Finding Patient visits in EMR using LUXID®.

Luca Toldo, Alexander Scheer

Department of Knowledge Management, Merck KGaA
Frankfurterstraße 250, 64293 Darmstadt, Germany
`firstname.lastname@merckgroup.com`

Abstract. *INTRODUCTION*: Free text sections of the Electronic Medical Records (EMR) contain information that cannot be appropriately constrained in the structured forms. Several studies have shown the potential utility in mining EMR free texts for identifying adverse events (e.g. EU-PSIP, EU-ALERT), and large public-private research projects (e.g. IMI-EHR4CR, CLOUD4HEALTH) aim at mining them further, e.g. for clinical trial optimisation and pharmacovigilance purposes.
AIM: The purpose of this work has been to assess the performance of LUXID®, an off-the-shelve commercial natural language processing system, using the dictionary- and rule-based Medical Entity Relationships Skill Cartridge®and KNIME as automation workflow engine for result combination and formatting, on the University of Pittsburgh BLULab NLP Repository benchmark, in the context of the TREC 2011 Medical Records Retrieval Track (TREC-MED2011).
RESULTS: The system here described achieved the best score for one of the 34 queries (defined as query 111) and overall classified as top 7th-8th (according to the scoring used) in the manual track of TREC-MED2011. More than 80% of the queries of TREC-MED2011 could be appropriately processed automatically. Performance of manually interpreted queries did not differ substantially from those automatically processed. More than 60% of the queries submitted by our system delivered a performance above or on the median of all participants. Very high precision of the system, delivering in certain cases a very low number of hits, correlated statistically with the overall performance.
CONCLUSIONS: Initial results, error analysis are reported and strategies for improvements of the system are outlined; fully supporting the appropriateness in using this technology for identifying patients matching inclusion/exclusion criteria using plain text from unstructured EMR.

Keywords: LUXID®, Medical Entity Relationships Skill Cartridge®, inclusion/exclusion criteria, EMR, EHR

1 Introduction

"An electronic health record (EHR) is a real-time, point-of-care, patient-centric information resource for clinicians that represents a major domain of health information technology. More recently, an EHR has been defined as "a longitudinal electronic record of patient health information, produced by encounters in one or more care settings." It includes patient information such as a problem list, orders, medications, vital signs, past medical history, notes, laboratory results, and radiology reports, among other things. The EHR generates a complete record of a clinical patient encounter or episode of care and underpins care-related activities such as decision making, quality management, and clinical reporting. Some authors distinguish between the terms EHR and electronic medical record (EMR), with EMR focusing on ambulatory care systems. "[16]. Electronic handling of personal medical information has the potential of greatly improving the quality and efficiency of patient care, decreasing medical errors[19], and revolutionising medical research[17]. Although the use of EHR have structural and process benefits[11], direct benefits for the patients might take a while to be achieved[15] and patient portals had little effect on patient empowerment[1]. It has been demonstrated that spontanous adverse events are greatly under-reported and EHR-based, triggered adverse drug events reporting has the potential to solve this problem[13]. Furthermore, electronic screening has been shown to improve efficiency in clinical trial recruitment, and natural language processing is expected to enhance that process further[14].

Hospital medical records are an important repository of medical knowledge, that could be used to improve drug safety monitoring[8] and patient safety through prevention of potential adverse drug events[4].Phenotypic information extracted from free text of EHR has been used to discover disease correlations and stratify patient cohorts[5]. A large amount of information in the EMRs is stored as numerical format, such as measurements of blood parameters, blood pressure, weight, drug administrations, etc[19]. However, "In electronic medical records (EMRs), medication data are often recorded in narrative clinical notes. For example, hospital discharge summaries usually contain some instructions on medications after discharge (eg, "Will start Orapred x 5 days and increase Pulmicort to 0.5 mg inh BID"), and outpatient clinic visits often document medication changes."[18]. In spite of errors in the EMRs, the mining of those databases has been shown to be very valuable[12].

