

Efficient Results in Semantic Interoperability for Health Care

Findings from the Section on Knowledge Representation and Management

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

¹ Normandie Universités, Univ. Rouen, NormaSTIC FR CNRS 3638, IRIB and LITIS EA 4108, Information Processing in Biology & Health, Saint Étienne du Rouvray, France

² INSERM, UMR_S 1142, LIMICS, Paris, France; Sorbonne Universités, Univ. Paris 06, Paris, France

³ AP-HP, Dept. of Clinical Research and Development, Paris, France

Summary

Objectives: To summarize excellent current research in the field of Knowledge Representation and Management (KRM) within the health and medical care domain.

Method: We provide a synopsis of the 2016 IMIA selected articles as well as a related synthetic overview of the current and future field activities. A first step of the selection was performed through MEDLINE querying with a list of MeSH descriptors completed by a list of terms adapted to the KRM section. The second step of the selection was completed by the two section editors who separately evaluated the set of 1,432 articles. The third step of the selection consisted of a collective work that merged the evaluation results to retain 15 articles for peer-review.

Results: The selection and evaluation process of this Yearbook's section on Knowledge Representation and Management has yielded four excellent and interesting articles regarding semantic interoperability for health care by gathering heterogeneous sources (knowledge and data) and auditing ontologies. In the first article, the authors present a solution based on standards and Semantic Web technologies to access distributed and heterogeneous datasets in the domain of breast cancer clinical trials. The second article describes a knowledge-based recommendation system that relies on ontologies and Semantic

Web rules in the context of chronic diseases dietary. The third article is related to concept-recognition and text-mining to derive common human diseases model and a phenotypic network of common diseases. In the fourth article, the authors highlight the need for auditing the SNOMED CT. They propose to use a crowd-based method for ontology engineering.

Conclusions: The current research activities further illustrate the continuous convergence of Knowledge Representation and Medical Informatics, with a focus this year on dedicated tools and methods to advance clinical care by proposing solutions to cope with the problem of semantic interoperability. Indeed, there is a need for powerful tools able to manage and interpret complex, large-scale and distributed datasets and knowledge bases, but also a need for user-friendly tools developed for the clinicians in their daily practice.

Keywords

Biomedical ontologies; information storage and retrieval; knowledge representation; vocabulary controlled

Yearb Med Inform 2016;184-7

<http://dx.doi.org/10.15265/IY-2016-051>

Published online November 10, 2016

heterogeneous datasets and knowledge bases constitute important resources from which new knowledge may be acquired for health care. Several types and projects for semantic interoperability exist [3], mainly related to data models and to knowledge representation. For example, the Smart Open Services for European Patients and SemanticHealthNet projects [4] define a framework for semantic interoperability in EHRs and the Electronic Health Record for Clinical Research (EHR4CR) project develops solutions for clinical research [5]. It is also well-established that ontologies play a central role for knowledge sharing and management. Recent works developed ontologies in bio-medical sub-domains. In the context of neurosciences, Batrancourt et al. [6] proposed a multi-layer ontology for instruments that assess brain and cognitive functions and behavior. It extends a developed ontology that allows sharing brain images and region of interest in neuroimaging. In [7] Wang et al. developed an application designed for users for ontology-based image navigation. It relies on the radiology knowledge source RadLex (Radiology Lexicon) [8] and the Annotation and Image Markup standard [9] and is applied to the context of magnetic resonance imaging of the brachial plexus. The developed strategy for harmonizing imaging data and anatomic metadata using knowledge representation and the resulted web-based service allows radiologists to explore both images and anatomic rela-

