, an analysis of counseling conversations to improve support to patients with mental disorder and grounding of gradable adjectives.

Conclusions: Clinical NLP continued to thrive in 2016, with an increasing number of contributions towards applications compared to fundamental methods. Fundamental work addresses increasingly complex problems such as lexical semantics, coreference resolution, and discourse analysis. Research results translate into freely available tools, mainly for English.

Keywords

Awards and prizes; decision making, computer-assisted; medical informatics/trends; natural language processing; semantics

Yarb Med Inform 2017:228-34

<http://dx.doi.org/10.15265/IY-2017-027>

Published online August 18, 2017

Introduction

Clinical natural language processing (NLP) is defined as NLP applied to clinical texts or aimed at a clinical outcome. This encompasses NLP applied to texts in Electronic Health Records (EHRs), which is the case of the bulk of information extraction for decision support or clinical research. We also consider as clinically relevant applications and research addressing the analysis of patient-authored text or speech for public health or diagnosis purposes. This year's survey paper reports on the increasing variety of texts used for clinical NLP, including social media [1]. The best papers selected this year address increasingly complex problems such as lexical semantics [2-3], coreference resolution [4], and discourse analysis [5]. In addition, they show how research results can effectively translate into freely available tools [4-6]. More specifically, the best papers offer a framework for abbreviation disambiguation [6] and coreference resolution [4], classification of clinically useful sentences [2], analysis of counseling conversations to improve support to patients with mental disorders [5], and grounding of gradable adjectives [3]. They provide a contribution that ranges from emerging original foundational methods [3] to transitioning solid established research results to a practical clinical setting [6].

Method

Papers were retrieved using the same search strategies as in 2016, relying on PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) and

the Association of Computational Linguistics (ACL) Anthology Searchbench (<http://aclasb.dfki.de/>). The PubMed query used minimal metadata and free-text keywords: (English[LA] AND journal article[PT] AND 2016[dp] AND hasabstract[text]) AND ((medical OR clinical OR natural) AND "language processing"). The ACL Anthology query restricted our selection to the most selective journals (TACL, Computational Linguistics), conferences (ACL, EMNLP, NAACL, COLING), and workshops (ACL BioNLP). It used the free text keywords `medical`, `clinical` and `health`. The collection of papers using these queries brought back 335 titles and abstracts from MEDLINE and 33 from ACL Anthology, resulting in a total selection of 368 papers.

Then, we used the results of the 2016 selection to train a logistic regression classifier to automatically rank the selected papers from most relevant to select to less relevant [7]. One section editor (AN) then used the BibReview tool (<https://pypi.python.org/pypi/BibReview>) to classify the papers based on titles and abstracts into four categories: (1) Off Topic (OT) which identified papers that focused on topics outside of the scope of clinical NLP, such as biological natural language processing, knowledge representation, psycholinguistics, or even image processing; (2) No (N) for papers that did not provide a contribution to either NLP methodology or clinical outcome. Review papers and correspondence were included in this category in order to keep only original research contributions; (3) Maybe (M) for papers that offered a contribution to NLP aimed at a clinical outcome.

Similarly to last year, papers reporting on participation to NLP challenges were included in this category because even though challenges provide valuable contributions to the field, challenge papers are usually polished working notes that report on work that has not reached the level of maturity expected from a “best paper”; (4) Yes (Y) for papers that did so outstandingly, or with high novelty.

The 31 papers initially selected in the Y category were grouped by broad topics and ranked by both section editors. The full text of the top 25 papers was then browsed to refine this selection in order to cover each of the topics while ensuring that the final selection comprised a variety of topics, authors, and venues. In the list of references provided at the end of the synopsis, a star (*) indicates papers that were in the final selection of the 15 candidate best papers.

Results

We present an overview of clinical NLP publications that cover the topics addressed by the research community in 2016. Table 1 lists the five papers selected as the best papers; a summary of each paper appears in the appendix of this synopsis.

Applications of Clinical NLP are Increasing in Number and Diversity

Applications of NLP by far outnumber research on foundational methods: 71% of clinical NLP articles reviewed reported on an application of NLP and 29% on foundational methods.

NLP tools ranging from basic keyword/dictionary-based extraction to advanced concept extraction (e.g. using cTAKES, PyContext or MEDLEE) are used for direct analysis of text of clinical interest or as part of a pipeline extracting features for state-of-the-art classification tools. These solutions are overwhelmingly applied to radiology reports, which are an easier

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2017 in the section ‘Clinical Natural Language Processing’. The articles are listed in alphabetical order of the first author’s surname.