Collaborative research projects such as EU-ALERT[1] and EU-PSIP[2], have fostered the medical research and development by bringing together academia and industries on various techniques to automatically analyse information from electronic health records for research purposes. Further projects such as IMI-EHR4CR[3] and CLOUD4HEALTH [4] are aiming at further progressing the field, the latter particularly with a focus on the free text parts of the electronic health records. Competitive assessments such as TREC-MED[5], and I2B2[6] have pro-

[1] EU-Adverse Drug Reaction, `http://www.alert-project.org/`
[2] EU-Patient Safety in Intelligent Procedures,`http://www.eu-psip.org/`
[3] Electronic Health Records for Clinical Research, `http://www.ehr4cr.eu/`
[4] `http://www.trusted-cloud.de/documents/cloud4health_Gesamtpraesentation.pdf`
[5] Text REtieval Conference Medical Track, `http://trec.nist.gov/`
[6] Informatics for Integrating Biology and the Bedside, `https://www.i2b2.org/`

vided ground for open development, evaluation, and benchmarking of medical information extraction technologies.

Forster et al. [9] applied a commercial search engine[7] to scan discharge summaries for the presence of 104 terms that potentially indicate an adverse effect. They indicated the potential use of the automatic system to replace expensive manual searches performed by the medical professionals. Brown et al. [6] used a general purpose indexing engine on the U.S. Department of Veterans Affairs national scale hospital information system, whose content is mostly unstructured text, to automatically encode clinical concepts in SNOMED CT[8]. Meanwhile, major text mining products (e.g. I2E[9], LUXID®[10], SAS®Text Miner[11]) have been used to extract medical findings from texts.

This paper reports on the application of the Medical Entity Relationships Skill Cartridge®(MER) [10][12], LUXID®text mining engine, and KNIME[3][13] Open Source technology to semi-automatically find patient visits matching the 34 queries of the TREC-MED2011 dataset.

MER uses the MeSH Medical Subject Headings thesaurus[14] and other taxonomies, as well as heuristics to find entities like Cell and tissue terms, clinical trial terms, diagnostic terms, disorder terms and others; and to find relationships such as Cell, Cell disease, Diagnosis, molecular targe, therapy, adverse event and others[15].

Our system includes a dictionary- and rule-based approach with linguistic variant generation to identify medical entities and relationships[2], a lucene [16] based search engine, and a graphical workflow platform (KNIME) to automatically identify the patient visits and to format the results according to TREC-MED2011 track requirements. The results achieved are very positive and support further efforts in the refinement of the approach to identify patients matching inclusion/exclusion criteria using deidentified free text.

2 Methods

Figure 1 shows an illustration of the procedure adopted in this work. Basically, the TREC-MED2011 dataset (hereafter referred as "corpus") has been processed with LUXID®engine, indexing it with MER and making it accessible also as full text index with the integrated open source lucene engine.

The queries have then been processed according to two different protocols: *A*utomatic: the query is processed with the MER, entities are recognised by the

[7] `http://dtsearch.com/`
[8] `http://www.snomed.org/`
[9] `http://www.linguamatics.com/`
[10] `http://www.temis.com/`,`http://www.mondeca.com/Research/Projects/`
[11] `http://www.sas.com/`
[12] `https://clara.uib.no/files/2010/09/Geissler.pdf`
[13] `http://www.knime.org/`
[14] `http://www.nlm.nih.gov/mesh/`
[15] `https://clara.uib.no/files/2010/09/Geissler.pdf`
[16] `http://lucene.apache.org/`

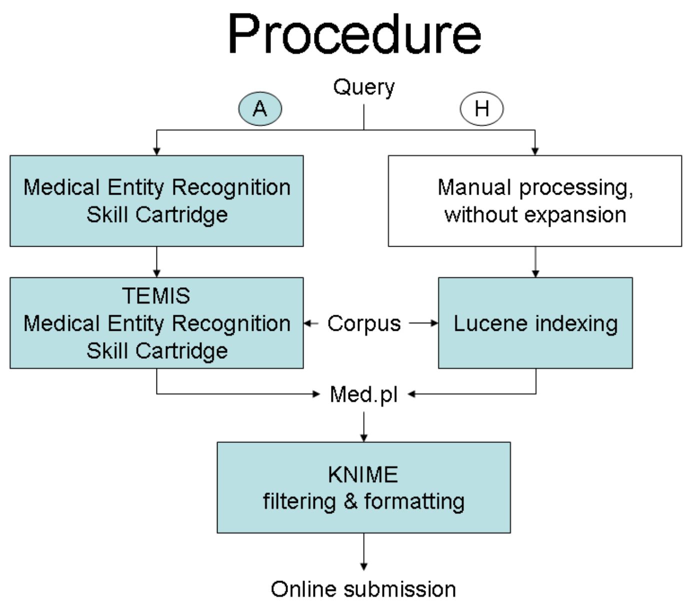

Fig. 1. Overall strategy used in this work.