Introduction

As already mentioned [1-2], main ongoing works on Knowledge Representation and Management in the biomedical domain are related to semantic interoperability *i.e.* the ability of heterogeneous systems to exchange information between them in order to enable the use of this exchanged

information when well-informed acts and decisions for the well-being of the patients are needed. Indeed, health institutions have to deal with multiple types of data, in heterogeneous formats and from different sources, such as Electronic Health Records (EHRs), clinical images and reports, or genome sequences. On the other hand, the availability of distributed and

tions. In the context of surgery and human computer interaction in clinical practice, and to cope with unwanted consequences thanks to an investigation model, Machno et al. [10] showed the several advantages of using an ontology as information model instead of a traditional database. Several recent works are related to the extension of existing ontologies. For example, the Disease Ontology [11] is another resource developed to address common and rare diseases and available online in Web Ontology Language (OWL) and Open Biomedical Ontologies format. Wu et al. [12] recently developed the Disease Ontology cancer project, to complete the Disease Ontology, in which concepts and terms are gathered from several sources (such as The Cancer Genome Atlas) and linked to other resources (such as Online Mendelian Inheritance in Man, Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT)). The ontologies and terminologies included in the Unified Medical Language System [13] constitute resources in [14] to create the Disease Manifestation Network, a complementary phenotype network using manifestation data. The objective is to contribute in linking complex diseases to their genetic basis by strengthening candidate disease-gene selection methods [15]. In this context, text-mining is used by Groza et al. [16] to complete the Human Phenotype Ontology (HPO) [17], mainly focused on rare diseases, by adding phenotypic annotations for common diseases. Text-mining performed on several databases and full-text articles is also used by Liu et al. [18] for discovering relationships between biomedical entities including, among others, diseases, genes, drugs, toxins, proteins and metabolites. The developed tool PolySearch2 (<http://polysearch.ca>) relies on the latest web technology standards and uses Elasticsearch for managing document repositories and cache results. In the domain of heart failure summary, Martinez-Costa et al. [19] showed how semantic patterns and ontologies (domain ontology, top ontology and information entity ontology), all represented in OWL, enable querying heterogeneous representations of patient information in EHRs.

Method

The best paper selection for the section Knowledge Representation and Management follows a generic method, commonly used in all sections of the 2016 IMIA Yearbook and since 2013. As for the last three years, the search is performed on MEDLINE by querying PubMed. The Boolean query includes MeSH descriptors related to the domain of knowledge representation and management in the context of medical informatics with a restriction to international peer-reviewed journals. Only original research articles published in 2015 (from 01/01/2015 to 12/31/2015) were considered; we excluded the following publications types: reviews, editorials, comments, letters to the editors. We limited the search on the major MeSH descriptors (for example “biomedical ontologies [MAJR]”) to avoid a large set of articles and we completed it by non-MeSH terms searched on the titles and abstracts of the articles (for example “terminologies [TIAB]”). However, there was no restriction on the top international peer-reviewed journals of the Knowledge Representation and Management section (by using the 2-year Impact Factors).

Results

This year, the PubMed query related to the field Knowledge Representation and Management resulted in a set of 1,432 articles (a first search in September 2015 yielded 1,288 articles completed by a set of 144 articles in January to consider the articles published between September and December 2015). The articles were evaluated separately by each section editor (LFS & JC) using the BibReview tool and the generic method described by Lamy et al. in [20]. When started, BibReview loads a PubMed file (in XML format, generated from the query results) and shows all metadata for each article. A user can tag, thanks to the user-friendly interface, the articles as “Accepted”, “To Revise”, “Conflict” or “Reject” (according to the text, the abstract or the title of the publication). The results of several reviewers can be merged and the results can be

filtered. This year, 31 articles were tagged as “Accepted” by both section editors from which a set of 15 articles for peer-review is composed. A complementary list of articles is also needed to avoid doubletons with other IMIA Yearbook sections, mainly within the Natural Language Processing and Clinical Informatics sections. Several field specialized reviewers, considered the 15 articles as candidates for inclusion. The section editors as well as the 2016 IMIA Yearbook editors and managing editors also evaluated each article. The best papers are ranked according to criteria of: topic significance, coverage of literature, quality of research, results and presentation [21]. Finally, after evaluation, top-four papers [16-22-23-25] were retained by the section editors.

