

Knowledge Representation and Management. From Ontology to Annotation

Findings from the Yearbook 2015 Section on Knowledge Representation and Management

J. Charlet^{1,2}, S.J. Darmoni^{2,3}, Section Editors for the IMIA Yearbook Section on Knowledge Representation and Management

¹ AP-HP, Dept. of Clinical Research and Development, Paris, France

² INSERM, U1142, LIMICS, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, Paris, France; Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), Villetaneuse, France

³ Department of Biomedical Informatics, Rouen University Hospital, Normandy & TIBS, LITIS EA 4108, Institute for Research and Innovation in Biomedicine, Rouen, France

Summary

Objective: To summarize the best papers in the field of Knowledge Representation and Management (KRM).

Methods: A comprehensive review of medical informatics literature was performed to select some of the most interesting papers of KRM published in 2014.

Results: Four articles were selected, two focused on annotation and information retrieval using an ontology. The two others focused mainly on ontologies, one dealing with the usage of a temporal ontology in order to analyze the content of narrative document, one describing a methodology for building multi-lingual ontologies.

Conclusion: Semantic models began to show their efficiency, coupled with annotation tools.

Keywords

Ontology, knowledge representation, annotation

Yearb Med Inform 2015;10:134-6

<http://dx.doi.org/10.15265/IY-2015-038>

Published online August 13, 2015

Introduction

The year 2014 has produced a large amount of publications related to Knowledge Representation and Management, in particular several articles on ontology-based annotation.

KRM focused on developing techniques to be used and leveraged in other medical informatics domain. This year again, the selected articles for the KRM section were covering several sub-domains: e.g. ontology-based annotation (the main one), terminology and ontology mapping, data integration, ontology for clinical decision support system (CDSS).

The aim of this section was to select and present some of the best papers published in 2014 in the KRM domain, based either on their impact or their novelty approach in the knowledge representation and management field.

About the Paper Selection

The selection of papers is the result of a comprehensive literature search: section editors have pre-selected 15 papers [1-15] after a complex query from PubMed retrieving more than 1,000 articles, then 100 articles after a first selection based on title and abstract of these articles. Five reviewers

reviewed the pre-selected papers to select the best four final papers (see Table 1) [1-4].

A brief content summary of these four selected papers can be found in the appendix of this synopsis. Among the 11 other selected papers,

- Palombi et al. [12] have developed an ontology-based tool to query and perform reasoning on complex anatomical models. This ontology (My Corporis Fabrica) is an extension of the Foundational Model of Anatomy (FMA), which is considered as the reference ontology in this domain: anatomy. In the same domain, Nichols et al. [5] has proposed an enhancement of the FMA for the neuroanatomical domain. These two papers highlight the necessity to create a solid base (in this case, the creation of the FMA) in order to allow future improvements in specific sub-domains.
- Aguita et al. [8] has proposed a view-oriented approach to align RDF-based repositories. In the literature, most of the mappings are based on one to one correspondences. This new approach takes into account RDF subgraph to propose more complex mappings. The same topic of inter-terminology mappings has been studied by Kim et al. [14]. They were focusing on nursing issues, using International Classification for Nursing Practice (ICNP) and the Systematized

Nomenclature of Medicine–Clinical Terms (SNOMED–CT).

- Soldateva et al. [13] have developed an ontology to capture as much as possible the semantics of biomedical protocols (e.g. good laboratory practice or good manufacturing practice).
- Ning Xue et al. [15] have proposed a robust and automated model-based semantic registration for the multimodal alignment of the knee bone and cartilage from three-dimensional MR image data. The semantics similarity was based on Dice distance. Harispe et al. [11] have reviewed all domain-specific semantic similarity measures that have been recently defined. These authors proposed a unifying framework to improve the understanding of these semantic measures. Gøeg et al. [9] has evaluated Lin similarity estimates and Sokal and Sneath similarity with two aggregation techniques to cluster clinical models from electronic health records based on SNOMED-CT.
- Doulaverakis et al. [7] have developed a semantic-enabled drug recommendations discovery framework to optimize personalized drug prescription via the discovery of new drug-drug interactions or drug-disease interactions.
- López-García et al. [6] have evaluated cross-domain targeted ontology subsets for annotation using a subset of drugs from RxNorm using the UMLS Metathesaurus, the NDF-RT cross-ontology, and the CORE problem list subset of SNOMED CT. The wide range in recall (21-69%) is strongly suggesting that more research is needed in this field. In the same domain of annotation, Chakrabarti et al. [10] has proposed to use statistical algorithms for ontology-based annotation of scientific literature, in particular a probabilistic framework

Conclusions and Outlook

In 2014, ontology and terminology based annotation appears as a major tool in medical informatics... and a step further: in several countries, including France, several tools became commercial products to index mil-

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2015 in the section 'Knowledge Representation and Management'. The articles are listed in alphabetical order of the first author's surname.