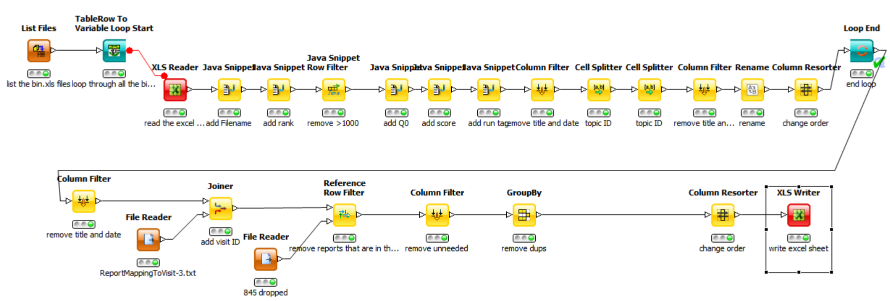

Fig. 2. From raw results to final submission:KNIME workflow

tagger. The corpus is then queried with those entities. *M*anual: In case that the Automatic procedure did not returned any entity, then the query has been processed manually, compiling a short full text pattern. The list of visit identifiers returned by the system have been downloaded and processed by a KNIME workflow to finally generate the outputs required by the TREC-MED organisers, after validation with the Med.pl script provided by TREC-MED organisers. In order to evaluate the performance of the various submissions, the organisers computed several performance indicators, using the trec-eval software[17]. These included: *bref* or binary preference,computes a preference relation of whether judged relevant documents are retrieved ahead of judged irrelevant documents[7];*r-prec* is the precision after R documents [7]. To facilitate interpretation of results, score classes "below, median, above" were computed by us according to table 1, and rows were colour coded: red rows are those for which none of the visit identifiers were judged positive in the corpus; the blue row is the one where LUXID®delivered the top score of all participants.

Table 1. Result scoring procedure.

class	function
below	if Merck-score lower than 90% of the TREC-MED median
above	if TREC median lower than 90% of Merck-score
median	in any other case not covered above

3 Results

Table 2 contains the primary results of this work.

3.1 Resource usage

Loading, parsing, semantic tagging with MER and indexing of the full text of the TREC-MED corpus of over 100,000 medical records required 72 hours, on a dedicated 4 CPU windows server. A limit of 10 minutes has been applied on manual query processing (including search), thus a total of 1 hour was needed for this step. Construction of the KNIME workflow required 30 minutes, including the processing with the Med.pl script provided by the TREC-MED organisers. Query processing and export from LUXID®required approximately 5 minutes per query.

3.2 Query processing, Automatic versus Manual

As reported in Table 2, more than 80% of the queries were processed automatically and delivered consistent results. Only 6 queries could not be automatically

[17] http://trec.nist.gov/trec_eval/index.html

mapped and required manual intervention, however, their performance did not differ much from those generated automatically.

3.3 Relative performance, comparison with other systems

Our approach achieved the best score of the TREC-MED competition for one of the queries (id:111), while 29.4% of the queries achieved a performance statistically significantly better than the median, and 29.4% approximately on the median of the whole competition. The method here described ranked as top 7-8 out of 18 manual runs[20].

3.4 Absolute performance measurements

Inspite of the excellent ranking in the TREC-MED competition, shown in Table 2, the absolute performance has been less than optimal with approximately 40% of the queries. Figure 3 reports for each of the runs and overall the precision and recall curves. These results correlate well with the performance indicators of Table 2 and show that the system is reliable although not maximally efficient (low precision at high recall values). These measurements are key indicators that will guide further refinements of the application, aimed at increasing the precision while retaining high recall.

