

Decision Support, Knowledge Representation and Management: Structuring Knowledge for Better Access

Findings from the Yearbook 2008 Section on Decision Support, Knowledge Representation and Management

A.-M. Rassinoux, Managing Editor for the IMIA Yearbook Section on Decision Support, Knowledge Representation and Management

University Hospitals of Geneva, Geneva, Switzerland

Summary

Objectives: To summarize current outstanding research in the field of decision support, knowledge representation and management.

Method: Synopsis of the articles selected for the IMIA Yearbook 2008.

Results: Five papers from international peer reviewed journals have been selected for the section on decision support, knowledge representation and management. They address a wide range of topics such as the recognition and extraction of negation or time from clinical narratives, the use of ontological elements to reduce the complexity of natural language processing applications or to strengthen the precision of document retrieval as well as the benefits of integrating clinical decision support within computer provider order entry.

Conclusions: The best paper selection brings to light that whatever the methodological approach used in decision support, knowledge representation and management, all applications benefit from manipulating information that is expressed in both a meaningful and structured way. In order to combine the flexibility and expressive power of natural language with the computational tractability of structured data, the electronic health record based on structured narrative offers new perspectives.

Keywords

Natural language processing, knowledge representation, ontologies, information retrieval, reasoning

Geissbuhler A, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2008. *Methods Inf Med* 2008; 47 Suppl 1:80-2

Introduction

Healthcare is a knowledge intensive domain and, over the years, we observe an increase of the amount and diversity of knowledge, along with new potential consumers of such health-related knowledge. The topic of the IMIA Yearbook 2008, "Access to Health Information", reflecting on the usage and sharing of such clinical information, implies to take into account the level of granularity chosen for storing information [1, 2]. Whereas the expressive power of free-text narrative is more suitable for human presentation, structuring information, in a way that allows a rapid and efficient way of accessing and retrieving it, is a pre-requisite for further automated processing. The Electronic Health Record (EHR) reflects this duality as clinical information is organized at various levels, going from medical records written in a free text form to detailed statements stored into databases [3, 4, 5].

Best Paper Selection

The best paper selection process has elected five papers (see Table 1) for the section decision support, knowledge representation and management. A brief content summary of each selected ar-

ticle is given in the appendix of this synopsis. All these papers, either related to natural language processing (NLP) [6, 7], ontologies [8], information retrieval [7, 9] or reasoning tasks [10], somehow refer to the importance of structuring information as highlighted in the above introduction.

As negation and temporal information are intrinsically related to many medical reasoning tasks, their recognition and extraction from clinical narratives constitute significant research fields. Zhou and Hripcsak [11] focus on detecting temporal information in medical texts and emphasize the level of complexity engendered by medical narrative data compared to structured data. The authors see a promising future direction, for higher level of temporal reasoning tasks, by combining these two kinds of data. In [12], Huang and Lowe take advantage of the whole syntactical structural information of a sentence, described through parse trees generated from full NLP parsing, to automatically locate negated concepts in clinical radiology reports. Such a syntactical structure has proved to be efficient for delimiting the scope of negation. Moreover, relying on existing structured knowledge resources such as ontologies to reduce the complexity of text processing tasks is one of the purposes of

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2008 in the section 'Decision Support, Knowledge Representation and Management'. The articles are listed in alphabetical order of the first author's surname.

Section
Decision Support, Knowledge Representation and Management
<ul style="list-style-type: none"> ▪ Fan JW, Friedman C. Semantic classification of biomedical concepts using distributional similarity. <i>J Am Med Inform Assoc</i> 2007;14(4):467-77. ▪ Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. <i>J Am Med Inform Assoc</i> 2007;14(3):304-11. ▪ Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. <i>J Am Med Inform Assoc</i> 2007;14(1):29-40. ▪ Moskovitch R, Martins SB, Behiri E, Weiss A, Shahar Y. A comparative evaluation of full-text, concept-based, and context-sensitive search. <i>J Am Med Inform Assoc</i> 2007;14(2):164-74. ▪ Zhou L, Hripcsak G. Temporal reasoning with medical data—a review with emphasis on medical natural language processing. <i>J Biomed Inform</i> 2007;40(2):183-202.

Fan and Friedman [13]. After identifying ontological concepts that most frequently occur in natural language, a distributional similarity approach is used to reclassify and validate ontology for NLP applications. Furthermore, Moskovitch et al. [14] argue that using structured, labelled documents as well as concept-based search can greatly improve the performance of traditional full-text search. Finally, the paper of Kuperman et al. [15] focuses on the medication-related benefits of integrating clinical decision support (CDS) within computer provider order entry (CPOE). Reasoning about patients' clinical data and conditions also implies to structure such data into the EHR in an accurate and comprehensive way.

