

Knowledge Representation and Management: Towards Interoperable Medical Terminologies

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

University Hospitals of Geneva, Geneva, Switzerland

Summary

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

Method: Synopsis of the articles selected for the IMIA Yearbook 2009.

Results: Four excellent papers have been selected for the section knowledge representation and management. All these papers are concerned with terminological systems whether it is to detect equivalent concept definitions, to structure narratives, to compare existing lay and professional terms or to reformulate the SNOMED's logical formalism towards a more expressive language.

Conclusions: The ability of one computer system to access and use the resources of another system becomes crucial in a world where the amount of electronically stored data increases continuously. The selected papers for the section knowledge representation and management corroborate that interoperable medical terminologies play a growing and strategic role in Electronic Health Record (EHR) systems. Indeed, they allow data in clinical systems to be defined in a more uniform and granular manner, thus leading to flexible, semantically interoperable and trust-worthy health information systems.

Keywords

Description logic, concept definitions, structured narrative, interoperable terminology

Yearb Med Inform 2009:99-102

Introduction

With years, concept-oriented controlled terminologies have proved to be well suited for representing, in a structured and formal way, the entities and relationships that characterize a given biomedical domain. They support a large spectrum of purposes whether it is to store patient information, to provide decision support or to sustain resource management as well as clinical research. In order to hold up this growing range of purposes, medical terminologies have become larger and therefore more complex and more diversified. The emerging challenge nowadays is to update and harmonize existing terminological standards in order to insure their semantic compatibility and interoperability [1, 2].

The topic of the IMIA Yearbook 2009: "Biomedical Informatics: Closing the Loops" is also relevant for medical terminologies. How can we close the loop for terminologies in order to guarantee sustainable and interoperable infrastructure in healthcare? Rendering terminologies easily accessible and useful for healthcare providers implies to consider both basic aspects of interoperability. On the one hand, the ability to access data, referring to syntactic interoperability, aims at representing terminology in a more expressive formalism. Description logic has proved to be particularly compliant for this purpose [3]. On the other hand, the ability to understand the data once retrieved, referring to semantic interoperability [4], compels

terminologies to provide more definitions, ideally one for each non-primitive concept. In order to reach these interoperable levels, a more advanced management of existing terminologies is required, in particular to maintain, audit as well as to improve their content, as emphasized by the recent literature covering these aspects [5, 6, 7, 8, 9, 10].

Best Paper Selection

Four outstanding papers have been elected for the section knowledge representation and management of the IMIA Yearbook 2009 (see Table 1). A brief content summary of each selected article can be found in the appendix of this synopsis. These papers draw attention to the various aforementioned terminological issues.

Unambiguously and accurately defining medical concepts is a prerequisite to ensure reliable and truthful communication and cooperation between actors involved in patient's care, whether they are persons or applications. Cornet and Abu-Hanna [11] present a method for detecting equivalent concept definitions embedded into description-logic-based medical terminological systems. Duplicate definition as well as underspecification greatly reduces the powers of knowledge-based inference of a terminological system and therefore its semantic interoperability.

The paper of Johnson et al. [12], already quoted in the IMIA Yearbook 2008 [13], proposes a design in

which unstructured text and coded data are fused into a single model called structured narrative. A real time natural language processing module, coupled with an ontology of document types and sections, are used to analyze clinical documents and identify relevant concepts and their modifiers. These resulting semantically annotated documents are designed to interoperate more easily with other clinical information systems.

Keselman et al. [14] report on the results of a large study for identifying and characterizing consumer health terms, which are not found in professional controlled vocabularies. Even within the same spoken natural language, people do not convey the same meaning through a particular terminology. This is why, filling the terminology gap between consumer and professional health discourse will also contribute to the interoperability of healthcare information systems.

Finally, to facilitate interoperability of a huge terminological system such as SNOMED, the fourth paper [15] describes a feasibility study that aims at reformulating the SNOMED's logical formalism into more expressive description logics. Adding a more formal and comprehensive framework to SNOMED will henceforth improve its scalability and reusability.