Section
Knowledge Representation and Management
<ul style="list-style-type: none"> ▪ Choi S, Choi J, Yoo S, Kim H, Lee Y. Semantic concept-enriched dependence model for medical information retrieval. <i>J Biomed Inform</i> 2014 Feb;47:18-27. ▪ Clark K, Sharma D, Qin R, Chute CG, Tao C. A use case study on late stent thrombosis for ontology-based temporal reasoning and analysis. <i>J Biomed Semantics</i> 2014 Dec 11;5(1):49. ▪ Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, Mougín F. Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease. <i>J Biomed Inform</i> 2014 Apr;48:171-82. ▪ Funk C, Baumgartner W Jr, Garcia B, Roeder C, Bada M, Cohen KB, Hunter LE, Verspoor K. Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. <i>BMC Bioinformatics</i> 2014;15:59.

lions of reports to improve and audit medical coding and to enhance automatic detection of potential patients in clinical trials.

Acknowledgements

We would like to thank Martina Hutter for her support and the reviewers for their participation in the selection process of the IMIA Yearbook.

References

1. Choi S, Choi J, Yoo S, Kim H, Lee Y. Semantic concept-enriched dependence model for medical information retrieval. *J Biomed Inform* 2014 Feb;47:18-27.
2. Funk C, Baumgartner W Jr, Garcia B, Roeder C, Bada M, Cohen KB, et al. Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinformatics* 2014;15:59.
3. Clark K, Sharma D, Qin R, Chute CG, Tao C. A use case study on late stent thrombosis for ontology-based temporal reasoning and analysis. *J Biomed Semantics* 2014 Dec 11;5(1):49.
4. Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, et al. Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease. *J Biomed Inform* 2014 Apr;48:171-82.
5. Nichols BN, Mejino JL, Detwiler LT, Nilsen TT, Martone ME, Turner JA, et al. Neuroanatomical domain of the foundational model of anatomy ontology. *J Biomed Semantics* 2014 Jan 8;5(1):1.
6. López-García P, Lependu P, Musen M, Illarramendi A. Cross-domain targeted ontology subsets for annotation: the case of SNOMED CORE and RxNorm. *J Biomed Inform* 2014 Feb;47:105-11.
7. Doulaverakis C, Nikolaidis G, Kleontas A, Kompatsiaris I. Panacea, a semantic-enabled drug recommendations discovery framework. *J Biomed Semantics* 2014 Mar 6;5(1):13.

8. Anguita A, García-Remesal M, de la Iglesia D, Graf N, Maojo V. Toward a view-oriented approach for aligning RDF-based biomedical repositories. *Methods Inf Med* 2015;54(1):50-5.
9. Gøeg KR, Cornet R, Andersen SK. Clustering clinical models from local electronic health records based on semantic similarity. *J. Biomed Inform* 2015 Apr;54:294-304.
10. Chakrabarti C, Jones TB, Luger GF, Xu JF, Turner MD, Laird AR, et al. Statistical algorithms for ontology-based annotation of scientific literature. *J Biomed Semantics* 2014 Jun 3;5(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G):S2.
11. Harispe S, Sánchez D, Ranwez S, Janaqi S, Montmain J. A framework for unifying ontology-based semantic similarity measures: a study in the biomedical domain. *J Biomed Inform* 2014 Apr;48:38-53.
12. Palombi O, Ulliana F, Favier V, Léon JC, Rousset MC. My Corporis Fabrica: an ontology-based tool for reasoning and querying on complex anatomical models. *J Biomed Semantics* 2014 May 6;5:20.
13. Soldatova LN, Nadis D, King RD, Basu PS, Haddi E, Baumlé V, et al. EXACT2: the semantics of biomedical protocols. *BMC Bioinformatics*. 2014;15 Suppl 14:S5.
14. Kim TY, Hardiker N, Coenen A. Inter-terminology mapping of nursing problems. *J Biomed Inform* 2014 Jun;49:213-20.
15. Xue N, Doellinger M, Fripp J, Ho CP, Surowiec RK, Schwarz R. Automatic model-based semantic registration of multimodal MRI knee data. *J Magn Reson Imaging* 2015 Mar;41(3):633-44.
16. Baneyx A, Charlet J, Jaulent MC. Building an ontology of pulmonary diseases with natural language processing tools using textual corpora. *Int J Med Inform* 2007 Feb-Mar;76(2-3):208-15.