Section
Clinical Natural Language Processing
<ul style="list-style-type: none"> ▪ Althoff, T, Clark K, Leskovec, J. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. <i>Trans Assoc Comput Linguist</i> 2016 (4):463-76. ▪ Kilicoglu, H, Demner-Fushman, D. Bio-SCoRes: A Smorgasbord Architecture for Coreference Resolution in Biomedical Text. <i>PLoS One</i> 2016 Mar 2;11(3):e0148538. ▪ Morid, MA, Fiszman, M, Raja, K, Jonnalagadda, SR, Del Fiol, G. Classification of clinically useful sentences in clinical evidence resources. <i>J Biomed Inform</i> 2016 Apr;60:14-22. ▪ Shivade C, de Marneffe MC, Fosler-Lussier E, Lai AM. Identification, characterization, and grounding of gradable terms in clinical text. <i>Proceedings of the 15th Workshop on Biomedical Natural Language Processing</i>. 2016:17-26. ▪ Wu, Y, Denny, JC, Rosenbloom, ST, Miller, RA, Giuse, DA, Wang, L, Blanquicett, C, Soysal, E, Xu, J, Xu, H. A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). <i>J Am Med Inform Assoc</i> 2017 Apr 1;24(e1):e79-e86.

source for efficient NLP than for instance discharge summaries: the latter are much more diverse and extended in terms of vocabulary, language structures, and types of information [8]. Two reviews of NLP of radiology reports were published in 2016: Cai et al., [9] included a tutorial intended for the Radiographics journal audience, while Pons et al., [10] performed a systematic review of “publications describing NLP methods that support practical applications in radiology.”

We observe a continued diversification of the types of texts used for clinical NLP work, which extends this year to resident pages [11], text message conversations [5], consumer product reviews [12], while social media and Twitter in particular continue to emerge as a strong source for public health monitoring for a number of issues including drug abuse [13], alcohol [14], adverse drug effects [15], and drug repurposing [16], and also motivate work aiming at identifying tweets that generally report a personal health experience [17].

Another emerging trend of research seeks to use NLP methods towards patient empowerment by making health literature more accessible through improved readability [18] or text simplification [19], making progress in clinical question answering for patients [20-21]. Conversely, some efforts seek insight from patients’ experiences to improve the delivery of healthcare. For instance, Hawkins et al., and Ranard et al.,

automatically extracted patient-perceived quality of care in hospitals [22-23], while Bahk et al., [24] and Strelakova et al., [25], respectively, analyzed patient sentiment on vaccination and cancer care options.

Assessing the quality of EHR content is another topic where NLP plays a role. Scholte et al., [26] noted the need for NLP to process text fields to evaluate the quality of physiotherapy care. Ford et al., [27] studied the delay in coding rheumatoid arthritis and they observed that over one fifth of patients had text entries for disease-modifying anti-rheumatic drugs more than 14 days before rheumatoid arthritis was coded. Kaufman et al., [28] assessed the feasibility of using dictation followed by NLP to enter content into the EHR.

Text Classification and Information Extraction Remain Strong Applications

A large part of the reported research aims at categorizing a patient’s clinical record into predefined categories. This is often done for phenotyping: obesity [29], axial spondyloarthritis (AxSpA) [30], childhood obesity [31], stroke risk factor [32], multiple sclerosis [33], hypertension [34], sometimes state-wide and in “real time” as in [35] for diabetes mellitus, and sometimes to distinguish between finer-grained cancer

information types as in [36]. Generic methods are designed to facilitate adaptation to new phenotypes [37].

More generally, text classification is used to identify patient records for the purpose of retrospective or prospective studies aiming at improving care pathways for a category of patients, or at characterizing patient cohorts. A wide range of topics were addressed in 2016: to predict the protocol and priority of MRI brain protocols [38], to identify heart failure patients with ineffective self-management status [39], to predict suicidal ideation and heightened psychiatric symptoms [40], to detect long-bone fractures [41], abdominal aortic aneurysms [42], liver cirrhosis [43], and the region containing an abnormality [44] in radiology reports, to predict the diagnosis of breast cancer in mammography reports [45], to identify pediatric traumatic brain injury in CT reports [46], hepatocellular cancer in pathology and radiology reports [47], cerebral aneurysms [48], first episode psychosis [49], non-alcoholic liver fatty disease [50], celiac disease [51], tree stand falls [52], as well as acute coronary syndrome from admission records in Chinese [53].

Classification is also applied to predict pre- or post-discharge events such as patients that will be medically ready for discharge from the neonatal intensive care unit in the subsequent 2-10 days [54], to predict suicide [55-56], to forecast the daily bed needs [57], to predict patients at high risk for high imaging utilization based on radiology reports [58], early psychiatric readmission [59], opioid abuse of patients considered for opioid therapy [60], and to produce a predictive risk report for hospitalized heart failure patients [61].