Conclusions and Outlook

Existing medical terminologies are more and more exploited by clinical actors whether they are persons or computer-based applications, as emphasized by the various works described in the set of best papers, selected for the section knowledge representation and management of the IMIA Yearbook 2009. The different national languages as well as the various intended usages of terminologies bring to light that no single, common terminology will ever exist. However, harmonizing concept representation, languages and

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2009 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"> ▪ Cornet R, Abu-Hanna A. Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions. <i>Int J Med Inform</i> 2008 May;77(5):336-45. ▪ Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P. An electronic health record based on structured narrative. <i>J Am Med Inform Assoc</i> 2008 Jan-Feb;15(1):54-64. ▪ Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, Zeng-Treitler Q. Consumer health concepts that do not map to the UMLS: where do they fit? <i>J Am Med Inform Assoc</i> 2008 Jul-Aug;15(4):496-505. ▪ Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. <i>J Am Med Inform Assoc</i> 2008 Nov-Dec;15(6):744-51.

methodologies emerges as the basic challenge towards interoperable information systems. This is why, collaborative efforts, for achieving true interoperability of existing medical terminologies on a multilingual basis, are crucial and are promised to a rich future [16, 17].

Acknowledgement

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

References

1. Russell A, Hamm, Sarah E, Knoop, Peter Schwarz, Aaron D. Block, Warren L. Davis IV. Harmonizing Clinical Terminologies: Driving Interoperability in Healthcare. *Stud Health Technol Inform* 2007; 129(Pt 1):660-3.
2. Lopez DM, Blobel B. Enhanced Semantic Interoperability by Profiling Health Informatics Standards. *Methods Inf Med* 2009;48:170-7.
3. Bodenreider O, Smith B, Kumar A, Burgun A. Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies. *Artif Intell Med* 2007;39: 183-95.
4. Lopez DM, Blobel B. A development framework for semantically interoperable health information systems. *Int J Med Inform* 2009;78:83-103.
5. Rogers JE. Quality Assurance of Medical Ontologies. *Methods Inf Med* 2006;3:267-74.
6. Gambarte ML, Osornio AL, Martinez M, Reynoso G, Luna D, de Quiros FG. A practical approach to advanced terminology services in health information systems. *Stud Health Technol Inform* 2007; 129(Pt 1):621-5.
7. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. A Model for Evaluating Interface Terminologies. *J Am Med Inform Assoc* 2008; 15(1):65-76.
8. Bakhashi-Raiez F, Cornet R, De Keizer NF. Development and Application of a Framework for Maintenance of Medical Terminological Systems. *J Am Med Inform Assoc* 2008;15(5):687-700.
9. Bodenreider O. Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. In: Geissbuhler A, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2008. Methods Inf Med 2008; 47 Suppl 1:67-79.
10. Baorto D, Li L, Cimoni JJ. Practical experience with the maintenance and auditing of a large medical ontology. *J Biomed Inform*. In Press 2009.
11. Cornet R, Abu-Hanna A. Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions. *Int J Med Inform* 2008;77(5):336-45.
12. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P. An electronic health record based on structured narrative. *J Am Med Inform Assoc* 2008; 15(1):54-64.
13. Rassinoux AM. Decision Support, Knowledge Representation and Management: Structuring Knowledge for Better Access. In: Geissbuhler A, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2008. Methods Inf Med 2008; 47 Suppl 1:80-2.
14. Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, Zeng-Treitler Q. Consumer health concepts that do not map to the UMLS: where do they fit? *J Am Med Inform Assoc* 2008;15(4):496-505.
15. Rector AL, Brandt S. Why do it the hard way? The case for an expressive description logic for SNOMED. *J Am Med Inform Assoc* 2008;15(6):744-51.
16. Donnelly K. Multilingual documentation and classification. *Stud Health Technol Inform* 2008;134: 235-43.
17. Rodrigues JM, Kumar A, Bousquet C, Trombert B. Standards and biomedical terminologies: the CEN TC 251 and ISO TC 215 categorial structures. A step towards increased interoperability. *Stud Health Technol Inform* 2008;136:857-62.

Correspondence to:
 Anne-Marie Rassinoux, Ph. D.
 University Hospitals of Geneva
 Service of Medical Informatics
 Unit of Clinical Informatics
 4, Rue Gabrielle-Perrin-Gentil
 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 2009, Section Knowledge Representation and Management*

Cornet R, Abu-Hanna A

Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions

Int J Med Inform 2008; 77(5):336-45

Duplicate concept definitions as well as underspecified concepts reduce the quality of terminologies as they may severely hamper reasoning and knowledge-based querying. Automatic discovery of these equivalently defined concepts is then challenging and possible for medical terminological systems (TSs) that are represented in an unambiguous way, through description logic (DL). The method described by the authors, consists of auditing defined concepts (also called non-primitive concepts as they are described through necessary and sufficient conditions) using a fully automated DL reasoner. Such DL reasoning systematically picks out equivalent concepts which are further interpreted by modelers either as redundant, i.e. defined more than once, or underspecified. A case study, carried out on the DICE knowledge base, has

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

permitted to highlight that around 25% of non-primitive concepts have definitions that are logically equivalent to those of other concepts. Analysis of these later reveals that most of these concepts need to be more completely defined either by adding existing relations or by introducing new relations into the terminological system.

Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P

An electronic health record based on structured narrative

J Am Med Inform Assoc 2008; 15(1):54-64

Representing health information in highly structured ways is of paramount interest to support subsequent computer processing such as decision support or quality assurance. The authors put forward a single model called "structured narrative" that combines the advantages of both unstructured text and coded data. Indeed, narrative is more accurate, more comprehensive, more familiar to clinicians, whereas coded data are highly reusable and thus more adapted for resource management and billing. In the proposed structured narrative model, texts are transformed into Extensible Markup Language (XML) documents thanks to a real time natural language processing (NLP) module that allows texts to be marked up with coded information at any level of detail. Indeed, these detailed semantic markups apply both to the gross structural level of documents (such as sections, fields, paragraphs and lists to any level of nesting), and to the fine structural level (such as medical concepts and their modifiers and relationships). All these produced XML documents are then stored within a database system that supports multiple ways to browse the patient's collection of documents by means of both a document ontology and inference engine that help filtering the results. The great benefit of this approach is that the coded information obtained through this model may en-

able efficient reuse of narrative material while reducing the effort required by physicians to enter coded data.

Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, Zeng-Treitler Q

Consumer health concepts that do not map to the UMLS: where do they fit?

J Am Med Inform Assoc 2008; 15(4):496-505

Nowadays, health consumers are more and more expected to act as partners in their healthcare. But this empowerment of healthcare consumers is hampered by the gap existing between lay and professional health terminologies. Indeed, lay people have difficulties understanding professional health discourse and this affects their ability to browse health-related websites and to communicate with their physicians. The study presented by Keselmen et al. aims first at identifying consumer health terms that are not found in professional controlled vocabularies such as the Unified Medical Language System (UMLS) Metathesaurus. Second, these unmapped consumer health concepts are further examined in order to determine if they relate in some way to the existing UMLS, or if they are truly novel or genuine lay concepts. These issues are considered for the Open Access and Collaborative Consumer Health Vocabulary (OAC CHV) initiative, with which the authors of this paper are affiliated. Of the 1046 terms extracted by the OAC CHV development team, only 64 could not be mapped to existing UMLS concepts. However, 47 of these terms denoted concepts that could be represented in professional medical discourse whereas the remaining 17 terms were classified as primarily lay. The authors conclude that while non-mapping terms represent a small, but non negligible proportion of health terms used by lay people, they constitute a valuable resource for understanding the structure of lay conceptual knowledge in health. Therefore, studies considering other domains and settings, where deviations from traditional

professional views are likely to abound, must be further conducted.

Rector AL, Brandt S

Why do it the hard way? The case for an expressive description logic for SNOMED

J Am Med Inform Assoc 2008; 15(6):744-51

Several countries around the world have selected SNOMED-CT as the reference terminology for their national health care institutions. Since its initial development in 1990s, SNOMED has kept the same underlying structure based on the Ontylog formalism, for which, several restrictions and ontological problems have been clearly emphasized by researchers using it. In

this paper, Rector and Brandt advocate the reformulation of the SNOMED's logical formalism towards a more expressive schema such as the W3C standard Web Ontology Language (OWL). The use of a more expressive language will lead to a more regular and uniform representation of concepts including negation and contexts as well as to a better defined and more consistent semantics allowing the recognition of equivalence between findings and observables. Moreover, clarifying the meanings behind the classes allows these later to be automatically classified on the basis of their definitions in the ontology, thus leading to more flexible hierarchies that can be further

partitioned into OWL modules adapted for different use cases. These modules are also easier to maintain and exchange than would be a monolithic ontology. The authors conclude that a feasibility study, based on a limited subset of SNOMED is strongly advocated to corroborate the practicability of reformulating SNOMED into OWL. In particular, they put ahead the difficulties linked to the huge size of SNOMED and the many errors and irregularities of the existing classification content as well as the fact that not only the ontology itself but also the applications and metadata already based on, need to be migrated to an OWL environment.