The first article [22] is related to the problem of gathering patient data for clinical research on breast cancer from several heterogeneous sources. Alonso-Calvo et al. present a standard-based solution founded on Resource Description Framework (RDF) triples, standard medical vocabularies (SNOMED CT, LOINC and Gene Names and their mappings) and SPARQL language for querying. A common information model, composed by a common data model on EHR and a core dataset based on the vocabularies, constitutes a semantic interoperability layer across the systems. The solution is tested efficiently (acceptable time for query expansion) on different types of queries and scenarios (patient screening, trial recruitment and retrospective analysis) in the context of several real scale European projects. In the second article [23], Semantic Web standards are used also for problem solving modeling and knowledge-based systems development from multiple sources. In the context of chronic kidney disease, Chi et al. describe a dietary consultation system founded on the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL) for knowledge inference. The system for food recommendation is tested and evaluated on real data patients by comparing the KBS inference and calculation to the human suggested dietary combinations. More accurate and faster results are obtained by the proposed system. The third article [16] concerns the work of Groza et al. in which the authors have developed a concept-recognition pro-

cedure that analyzes the frequencies of HPO disease annotations as identified in over five million PubMed abstracts indexed with MeSH terms. The objective is to complete the Human Phenotype Ontology (HPO) [17], mainly focused on rare diseases, by adding phenotypic annotations for common diseases leading to a semantic unification of common and rare diseases founded on semantic-similarity scores. They used an iterative procedure to optimize precision and recall of the identified terms (MeSH entries belonging to the Disease Ontology) and they derived disease models for 3,145 common human diseases comprising a total of 132,006 HPO annotations. The fourth article [24] is related to ontology verification. Mortensen et al. performed ontology verification by crowdsourcing on SNOMED-CT. Indeed, the SNOMED-CT includes logic-based definitions to represent terminological knowledge that are sometimes inconsistent and may lead to incoherent decision support systems or querying systems. The developed method combines micro-tasking with a Bayesian classifier and allows to identify errors in a subset of SNOMED CT comprising 200 taxonomic relationships. This method may be useful in situations where an expert is unavailable.

Table 1 lists the four papers selected as best papers for the section Knowledge Representation and Management. A brief content of each one can be found in the appendix of this synopsis.

Conclusions

The articles described herein, particularly the top-four, are characterized by their specific application. Indeed, each evaluation focused on how the work described satisfied goals that are directly important for clinical or translational medicine. In the recent research activities, EHRs are valuable sources of information for knowledge discovery. The integration of genomic data into EHRs as well as the development of genomic tests and their increasing clinical utility can change the medical decision process [15]. The recent research and development efforts are contributing to the challenge of impacting clinically the results and even going towards

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2016 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"> ▪ Alonso-Calvo R, Perez-Rey D, Paraiso-Medina S, Clearhout B, Hennebert P, Bucur A. Enabling semantic interoperability in multi-centric clinical trials on breast cancer. <i>Computer Methods Programs Biomed</i> 2015;118:322–9. ▪ Chi YL, Chen TY, Tsai WT. A chronic disease dietary consultation system using OWL-based ontologies and semantic rules. <i>J Biomed Inform</i> 2015;53:208–19. ▪ Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, Schriml LM, Kibbe WA, Schofield PN, Beck T, Vasant D, Brookes AJ, Zankl A, Washington NL, Mungall CJ, Lewis SE, Haendel MA, Parkinson H, Robinson PN. The Human Phenotype Ontology : semantic unification of common and rare disease. <i>Am J Hum Genet</i> 2015;97:111–24. ▪ Mortensen JM, Minty EP, Januszkyk M, Sweeney TE, Rector AL, Noy NF, Musen MA. Using the wisdom of the crowds to find critical errors in biomedical ontologies : a study of SNOMED CT. <i>J Am Med Inform Assoc</i> 2015;22:640–8.

a personalized medicine in the near future. In this context, data standards such as Logical Observation Identifiers Names and Codes or SNOMED-CT, which provide a highly comprehensive and detailed set of clinical terms used in many systems to enrich the information in EHRs, do remain good candidates for improving semantic interoperability [25].