Compared to text categorization, information extraction focuses on specific information elements found in clinical text. This includes cancer stage in patients with lung cancer [62], average weekly doses of drugs [63], adverse events in robotic surgery [64], indwelling urinary catheter and urinary symptoms [65], liver tumor characteristics from radiology reports [66], left ventricular ejection fraction from echocardiography reports [67], wound information (wound type, pressure ulcer stage, wound size, anatomic

location, and wound treatment) from free text clinical notes [39], and congestive heart failure medication information from Veteran Administration EHRs [68]. Clinical trials attract much attention for tasks that require specific information extraction for public health research, including comparative effectiveness research [69], mapping of disease research [70], categorizing adverse events by age and type [71], or characterizing cancer drug toxicity [72].

Several efforts addressing information extraction and text classification tasks benefit from an integrated approach. For instance, Botsis et al., developed a common decision support environment for medical product safety surveillance [73], while others used NLP to build databases of clinically useful information, such as mutation-disease associations [74], drug side-effects by parsing product labels [75], and genetic alteration information in cancer trials [76]. This includes work on annotated text corpora, e.g. a corpus of tweets reporting events related to the author's own health [17], and specialized vocabularies, for instance to detect substance abuse terms [77].

Foundational Methods of Clinical NLP Take Both Innovative and Consolidating Directions

In 2016, the efforts addressing foundational methods of clinical NLP continued to explore core topics that deserve sustained attention such as entity recognition and normalization or relation extraction. Emerging topics were also addressed, including discourse or lexical semantics [3-78].

Entity extraction is one core topic of clinical NLP that still represents a challenge at the level of mention extraction as well as entity linking or normalization for types of entity such as diseases [79]. The context of utterance of entities, such as speculation, continues to be explored including in languages other than English such as Chinese [80]. While personal health identifiers are entities which have received sustained interest, research directions in the field of de-identification

are switching from entity recognition to revisiting evaluation methods [81-82] and annotation efforts optimizing de-identification efforts [83].

By linking together multiple mentions of the same entity or event, anaphora and coreference resolution may improve the quality of information extraction. Progress was made on anaphora resolution in MEDLINE abstracts [84], and led to a general toolbox for coreference resolution [4] – described in more detail below – tested on both MEDLINE abstracts and the clinical texts of the i2b2 2011 shared task. While the intrinsic performance of coreference resolution can be evaluated on the i2b2 clinical texts [84-85], its actual impact on an information extraction task is a more concrete test of its relevance [66].

The development of NLP tools is an active area fueled by continued research efforts on these hard problems: we welcome three new open-source tools for concept recognition [86], co-reference resolution [4], and temporal analysis [87], which are important components for biomedical language processing. Kaggal et al., [88] described an institutional implementation of a big data-empowered clinical NLP infrastructure.

The challenge of the lack of annotated data for training supervised systems was addressed in 2016 by distant supervision to help extract relevant sentences from articles to assist the systematic review process [89].

More work was performed to compute language-related scores based on patient speech using NLP and Machine Learning methods. Clark et al., [90] created novel scores that are better predictors of the evolution from mild cognitive impairment to Alzheimer's disease. Takano et al., [91] identified linguistic features in recalled memories that helped distinguish between specific and non-specific memories in the Autobiographical Memory Test. Luo et al., [92] discriminated between adults with autism spectrum disorder and a control group. Althoff et al., [5] analyzed doctor-patient counseling interactions in text messages through a variety of techniques and evidence actionable conversation strategies that are liable to improve counseling practice. This work is described in more detail below.

Conclusion

Clinical Natural Language Processing continued to thrive in 2016, with an increasing number of contributions towards applications compared to fundamental methods. However, we note that fundamental work addresses increasingly complex problems such as lexical semantics, coreference resolution, and discourse analysis. Research results translate into freely available tools, mainly for English. The best papers of this year illustrate these trends.

References

In the reference list below, papers that were shortlisted as best paper candidates are marked with a *.

- Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text. *Yearb Med Inform* 2017;214-27.
- * Morid MA, Fiszman M, Raja K, Jonnalagadda SR, Del Fiol G. Classification of clinically useful sentences in clinical evidence resources. *J Biomed Inform* 2016 Apr;60:14-22.
- * Shivade C, de Marneffe MC, Fosler-Lussier E, Lai AM. Identification, characterization, and grounding of gradable terms in clinical text. *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*. 2016:17-26
- * Kilicoglu H, Demner-Fushman D. Bio-SCoRes: A Smorgasbord Architecture for Coreference Resolution in Biomedical Text. *PLoS One* 2016 Mar 2;11(3):e0148538.
- * Althoff T, Clark K, Leskovec J. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Trans Assoc Comput Linguist* 2016;4:463-76.
- * Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, Wang L, et al. A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). *J Am Med Inform Assoc* 2017;24(e1):e79-e86.
- Norman C, Leeflang ML, Zweigenbaum P, Névéol A. *Tri Automatique de la Littérature pour les Revues Systématiques. Traitement Automatique de la Langue Naturelle - TALN*. 2017:234-41.
- Friedman C. Towards a comprehensive medical language processing system: methods and issues. *Proc AMIA Annu Fall Symp* 1997:595-9.
- Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural Language Processing Technologies in Radiology Research and Clinical Applications. *Radiographics* 2016 Jan-Feb;36(1):176-91.
- Pons E, Braun LM, Humink MG, Kors JA. *Natural Language Processing in Radiology: A Systematic Review*. *Radiology* 2016 May;279(2):329-43.
- Smith AD, de Vos MS, Smink DS, Nguyen LL, Ashley SW. Text paging of surgery residents: Efficacy, work intensity, and quality improvement. *Surgery* 2016 Mar;159(3):930-7.
- Torii M, Tilak SS, Doan S, Zisook DS, Fan JW. Mining Health-Related Issues in Consumer Product Reviews by Using Scalable Text Analytics. *Biomed Inform Insights* 2016 Jun 20;8(Suppl 1):1-11.
- Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. *Drug Saf* 2016 Mar;39(3):231-40.
- Ranney ML, Chang B, Freeman JR, Norris B, Silverberg M, Choo EK. Tweet Now, See You In the ED Later? Examining the Association Between Alcohol-related Tweets and Emergency Care Visits. *Acad Emerg Med* 2016 Jul;23(7):831-4.
- Liu J, Zhao S, Zhang X. An ensemble method for extracting adverse drug events from social media. *Artif Intell Med* 2016 Jun;70:62-76.
- Rastegar-Mojarad M, Liu H, Nambisan P. Using Social Media Data to Identify Potential Candidates for Drug Repurposing: A Feasibility Study. *JMIR Res Protoc* 2016 Jun 16;5(2):e121.
- Calix RA, Gupta M, Jiang K. Construction of a Personal Experience Tweet Corpus for Health Surveillance. *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*. 2016:128-35.
- Wu DT, Hanauer DA, Mei Q, Clark PM, An LC, Proulx J, et al. Assessing the readability of ClinicalTrials.gov. *J Am Med Inform Assoc* 2016 Mar;23(2):269-75.
- Kauchak D, Leroy G. Moving Beyond Readability Metrics for Health-Related Text Simplification. *IT Prof* 2016 May-Jun;18(3):45-51.
- Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of consumer and professional questions. *J Am Med Inform Assoc* 2016 Jul;23(4):802-11.
- Wongchaisuwat P, Klabjan D, Jonnalagadda SR. A Semi-Supervised Learning Approach to Enhance Health Care Community-Based Question Answering: A Case Study in Alcoholism. *JMIR Med Inform* 2016 Aug 2;4(3):e24.
- Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ Qual Saf* 2016 Jun;25(6):404-13.
- Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, et al. Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care. *Health Aff (Millwood)* 2016 Apr;35(4):697-705.
- Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly Available Online Tool Facilitates Real-Time Monitoring Of Vaccine Conversations And Sentiments. *Health Aff (Millwood)* 2016 Feb;35(2):341-7.
- Strekalova YA, James VS. Language of Uncertainty: the Expression of Decisional Conflict Related to Skin Cancer Prevention Recommendations. *J Cancer Educ* 2016 Jan 19.
- Scholte M, van Dulmen SA, Neeleman-Van der Steen CW, van der Wees PJ, Nijhuis-van der Sanden MW, Braspenning J. Data extraction from electronic health records (EHRs) for quality measurement of the physical therapy process: comparison between EHR data and survey data. *BMC Med Inform Decis Mak* 2016 Nov 8;16(1):141.
- Ford E, Carroll J, Smith H, Davies K, Koeling R, Petersen I, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. *BMJ Open* 2016 Jun 28;6(6):e010393.
- * Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural Language Processing-Enabled and Conventional Data Capture Methods for Input to Electronic Health Records: A Comparative Usability Study. *JMIR Med Inform* 2016 Oct 28;4(4):e35.
- Figueroa RL, Flores CA. Extracting Information from Electronic Medical Records to Identify the Obesity Status of a Patient Based on Comorbidities and Bodyweight Measures. *J Med Syst* 2016 Aug;40(8):191.
- Walsh JA, Shao Y, Leng J, He T, Teng CC, Redd D, et al. Identifying axial spondyloarthritis in electronic medical records of United States Veterans. *Arthritis Care Res (Hoboken)* 2016 Nov 3.
- Lingren T, Thaker V, Brady C, Namjou B, Kennebeck S, Bickel J, et al. Developing an Algorithm to Detect Early Childhood Obesity in Two Tertiary Pediatric Medical Centers. *Appl Clin Inform* 2016 Jul 20;7(3):693-706.
- Mowery DL, Chapman BE, Conway M, South BR, Madden E, Keyhani S, et al. Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis. *J Biomed Semantics* 2016 May 10;7:26.
- Nelson RE, Butler J, LaFleur J, Knippenberg K, C Kamau AW, DuVall SL. Determining Multiple Sclerosis Phenotype from Electronic Medical Records. *J Manag Care Spec Pharm* 2016 Dec;22(12):1377-82.
- Teixeira PL, Wei WQ, Cronin RM, Mo H, Van-Houten JP, Carroll RJ, et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *J Am Med Inform Assoc* 2017 Jan;24(1):162-71.
- * Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. *JMIR Med Inform* 2016 Nov 11;4(4):e37.
- * Yala A, Barzilay R, Salama L, Griffin M, Sollenberger G, Bardia A, et al. Using machine learning to parse breast pathology reports. *Breast Cancer Res Treat* 2017 Jan;161(2):203-11.
- Yu S, Chakraborty A, Liao KP, Cai T, Ananthakrishnan AN, Gainer VS, et al. Surrogate-assisted feature extraction for high-throughput phenotyping. *J Am Med Inform Assoc* 2017 Apr 1;24(e1):e143-e149.
- Brown AD, Marotta TR. A Natural Language Processing-based Model to Automate MRI Brain Protocol Selection and Prioritization. *Acad Radiol* 2017 Feb;24(2):160-166