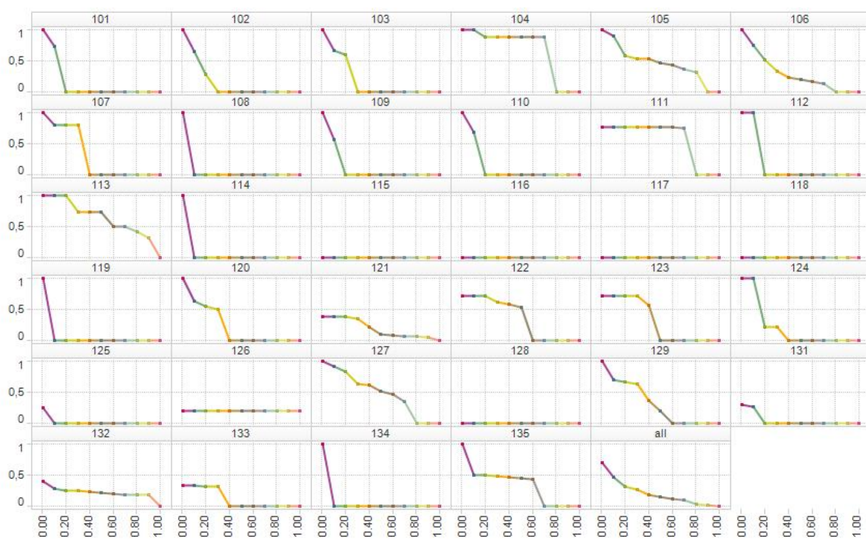

Fig. 3. Precision-recall measurements.

3.5 Error Analysis

Kruskal-Wallis analysis applied on the Table 2 is reported in Table 3 and shows that indeed a moderate correlation exists between the number of terms used in the search strategy and the categorical assignment. At the time of the runs, we were not able to obtain the absolute score for each document, and simply submitted the documents according to their rank. The system provided an intrinsic "cut" below which hits are not returned. This feature of the system that protects the users from low value hits, was also the reason for some of the low scores achieved in less than a third of the runs. Our system did not deliver any result matching the corpora expert annotation for 6 queries (id:115, 116, 117, 118, 126 and 128). An in depth error analysis for these queries is ongoing, however initial results show that the corresponding queries were done appropriately, and should have matched the entities identified by LUXID®or in the free text. Therefore, we expect the issue being in the scoring function of lucene or possibly in loading or indexing errors.

4 Conclusions

The Medical Entity Relationships Skill Cartridge®combined with the LUXID®user interface for search and analytics delivered the best performance in the TREC-MED competition for a specific query, and above median or median performance on over 2 thirds of the queries. These results, the first reported for this industrial text mining software on electronic health records, are very positive and support further efforts in the refinement of the approach to identify patients matching inclusion/exclusion criteria using deidentified free text.

Inspite of these excellent relative ranking, the automatic processing of the queries needs to be improved in order to achieve high rankings for 100% of the queries. Furthermore, the precision of the system needs to be improved, also increasing the number of document ids returned. Detailed error analysis and quantitative assessment with the valuable corpus will be the guiding principles to further improve the precision of the system, however the tagging is shown to be appropriate, with high precision and moderate recall.

5 Acknowledgements

We appreciate the University of Pittsburgh BLULab NLP Repository for having given access to the de-identified medical records, NIST and the anonymous physicians of the Oregon Health and Sciences University that have developed the 34 queries and to those that have done the relevance judging. We also thank Claudia Herrmann for valuable contributions to the manuscript. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

References

1. Ammenwerth E, Schnell-Inderst P, Hoerbst A(2011): Patient empowerment by electronic health records: first results of a systematic review on the benefit of patient portals. Stud Health Technol Inform 165:63-7
2. Aubry C, Gullemin-Lanne S, Lautier C(2002): Une methodologie et un environnement 'aide a la construction de composants de connaissance pour l'Extraction d' Information. CIFT'02: Colloque International sur la Fouille de Textes - Hammamet - Tunisie - 20-23 octobre 2002.
3. Berthold MR, Cebron N, Fill F et al.(2009): KNIME âĂŞ The Konstanz Information Miner. SIGKDD Explorations 11:26-31
4. Beuscart R,Chazard E(2011): PSIP Overview: Detection and prevention of adverse drug events. In: International Workshop Patient Safety through Intelligent Procedures in medication, Sofia, Bulgaria, 23 June 2011, pp. 1-3
5. Roque FS, Jensen PB, Schmock H et al.(2011): Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS Comput Biol 7:e1002141
6. Brown SH, Elkin PL, Rosenbloom ST et al.(2008): eQuality for All: Extending Automated Quality Measurement of Free Text Clinical Narratives. In: AMIA 2008 Symposium Proceedings, pp.71-75
7. Buckley C, Voorhees E(2005): Retrieval System Evaluation. In: Voorhees E and Harman D(eds) TREC: Experiment and Evaluation in Information Retrieval. MIT Press 2005.
8. Coloma PM, Schuemie MJ, Trifiro G et al.(2011): Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiology and drug safety 20:1âĂŞ11
9. Forster AJ, Andrade J, van Walraven C.(2004): Validation of a discharge summary term search method to detect adverse events. J Am Med Inform Assoc 12:200-206
10. Geissler S(2010): Information-extraction in a multi-purpose text mining platform. In: Terminology and resource harmonization. Common Language Resources and their Applications – a Marie Curie Initial Training Network. 13-17 September 2010, Bergen, Norway.https://clara.uib.no/training-program/terminology-course/
11. Holroyd-Leduc JM, Lorenzetti D, Straus SE et al.(2011): The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence. J Am Med Inform Assoc 18:732e737
12. Hripcsak G, Knirsch C, Zhou L et al.(2011): Bias Associated with Mining Electronic Health Records. Journal of Biomedical Discovery and Collaboration 6:48-52
13. Linder JA, Haas SJ, Iyer A et al.(2010): Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. Pharmacoepidemiology and drug safety 19:1211-1215
14. Thadani SR, Weng C, Bigger TJ et al.(2009): Electronic screening improves efficiency in clinical trial recruitment. J Am Med Inform Assoc 16:869-873
15. Sheikh A, Cornford T, Barber N et al.(2011): Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. BMJ 343:d6054. doi: 10.1136/bmj.d6054
16. Staggers N, Weir C, Phansalkar S(2008): Patient Safety and Health Information Technology: Role of the Electronic Health Record. Agency for Healthcare Research and Quality.