Acknowledgements

We would like to acknowledge the valuable support of Martina Hutter and all the reviewers in the evaluation process of the section Knowledge Representation and Management. We also would like to greatly thank the IMIA Yearbook 2016 editors and managing editors Marie-Christine Jaulent, Brigitte Séroussi and Christoph U Lehmann.

References

1. Griffon N, Charlet J, Darmoni SJ. Knowledge representation and management: towards an integration of a semantic web in daily health practice. *Yearb Med Inform* 2013;8(1):155–8.
2. Charlet J, Darmoni SJ. Knowledge Representation and Management. From Ontology to Annotation. *Yearb Med Inform* 2014;9(1):134–6.
3. Anguita A, Martin L, Pérez-Rey D, Maojo V. A review of methods and tools for database integration in biomedicine. *Curr Bioinform* 2010;5(4):253–69.
4. Langenhove PV, Rogala A, Olyslaegers T, Whitehouse D. eHealth Interoperability Framework Study. 2013; Technical report, European Communities.
5. Moor GD, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using Electronic Health Records for Clinical Research: the case of the EHR4CR project. *J Biomed Inform* 2015;53:162–73.
6. Batrancourt B, Dojat M, Gibaud B, Kassel G. A multilayer ontology of instruments for neurological, behavioral and cognitive assessments. *Neuroinform* 2015;13:93–110.
7. Wang KC, Salunkhe AR, Morrison JJ, Lee PP, Mejino JLV, Detwiler LT, et al. Ontology-based image navigation: exploring 3.0-T MR neurography of the brachial plexus using AIM and RadLex. *RadioGraphics* 2015;35:142–51.
8. Langlotz CP. RadLex: a new method for indexing online educational materials. *Radiographics* 2006;26:1595–7.
9. Channin DS, Mongkolwat P, Kleper V, Rubin DL. The annotation and image mark-up project. *Radiology* 2009;253(3):590–2.
10. Machno A, Jannin P, Dameron O, Korb W, Scheuermann G, Meixenberger J. Ontology for assessment studies of human-computer-interaction in surgery. *Artif Intell Med* 2015;63:73–84.
11. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease ontology 2015 update: an expanded database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res* 2014;43:D805–11.
12. Wu TJ, Schriml LM, Chen QR, Colbert M, Crichton DJ, Finney R, et al. Generating a focused view of disease ontology cancer terms for pan-cancer data integration and analysis. *Database* 2015;1–10.
13. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004;32:D267–70.
14. Chen Y, Zhang X, Zhang G, Xu R. Comparative analysis of a novel disease phenotype network based on clinical manifestations. *J Biomed Inform* 2015;53:113–20.
15. Lecroq T, Soualmia LF. Managing Large-Scale Genomic Datasets and Translation into Clinical Practice. *Yearb Med Inform* 2014;9(1):212–4.
16. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The Human Phenotype Ontology : semantic unification of common and rare disease. *Am J Hum Genet* 2015;97:114–24.
17. Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids*

- Res 2014;42:D966–74.
18. Liu Y, Liang Y, Wishart D. PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res* 2015;43:W535–42.
 19. Lamy JB, Séroussi B, Griffon N, Kerdelhué G, Jaulent MC, Bouaud J. Toward a formalization of the process to select IMIA Yearbook best papers. *Methods Inf Med* 2015;54(2):135–44.
 20. Ammenwerth E, Wolff AC, Knaup P, Ulmer H, Skonetzki S, van Bommel JH, et al. Developing and evaluating criteria to help reviewers of biomedical informatics manuscripts. *J Am Med Inform Assoc* 2003;10:512–4.
 21. Martínez-Costa C, Cornet R, Karlsson D, Schulz S, Kalra D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. *J Am Med Inform Assoc* 2015;22:565–76.
 22. Alonso-Calvo R, Perez-Rey D, Paraiso Medina S, Claerhout B, Hennebert P, Bucur A. Enabling semantic interoperability in multi-centric clinical trials on breast cancer. *Comput Methods Programs Biomed* 2015;118:322–9.
 23. Chi YL, Chen TY, Tsai WT. A chronic disease dietary consultation system using OWL-based ontologies and semantic rules. *J Biomed Inform* 2015;53:208–19.
 24. Mortensen JM, Minty EP, Januszzyk M, Sweeney TE, Rector AL, Noy NF, et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies : a study of SNOMED CT. *J Am Med Inform Assoc* 2015;22:640–8.
 25. Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. *J Am Med Inform Assoc* 2015;22:621–7.