39. Topaz M, Radhakrishnan K, Blackley S, Lei V, Lai K, Zhou L. Studying Associations Between Heart Failure Self-Management and Rehospitalizations Using Natural Language Processing. *West J Nurs Res* 2016 Sep 14.
40. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. *Comput Math Methods Med* 2016;2016:8708434.
41. Grundmeier RW, Masino AJ, Casper TC, Dean JM, Bell J, Enriquez R, et al; Pediatric Emergency Care Applied Research Network. Identification of Long Bone Fractures in Radiology Reports Using Natural Language Processing to support Healthcare Quality Improvement. *Appl Clin Inform* 2016 Nov 9;7(4):1051-68.
42. Morioka C, Meng F, Taira R, Sayre J, Zimmerman P, Ishimitsu D, et al. Automatic Classification of Ultrasound Screening Examinations of the Abdominal Aorta. *J Digit Imaging* 2016 Dec;29(6):742-8.
43. Chang EK, Yu CY, Clarke R, Hackbarth A, Sanders T, Esrailian E, et al. Defining a Patient Population With Cirrhosis: An Automated Algorithm With Natural Language Processing. *J Clin Gastroenterol* 2016 Nov/Dec;50(10):889-94.
44. Masino AJ, Grundmeier RW, Pennington JW, Germiller JA, Crenshaw EB 3rd. Temporal bone radiology report classification using open source machine learning and natural language processing libraries. *BMC Med Inform Decis Mak* 2016 Jun 6;16:65.
45. Bozkurt S, Gimenez F, Burnside ES, Gulkesen KH, Rubin DL. Using automatically extracted information from mammography reports for decision-support. *J Biomed Inform* 2016 Aug;62:224-31.
46. Yadav K, Sarioglu E, Choi HA, Cartwright WB 4th, Hinds PS, Chamberlain JM. Automated Outcome Classification of Computed Tomography Imaging Reports for Pediatric Traumatic Brain Injury. *Acad Emerg Med* 2016 Feb;23(2):171-8.
47. Sada Y, Hou J, Richardson P, El-Serag H, Davila J. Validation of Case Finding Algorithms for Hepatocellular Cancer From Administrative Data and Electronic Health Records Using Natural Language Processing. *Med Care* 2016 Feb;54(2):e9-14.
48. Castro VM, Dligach D, Finan S, Yu S, Can A, Abd-El-Barr M, et al. Large-scale identification of patients with cerebral aneurysms using natural language processing. *Neurology* 2017 Jan 10;88(2):164-8.
49. Gorrell G, Oduola S, Roberts A, Crai, T, Morgan C, Stewart R. Identifying First Episodes of Psychosis in Psychiatric Patient Records using Machine Learning. Proceedings of the 15th Workshop on Biomedical Natural Language Processing. 2016:196-205.
50. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and Validation of an Algorithm to Identify Nonalcoholic Fatty Liver Disease in the Electronic Medical Record. *Dig Dis Sci* 2016 Mar;61(3):913-9.
51. Chen W, Huang Y, Boyle B, Lin S. The utility of including pathology reports in improving the computational identification of patients. *J Pathol Inform* 2016 Nov 29;7:46.
52. VanWormer JJ, Holsman RH, Petchenik JB, Dhuey BJ, Keifer MC. Epidemiologic trends in medically-attended tree stand fall injuries among Wisconsin deer hunters. *Injury* 2016 Jan;47(1):220-5.
53. Hu D, Huang Z, Chan TM, Dong W, Lu X, Duan H. Utilizing Chinese Admission Records for MACE Prediction of Acute Coronary Syndrome. *Int J Environ Res Public Health* 2016 Sep 13;13(9).
54. Temple MW, Lehmann CU, Fabbri D. Natural Language Processing for Cohort Discovery in a Discharge Prediction Model for the Neonatal ICU. *Appl Clin Inform* 2016 Feb 24;7(1):101-15.
55. McCoy TH Jr, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. *JAMA Psychiatry* 2016 Oct 1;73(10):1064-71.
56. * Pestian JP, Sorter M, Connolly B, Brettonel Cohen K, McCullumsmith C, et al; STM Research Group. A Machine Learning Approach to Identifying the Thought Markers of Suicidal Subjects: A Prospective Multicenter Trial. *Suicide Life Threat Behav* 2017 Feb;47(1):112-21.
57. * Toerper MF, Flanagan E, Siddiqui S, Appelbaum J, Kasper EK, Levin S. Cardiac catheterization laboratory inpatient forecast tool: a prospective evaluation. *J Am Med Inform Assoc* 2016 Apr;23(e1):e49-57.
58. Hassanpour S, Langlotz CP. Predicting High Imaging Utilization Based on Initial Radiology Reports: A Feasibility Study of Machine Learning. *Acad Radiol* 2016 Jan;23(1):84-9.
59. Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry* 2016 Oct 18;6(10):e921.
60. Haller IV, Renier CM, Juusola M, Hitz P, Steffen W, Asmus MJ, et al. Enhancing Risk Assessment in Patients Receiving Chronic Opioid Analgesic Therapy Using Natural Language Processing. *Pain Med* 2016 Dec 29.
61. * Evans RS, Benuzillo J, Horne BD, Lloyd JF, Bradshaw A, Budge D, et al. Automated identification and predictive tools to help identify high-risk heart failure patients: pilot evaluation. *J Am Med Inform Assoc* 2016 Sep;23(5):872-8.
62. Warner JL, Levy MA, Neuss MN, Warner JL, Levy MA, Neuss MN. ReCAP: Feasibility and Accuracy of Extracting Cancer Stage Information From Narrative Electronic Health Record Data. *J Oncol Pract* 2016 Feb;12(2):157-8; e169-7.
63. Lu CC, Leng J, Cannon GW, Zhou X, Egger M, South B, et al. The use of natural language processing on narrative medication schedules to compute average weekly dose. *Pharmacoepidemiol Drug Saf* 2016 Dec;25(12):1414-24.
64. Alemzadeh H, Raman J, Leveson N, Kalbarczyk Z, Iyer RK. Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data. *PLoS One* 2016 Apr 20;11(4):e0151470.
65. Gundlapalli AV, Divita G, Redd A, Carter ME, Ko D, Rubin M, et al. Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients using natural language processing. *J Biomed Inform* 2016 Jul 9.
66. Yim WW, Kwan SW, Yetisgen M. Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction. *J Biomed Inform* 2016 Dec;64:179-91.
67. Xie F, Zheng C, Yuh-Jer Shen A, Chen W. Extracting and analyzing ejection fraction values from electronic echocardiography reports in a large health maintenance organization. *Health Informatics J* 2016 Jun 7.
68. Meystre SM, Kim Y, Gobbel GT, Matheny ME, Redd A, Bray BE, et al. Congestive heart failure information extraction framework for automated treatment performance measures assessment. *J Am Med Inform Assoc* 2017 Apr 1;24(e1):e40-e46.
69. Chang M, Chang M, Reed JZ, Milward D, Xu JJ, Cornell WD. Developing timely insights into comparative effectiveness research with a text-mining pipeline. *Drug Discov Today* 2016 Mar;21(3):473-80.
70. Atal I, Zeitoun JD, Névél A, Ravaud P, Porcher R, Trinquant L. Automatic classification of registered clinical trials towards the Global Burden of Diseases taxonomy of diseases and injuries. *BMC Bioinformatics* 2016 Sep 22;17(1):392.
71. Luo J, Eldredge C, Cho CC, Cisler RA. Population Analysis of Adverse Events in Different Age Groups Using Big Clinical Trials Data. *JMIR Med Inform* 2016 Oct 17;4(4):e30.
72. Luo J, Cisler RA. Discovering Outliers of Potential Drug Toxicities Using a Large-scale Data-driven Approach. *Cancer Inform* 2016 Oct 26;15:211-7.
73. Botsis T, Jankosky C, Arya D, Kreimeyer K, Foster M, Pandey A, et al. Decision support environment for medical product safety surveillance. *J Biomed Inform* 2016 Dec;64:354-62.
74. Mahmood AS, Wu TJ, Mazumder R, Vijay-Shanker K. DiMeX: A Text Mining System for Mutation-Disease Association Extraction. *PLoS One* 2016 Apr 13;11(4):e0152725.
75. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016 Jan 4;44(D1):D1075-9.
76. Xu J, Lee HJ, Zeng J, Wu Y, Zhang Y, Huang LC, et al. Extracting genetic alteration information for personalized cancer therapy from ClinicalTrials.gov. *J Am Med Inform Assoc* 2016 Jul;23(4):750-7.
77. Velupillai S, Mowery DL, Conway M, Hurdle J, Kious B. Vocabulary Development To Support Information Extraction of Substance Abuse from Psychiatry Notes. Proceedings of the 15th Workshop on Biomedical Natural Language Processing. 2016: 92-101.
78. * Pakhomov SV, Finley G, McEwan R, Wang Y, Melton GB. Corpus domain effects on distributional semantic modeling of medical terms. *Bioinformatics* 2016 Dec 1;32(23):3635-44.
79. * Limsopatham N, Collier N. Normalising Medical Concepts in Social Media Texts by Learning Semantic Representation. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016:1014-23.
80. * Zhang S, Kang T, Zhang X, Wen D, Elhadad N, Lei J. Speculation detection for Chinese clinical notes: Impacts of word segmentation and embedding models. *J Biomed Inform* 2016