17. Sundgren M, Schmidt A(2009): The Impact of the Re-use of EHR Data-A Pharmaceutical Industry Perspective. EuroRec 2009 Annual Conference
18. Xu H, Stenner SP, Doan D et al.(2010): MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 17:19e24
19. Zlabek JA, Wickus JW, Mathiason MA(2011): Early cost and safety benefits of an inpatient electronic health record. J Am Med Inform Assoc 18:169-72
20. Voorhees EM, Tong RM (2011) Overview of the TREC 2011 Medical Records Track. In the Twentieth text retrieval conference (TREC2011) Notebook, 2011.

Table 2. Performance of Automatic and Manual query processing. *qry* identifier of the TREC-MED query; *n terms* number of terms in the (manually or automatically) generated query; *n hits* number of documents returned by LUXID®; *bref* or binary preference;*r-prec* or precision after R documents. See Materials and Methods.

qry	n terms	n hits	bref	score	r-prec	score
101	1	25	0,1574	below	0,1622	below
102	2	167	0,253	below	0,2697	median
103	0	20	0,2292	above	0,25	above
104	2	12	0,7037	median	0,7778	median
105	1	2685	0,83	median	0,4621	median
106	4	3252	0,5416	above	0,3059	above
107	1	13	0,327	median	0,3478	above
108	2	27	0,0769	above	0,0769	median
109	2	38	0,1199	below	0,122	below
110	1	30	0,1455	below	0,1474	below
111	3	50	0,6304	above	0,7143	above
112	1	24	0,1918	below	0,1918	below
113	2	87	0,6071	above	0,5714	above
114	3	111	0,0727	below	0,0727	below
115	0	167	0	below	0	below
116	2	11	0	below	0	below
117	1	75	0	below	0	below
118	1	4	0	below	0	below
119	2	13	0,0855	below	0,087	below
120	0	204	0,3044	below	0,3248	median
121	3	1358	0,2594	above	0,3	above
122	0	75	0,4444	median	0,5	median
123	0	28	0,3646	below	0,4242	median
124	2	44	0,1667	above	0,1667	above
125	2	21	0,0561	median	0,0714	median
126	1	249	0	below	0	below
127	1	340	0,6544	below	0,5059	median
128	2	2	0	below	0	below
129	2	167	0,4507	above	0,3962	above
131	0	131	0,0967	below	0,1111	below
132	2	1691	0,77	below	0,2234	below
133	2	41	0,205	above	0,3	above
134	1	7	0,0562	below	0,0588	below
135	2	196	0,5613	above	0,4576	above

Table 3. Kruskal-Wallis analysis of the Table2.

numerical	categorigal	P-value
r-prec Merck	r-precscore	7.17E-005
bpref Merck	r-precscore	6.95E-004
r-prec Merck	medianscore	4.66E-003
r-prec Merck	medianscore	1.54E-002
n terms	medianscore	2.67E-002
n terms	r-precscore	7.96E-002