Conclusions and Outlook

The best paper selection for the IMIA Yearbook 2008 section on decision support, knowledge representation and management, has gathered articles related to a broad spectrum of methodological approaches [16], like NLP [11, 12, 13], ontologies [13, 14], information retrieval [14] or reasoning [11, 15]. All these papers are also concerned, to some extent, with medical information

representation, whether it relates to time [11], negation [12], biomedical terms [13], queries [14] or patient's clinical data [15]. Finally, as already mentioned in the previous IMIA Yearbook 2007 [17], NLP techniques play a growing key role for extracting structured data from narrative texts [18, 19]. Taking advantage of both the flexibility and expressive power of natural language with the computational tractability of structured data should lead to more powerful EHR based on structured narrative [20]. Thus, fusing unstructured text together with coded data promises to a fruitful area of research for the next coming years.

Acknowledgement

I greatly acknowledge the support of Martina Hutter and of the reviewers in the selection process of the IMIA Yearbook.

References

1. Kumar A, Smith B. Biomedical informatics and granularity. *Comp Funct Genomics* 2004;5:501-8.
2. Tange HJ, Schouten HC, Kester ADM, Hasman A. The granularity of medical narratives and its effect on the speed and completeness of information retrieval. *J Am Med Inform Assoc* 1998;5:571-82.
3. Lovis C, Baud RH, Planche P. Power of expression in the electronic patient record: structured data or narrative text? *Int J Med Inform* 2000;58-59:101-10.
4. Bernstein K, Bruun-Rasmussen M, Vingtoft S. A method for specification of structured clinical content in electronic health records. *Stud Health Technol Inform*. 2006;124:515-21.
5. Los RK, van Ginneken AM, de Wilde M, van der

- Lei J. OpenSDE: Row modeling applied to generic structured data entry. *J Am Med Inform Assoc* 2004; 11(2):162-5.
6. Friedman C, Hripcsak G. Natural language processing and its future in medicine. *Acad Med* 1999;74(8):890-5.
7. Chen L, Friedman C. Extracting phenotypic information from the literature via natural language processing. *Medinfo* 2004;11(Pt2):758-62.
8. Stenzhorn H, Beisswanger E, Schulz S. Towards a top-domain ontology for linking biomedical ontologies. *Medinfo* 2007;12(Pt2):1225-9.
9. Hung PW, Johnson SB, Kaufman DR, Mendonça EA. A multi-level model of information seeking in the clinical domain. *J Biomed Inform* 2008;41:357-70.
10. Augusto JC. Temporal reasoning for decision support in medicine. *Artif Intell Med* 2005;33(1):1-24.
11. Zhou L, Hripcsak G. Temporal reasoning with medical data - a review with emphasis on medical natural language processing. *J Biomed Inform* 2007;40(2):183-202.
12. Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. *J Am Med Inform Assoc* 2007;14(3):304-11.
13. Fan JW, Friedman C. Semantic classification of biomedical concepts using distributional similarity. *J Am Med Inform Assoc* 2007;14(4):467-77.
14. Moskovitch R, Martins SB, Behiri E, Weiss A, Shahar Y. A comparative evaluation of full-text, concept-based, and context-sensitive search. *J Am Med Inform Assoc* 2007;14(2):164-74.
15. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc* 2007;14(1):29-40.
16. Brill B. Decision Support, Knowledge Representation and Management: A broad methodological spectrum. In Haux R, Kulikowski C, editors. *IMIA Yearbook of Medical Informatics 2006*. *Methods Inf Med* 2006; 45 Suppl 1:81-3.
17. Meystre S. Electronic patient records: Some answers to the data representation and reuse challenges. In Geissbuhler A, Haux R, Kulikowski C, editors. *IMIA Yearbook of Medical Informatics 2007*; 47-8.
18. Fujii H, Yamagishi H, Ando Y, Tsukamoto N, Kawaguchi O, Kasamatsu T, et al. Structuring of free-text diagnostic report. *Medinfo*, 2007; 12(Pt1):669-73.
19. Chen ES, Hripcsak G, Friedman C. Disseminating natural language processed clinical narratives. *AMIA Annu Symp Proc*, 2006. p. 126-30.
20. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, et al. An electronic health record based on structured narrative. *J Am Med Inform Assoc*, 2008;15(1): 54-64.