Correspondence to:

Dr Jean Charlet
LIMICS - INSERM U1142
Campus des Cordeliers
15, rue de l'école de médecine
75006 Paris, France
Tél. +33 1 44 27 91 09
E-mail: Jean.Charlet@upmc.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2014, Section Knowledge Representation and Management

Choi S, Choi J, Yoo S, Kim H, Lee Y

Semantic concept-enriched dependence model for medical information retrieval

J Biomed Inform 2014 Feb;47:18-27

This paper is about semantic information retrieval, where semantics is not anymore based on semantic expansion in the query but on semantic concept-based term-dependence to improve the ranking of the retrieved resources. Based on a clinical document corpus (TREC Medical records track) and a medical literature corpus (OHSUMED), a leave-one-out cross validation was performed. The semantic concept-enriched dependence model (SCDM) proposed by this study consistently outperformed other state-of-the-art retrieval methods.

Funk C, Baumgartner W Jr, Garcia B, Roeder C, Bada M, Cohen KB, Hunter LE, Verspoor K

Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters

BMC Bioinformatics 2014 Feb 26;15:59

This study is a benchmark on terminology and ontology-based annotators. These three annotators (MetaMap, NCBO Annotator,

and ConceptMapper) were evaluated on eight biomedical ontologies in the Colorado Richly Annotated Full-Text (CRAFT) Corpus. ConceptMapper provides the highest F-measure of seven out of eight ontologies.

Actually, we can note that the generic ConceptMapper tool generally provides the best performance on the concept normalization task, despite not being specifically designed for use in the biomedical domain. Flexibility it provides in controlling precisely how terms are matched in text makes it possible to adapt it to the varying characteristics of different ontologies.

Clark K, Sharma D, Qin R, Chute CG, Tao C

A use case study on late stent thrombosis for ontology-based temporal reasoning and analysis

J Biomed Semantics 2014 Dec 11;5(1):49

In this paper, the authors show how they have applied the Clinical Narrative Temporal Relation Ontology (CNTRO) and its associated temporal reasoning system (the CNTRO Timeline Library) to trend temporal information within medical device adverse event report narratives.

238 narratives documenting occurrences of late stent thrombosis adverse events from the Food and Drug Administration's (FDA) Manufacturing and User Facility Device Experience (MAUDE) database were annotated and evaluated using the CNTRO Timeline Library to identify, order, and calculate the duration of temporal events.

The CNTRO Timeline Library had a 95% accuracy in correctly ordering events within the 238 narratives. Other results are proposed in the paper. It is important to notice that, in

the project, the annotation was manual. A perspective for the authors is to develop an automatic annotation process and it seems evident that the approach/tools these same tools could be applied to other medical device adverse event narratives in order to identify currently unknown temporal trends.

Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, Mouglin F

Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease

J Biomed Inform 2014 Apr;48:171-82

Ontologies are useful tools for sharing and exchanging knowledge. In this paper, the authors present a method for building a bilingual domain ontology from textual and termino-ontological resources intended for semantic annotation and information retrieval of textual documents. This method combines two approaches: ontology learning from texts and the reuse of existing terminological resources. It consists of four steps: (i) term extraction from domain specific corpora (in French and English) using textual analysis tools, (ii) clustering of terms into concepts organized according to the UMLS Metathesaurus, (iii) ontology enrichment through the alignment of French and English terms using parallel corpora and the integration of new concepts, (iv) refinement and validation of results by domain experts. The treatment of corpora follows the same methodology with the same tools than [16]. The work with parallel corpora is sophisticated and allows to build a real bilingual domain ontology.