Correspondence to:

Dr Lina F. Soualmia
 Normandie Universités, Rouen University and Hospital
 D2IM, LITIS EA 4108
 Information Processing in Biology & Health
 1, rue de Germont, Cour Leschevin porte 21
 76031 Rouen Cedex, France
 Tel : +33 232 885 869
 E-mail: Lina.Soualmia@chu-rouen.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2016, Section Knowledge Representation and Management

Alonso-Calvo R, Perez-Rey D, Paraiso-Medina S, Clearhout B, Hennebert P, Bucur A
Enabling semantic interoperability in multi-centric clinical trials on breast cancer

Computer Methods Programs Biomed 2015;118:322–9

In this paper, the authors present a standard-based solution to provide a uniform access endpoint to patient data involved in current clinical research. In this work, clinical trials' data on breast cancer are integrated as RDF triples with respects to medical vocabularies (mainly SNOMED-CT, LOINC for laboratory test and specific Gene Names related to breast cancer). With a specific mechanism where data are recorded in a Common Data Model linked to a Core Dataset, which gives access to semantic models, the SPARQL query language allows to retrieve efficiently the data. To validate their solution, the authors tested different types of queries and developed an end-user tool within a framework that requires efficient retrieval of integrated data. Several prototypes have been deployed and tested for: (i) patient screening, (ii) trial recruitment, and (iii) retrospective analysis. A site with virtual patients is also available for testing and analyzing models and queries.

Chi YL, Chen TY, Tsai WT

A chronic disease dietary consultation system using OWL-based ontologies and semantic rules

J Biomed Inform 2015;53:208–19

The work is developed in the context of chronic kidney disease dietary consultation. The system developed is founded on the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL) for knowledge inference. The system development needs to integrate multiple knowledge sources for problem solving modeling and knowledge-based system (KBS) development. The KBS development is conducted with several design elements : problem scenarios, domain ontology, task ontology and semantic rules. A set of 84 real case patients was used to evaluate the system in recommending appropriate food serving amounts from different food groups for balanced key nutrient ingestion. This paper is interesting for precision descriptions, particularly dietary rules coded in SWRL.

Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, Schriml LM, Kibbe WA, Schofield PN, Beck T, Vasant D, Brookes AJ, Zankl A, Washington NL, Mungall CJ, Lewis SE, Haendel MA, Parkinson H, Robinson PN

The Human Phenotype Ontology: semantic unification of common and rare disease

Am J Hum Genet 2015;97:111–24

The goal of the study related in this paper is to complete the Human Phenotype Ontology (HPO), mainly focused on rare diseases, by adding phenotypic annotations for common diseases. The authors have developed a concept-recognition procedure that analyzes the frequencies of HPO disease annotations as identified in over five million PubMed abstracts by employing an iterative procedure to optimize precision and recall of the identified terms. They derived disease models for 3,145 common human diseases comprising a total of 132,006 HPO annotations.

An efficient concept recognition system was developed to generate and validate the common-disease annotations. The platform the authors made available together with the data is in itself a valuable resource for the community.

Mortensen JM, Minty EP, Januszzyk M, Sweeney TE, Rector AL, Noy NF, Musen MA

Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT

J Am Med Inform Assoc 2015;22:640–8

The goal of the study presented in this paper aims at performing ontology verification by crowdsourcing. The authors developed a methodology that uses micro-tasking combined with a Bayesian classifier. They then conducted a prospective study in which both the crowd and domain experts verified a subset of SNOMED-CT comprising 200 taxonomic relationships. The crowd can indeed identify errors in SNOMED CT that experts also found. The authors note that such a methodology may be useful in situations where an expert is unavailable.