Correspondence to:

Anne-Marie Rassinoux, Ph. D.
University Hospitals of Geneva
Service of Medical Informatics
Unit of Clinical Informatics
24, rue Micheli-du-Crest
CH-1211 Geneva 14, Switzerland
Tel: +41 22 372 6293
Fax: +41 22 372 8680
E-mail: anne-marie.rassinoux@sim.hcuge.ch

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2008, Section Decision Support, Knowledge Representation and Management*

Fan JW, Friedman C

Semantic classification of biomedical concepts using distributional similarity

J Am Med Inform Assoc 2007; 14(4):467-77

As more and more textual biomedical knowledge becomes available electronically, natural language processing (NLP) applications have proved to be useful for extracting and managing biomedical entities and relations from texts [6]. In order to reduce the complexity of recognizing large number of biomedical terms, existing large biomedical knowledge resources such as the Unified Medical Language System (UMLS) can serve as a repository. The approach proposed by Fan and Friedman [13] focuses on re-classifying and validating UMLS concepts for NLP applications by applying a reproducible method based on distributional similarity. For this, syntactic dependencies from shallow parsing are used to automatically classify UMLS concepts into seven broader semantic classes selected as more suitable for the needs of NLP applications. The results demonstrate the feasibility of automatic classification with an estimated lowest error rate of 0.198 and also show that increasing the number of syntactic dependencies strengthens the classification accuracy.

Huang Y, Lowe HJ

A novel hybrid approach to automated negation detection in clinical radiology reports

J Am Med Inform Assoc 2007; 14(3):304-11

Negation is frequently encountered in clinical documents and ignoring it in automated indexing systems can be an important source of poor precision, especially when considering further clinical document retrieval. In [12], Huang and Lowe propose a novel hybrid approach that combines regular expression matching with grammatical parsing to automatically detect negation in clinical radiology reports. Both a classification and a grammar for negation, with good coverage, were developed and using the whole syntactical structure of sentences within parse trees leads to a good precision in identifying negated phrases.

Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW
Medication-related clinical decision support in computerized provider order entry systems: a review

J Am Med Inform Assoc 2007; 14(1):29-40

Kuperman et al. review clinical decision support (CDS) for electronic prescribing within computer provider order entry (CPOE) systems [15]. They argue that CDS provides added value when a CPOE system is in place as it allows improving the quality and safety of medication prescribing as well as reducing its cost. The authors distinguish two stages of medication-based decision support which are basic and advanced, the second one being implemented only after the first one is in place. For instance, whereas drug-drug interaction checking is definitively part of basic CDS, issues related to drug-disease contraindication or drug-pregnancy checking are included in advanced CDS as they require more information checking and controls that are not always trivial to put in place. Recommendations for delivering effective medication-related CDS within CPOE systems are finally addressed to every individual and organization that is involved in applying information technology to improve pharmacotherapy quality and safety within patient care.

Moskovitch R, Martins SB, Behiri E, Weiss A, Shahar Y

A comparative evaluation of full-text, con-

cept-based, and context-sensitive search

J Am Med Inform Assoc 2007; 14(2):164-74

Moskovitch et al. present a study [14] in which they compare three approaches to document retrieval: the traditional full-text search using a set of keywords, the concept-based search using documents that are pre-indexed by a conceptual hierarchy and the context-sensitive search that uses structured, labelled documents. A clinical guidelines corpus was used to evaluate these methods. Although none of these approaches is really new, this comprehensive comparison highlights important features for effective document retrievals. In particular, the authors conclude that the more ontological elements used, the greater the resulting precision. This is why adding concept-based and context-sensitive searches over the use of free-text queries is suggested.

Zhou L, Hripcsak G

Temporal reasoning with medical data - a review with emphasis on medical natural language processing

J Biomed Inform 2007, 40(2):183-202

There is a long history of temporal representation and reasoning with medical data as depicted by the extensive overview of previous publications given by Zhou and Hripcsak in [11]. Whereas predominant research efforts on temporal reasoning in biomedical informatics focus on processing structured data, Zhou and Hripcsak aim attention at systems that deal with time from clinical narrative data using Medical Language Processing (MLP) techniques. Therefore, new challenges in processing temporal information in medical texts are highlighted such as the variability of temporal expressions, ambiguity, complex relations, implicit and pragmatic information, temporal fuzziness and uncertainty, as well as inconsistency, to name only a few. The authors conclude that integrating structured data information to the narrative information seems a promising future direction for improving the temporal reasoning task as well as retrieval or visualization of temporal information.

* The complete papers can be accessed in the Yearbook's full electronic version, provided that permission has been granted by the copyright holder(s)