- Apr;60:334-41.
81. Li M, Carrell D, Aberdeen J, Hirschman L, Kirby J, Li B, et al. Optimizing annotation resources for natural language de-identification via a game theoretic framework. *J Biomed Inform* 2016 Jun;61:97-109.
 82. * Scaiano M, Middleton G, Arbuckle L, Kolhatkar V, Peyton L, Dowling M, et al K. A unified framework for evaluating the risk of re-identification of text de-identification tools. *J Biomed Inform* 2016 Oct;63:174-83.
 83. Carrell DS, Cronkite DJ, Malin BA, Aberdeen JS, Hirschman L. Is the Juice Worth the Squeeze? Costs and Benefits of Multiple Human Annotators for Clinical Text De-identification. *Methods Inf Med* 2016 Aug 5;55(4):356-64.
 84. Kilicoglu H, Roseblat G, Fiszman M, Rindflesch TC. Sortal anaphora resolution to enhance relation extraction from biomedical literature. *BMC Bioinformatics* 2016 Apr 14;17:163.
 85. Liu S, Liu H, Chaudhary V, Li D. An Infinite Mixture Model for Coreference Resolution in Clinical Notes. *AMIA Jt Summits Transl Sci Proc* 2016 Jul 22;2016:428-37.
 86. Tseytlin E, Mitchell K, Legowski E, Corrigan J, Chavan G, Jacobson RS. NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinformatics* 2016 Jan 14;17:32.
 87. Lin C, Dligach D, Miller TA, Bethard S, Savova GK. Multilayered temporal modeling for the clinical domain. *J Am Med Inform Assoc* 2016 Mar;23(2):387-95.
 88. Kaggal VC, Elayavilli RK, Mehrabi S, Pankratz JJ, Sohn S, Wang Y, et al. Toward a Learning Healthcare System - Knowledge Delivery at the Point of Care Empowered by Big Data and NLP. *Biomed Inform Insights* 2016 Jun 23;8(Suppl 1):13-22.
 89. Wallace BC, Kuiper J, Sharma A, Zhu MB, Marshall IJ. Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision. *J Mach Learn Res* 2016;17.
 90. Clark DG, McLaughlin PM, Woo E, Hwang K, Hurtz S, Ramirez L, et al. Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment. *Alzheimers Dement (Amst)* 2016 Feb 15;2:113-22.
 91. Takano K, Ueno M, Moriya J, Mori M, Nishiguchi Y, Raes F. Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. *Behav Res Methods* 2017 Jun;49(3):835-52.
 92. Luo SX, Shinall JA, Peterson BS, Gerber AJ. Semantic mapping reveals distinct patterns in descriptions of social relations in adults with autism spectrum disorder. *Autism Res* 2016 Aug;9(8):846-53.

