

Chapter 11

Networks and Shared Tasks in Clinical Text Mining



Clinical text mining and healthcare text analytics are developed and discussed in several existing researcher networks and practical experiments are carried out in shared tasks. Some of the existing and previous networks are listed below.

- *The Health Natural Language Processing Center (hNLP)*, <http://center.healthnlp.org>.
 - *The UK healthcare text analytics network (Healtex)*, <http://healtex.org>.
 - *The HEalth teXt Analysis network in the Nordic and Baltic countries (HEXAnord)*, <http://dsv.su.se/hexanord>.
 - *The Informatics for Integrating Biology and the Bedside (i2b2)*, <https://www.i2b2.org>.
 - *The Natural Language Processing / Information Extraction (NLP/IE) Program*, <https://github.com/nlpie>.
 - *The Open Health Natural Language Processing Consortium (OHNLP)*, http://www.ohnlp.org/index.php/Main_Page.
 - *The Physionet: Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) Databases*, <https://physionet.org/mimic2/>.
- (All links accessed 2018-01-11.)

In these networks both tools and data sets can be found and downloaded for use in clinical text mining.

The shared tasks use clinical text sets that have been made available for the research community. All the data sets have been de-identified and all participants are obliged to sign confidentiality agreements.

The shared tasks are listed below (Huang and Lu 2015).

- *The Computational Medicine Center's (CMC) 2007 Medical Natural Language Processing Challenge* to automatically assign ICD-9 codes to radiology reports.

- *The CoNLL-2018*.¹ Shared task on the detection of sentences containing hedges and in-sentence resolution of hedge cues.
- *i2b2 series of shared tasks from 2010 to 2014*.
 - Extraction of clinical concepts.
 - Identification of clinical relations.
 - Resolution of co-reference.
 - De-identification of clinical discharge summaries.
 - Identification of heart disease risks.
 - Extraction of medication-related information.
 - Reuse of releases, clinical data sets for answering new clinical questions.
 - Prediction of obesity and its co-morbidity.
 - Classification of sentiment at sentence levels.
 - Prediction of smoking status.
 - Evaluation of clinical NLP software usability.
 - Identification of temporal event/expression.
 - Extraction of temporal relation (using the THYME corpora).
- *The Text REtrieval Conference (TREC) Medical*,² and Clinical Decision Support (CDS) tracks³ during 2011–2014.
 - The Medical track treats extraction of cohorts matching inclusion criteria.
 - The CDS track deals with retrieval of relevant documents for clinical decision making and support it is also called the Precision Medicine track.
- The ShARe/CLEF eHealth⁴ tasks, which consist of:
 - Normalisation of acronyms/abbreviations.
 - Disease template/attribute filling.
 - Recognition and normalisation of diseases.
 - Retrieval of relevant documents.
 - Interactive search systems for eHealth data.
 - Retrieval of relevant documents for query in different languages.
 (See also Fig. 11.1 for more details on the different ShARe/CLEF eHealth tasks.)
- *The SemEval task in 2014* on the recognition and normalisation of diseases (Pradhan et al. 2014).

¹Overview of the CoNLL-2010 Shared Task, <http://rgai.inf.u-szeged.hu/index.php?lang=en&page=conll2010st>. Accessed 2018-01-11.

²TREC Medical track, <http://trec.nist.gov/data/medical.html>. Accessed 2018-01-11.

³TREC Precision Medicine/Clinical Decision Support Track, <http://www.trec-cds.org>. Accessed 2018-01-11.

⁴The ShARe/CLEF eHealth tasks <https://sites.google.com/site/clefehealth>. Accessed 2018-01-11.

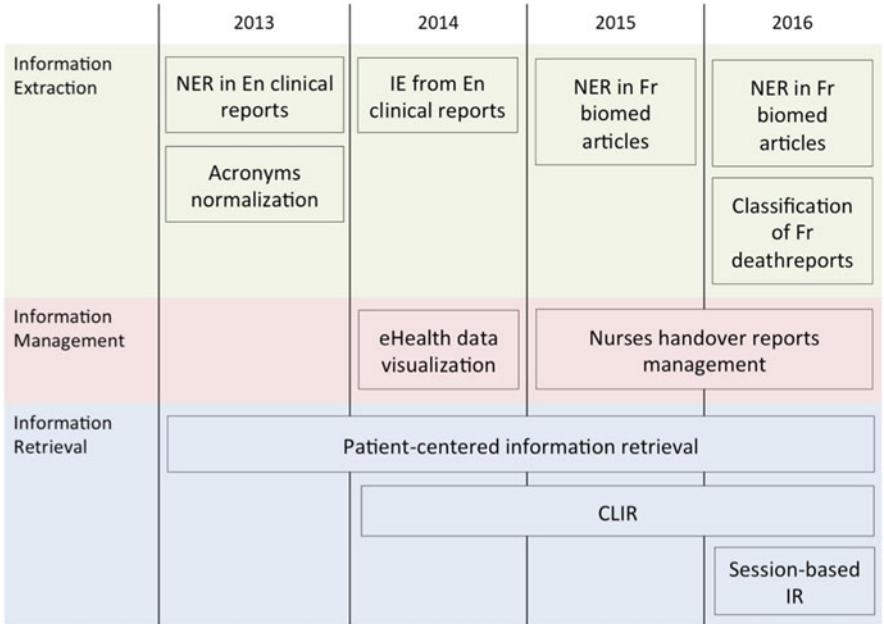


Fig. 11.1 Overview of all ShARe/CLEF eHealth evaluation tasks, from 2013 to 2016,^a (the figure is courtesy of Hanna Suominen, see Kelly et al. 2016). ^aCLEF eHealth evaluation tasks, <https://sites.google.com/site/clefehealth/home/clefehealthtasks.jpg>. Accessed 2018-01-11

- *NLP Challenges for Detecting Medication and Adverse Drug Events from Electronic Health Records.*⁵

Many of these shared tasks have been described in this book previously.

In Kelly et al. (2016) all previous ShARe/CLEF eHealth evaluation tasks are mentioned, and specifically for ShARe/CLEF eHealth 2016 all previous shared tasks are repeated. In Fig. 11.1 there is an overview of all ShARe/CLEF eHealth evaluation tasks from 2013 to 2016.

11.1 Conferences, Workshops and Journals

General conferences in natural language processing occasionally contain presentations on clinical text mining. Conferences in health informatics such as the American Medical Informatics Association (AMIA) and Medinfo World Congress

⁵NLP Challenges for Detecting Medication and Adverse Drug Events from Electronic Health Records, <http://bio-nlp.org/index.php/announcements/39-nlp-challenges>. Accessed 2018-01-11.

on Medical and Health Informatics etc., contain a larger number of articles on clinical text mining, usually in designated tracks at the conferences.

Workshops as International Workshop on Health Text Mining and Information Analysis (Louhi) and the Biomedical Natural Language Processing (BioNLP) at both the Association for Computational Linguistics (ACL) and at the Recent Advances in Natural Language Processing (RANLP) conferences are specially designated workshops in biomedicine and clinical text mining.

For scientific journals there a large number of fora such as the Journal of Biomedical Informatics, Artificial Intelligence in Medicine, Journal of Biomedical Semantics, International Journal of Medical Informatics, Yearbook of Medical Informatics, BMC Medical Informatics and Decision Making, Journal of the American Medical Informatics Association, Health Informatics Journal, Studies in Health Technology and Informatics and many more.

11.2 Summary of Networks and Shared Tasks in Clinical Text Mining

This section has described various networks and shared tasks in the area of clinical text mining and fora for research exchange such as conferences, workshops and scientific journals.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