Correspondence to:

A. Névéol, P. Zweigenbaum
LIMS
Campus universitaire, bât 508
Rue John von Neumann
F-91405 Orsay cedex
France
E-mail: {neveol,pz}@limsi.fr

Appendix: Content Summaries of Selected Best Papers for the 2017 IMIA Yearbook, Section Clinical Natural Language Processing

Althoff, T, Clark K, Leskovec, J

Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health

Trans Assoc Comput Linguist 2016;4:463-47

This paper presents an analysis of counselor/patient interactions in text message conversations collected from a free hotline. The authors characterize counseling conversation relying on a wide range of methods including discourse analysis, statistical language modeling, and sentiment analysis. They evidence actionable conversation strategies that are associated with better conversation outcomes: successful counselors exhibit an ability to adapt creatively to each new counseling situation, their reaction to ambiguity with check questions and appreciation language is well received, they conduct conversations efficiently by spending less time to understand the patient's issue and more time in problem-solving, and they facilitate a change in perspective. These results may be used towards improving practice recommendations and counselor training.

Kilicoglu, H, Demner-Fushman, D

Bio-SCoRes: A Smorgasbord Architecture for Coreference Resolution in Biomedical Text

PLoS One 2016 Mar 2;11(3):e0148538

This paper addresses coreference resolution, which is a fundamental and challenging task in natural language processing. The authors offer a very instructive and informed definition of coreference as well as a comprehensive literature review of the state of the art in coreference resolution. They describe and evaluate a general method for biomedical coreference resolution called

BioSCoRes, implemented in a publicly available toolkit. The evaluation broadly covers the biomedical domain by relying on two existing corpora (one clinical, one biological), as well as a newly developed and shared corpus of drug label inserts. It also offers a detailed performance report on each type of coreference. BioSCoRes obtains overall results that come close (on the clinical corpus) or exceed (on the other two corpora) the state-of-the art.

Morid, MA, Fiszman, M, Raja, K, Jonnalagadda, SR, Del Fiol, G

Classification of clinically useful sentences in clinical evidence resources

J Biomed Inform 2016 Apr;60:14-22

This paper presents a method for classifying sentences from a variety of evidence-based clinical decision support knowledge sources according to clinical usefulness. This work offers a specific definition of actionable, clinically useful sentences. Then, it proceeds to explore advanced NLP methods to extract rich features for sentence classification. A feature ablation study supports the proposed feature-rich approach and shows that the system performs with an F-measure of at least 73% on different text genres. This work is exemplary in exploring fundamental approaches towards the practical goal of providing real-time clinical information at the point of care, while setting up a technical framework that will facilitate the integration of the research results in a clinical setting.

Shivade C, de Marneffe MC, Fosler-Lussier E, Lai AM

Identification, characterization, and grounding of gradable terms in clinical text

Proceedings of the 15th Workshop on Biomedical Natural Language Processing 2016:17-26

This paper presents an analysis of gradable adjectives found in clinical text to qualify medical findings. The authors use existing methods to automatically identify gradable adjectives in clinical corpora and estimate prevalence at about 30% of adjectives.

Focusing on four clinical phenomena, the authors show that the gradable adjectives used to qualify these phenomena in a clinical corpus can be reliably associated to numerical value intervals using a probabilistic model. This very original work relies on a regular expression-based analysis of a clinical description of selected phenomena comprising a gradable adjective along with a numerical value. It offers a first step towards the interpretation of statements using gradable adjectives by grounding their meaning to value intervals, which would facilitate clinical decision-making.

Wu, Y, Denny, JC, Rosenbloom, ST, Miller, RA, Giuse, DA, Wang, L, Blanquicett, C, Soysal, E, Xu, J, Xu, H

A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD)

J Am Med Inform Assoc 2017 Apr 1; 24(e1):e79-e86

This paper presents an open source framework for clinical abbreviation recognition and disambiguation via entity linking to the

Unified Medical Language System (UMLS). This work builds on a large body of research on different aspects of abbreviation resolution including recognition of abbreviated entities in clinical text, the extraction of possible long forms or senses from knowledge bases, and the disambiguation of a given entity that leverages context to identify the unique sense of a given abbreviated entity. The overall framework offers performance that exceeds the state-of-the-art on two shared datasets. In addition, the tool may be tailored to specific needs by allowing the use of customized resources.