Knowledge Representation and Management: Towards an Integration of a Semantic Web in Daily Health Practice

N. Griffon1, J. Charlet2, 3, SJ. Darmoni1, Section Editors for the IMIA Yearbook 2013, Section on Knowledge Representation and Management

1 CISMeF-TIBS-LITIS EA 4108, Rouen University Hospital, Rouen, France
2 AP-HP, Dept. of Clinical Research and Development, Paris France
3 INSERM UMRS 872 eq. 20, CRC, Paris, France

Summary

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

Methods: A synopsis of the four selected articles for the IMIA Yearbook 2013 KRM section is provided, as well as highlights of current KRM trends, in particular, of the semantic web in daily health practice. The manual selection was performed in three stages: first a set of 3,106 articles, then a second set of 86 articles followed by a third set of 15 articles, and finally the last set of four chosen articles.

Results: Among the four selected articles (see Table 1), one focuses on knowledge engineering to prevent adverse drug events; the objective of the second is to propose mappings between clinical archetypes and SNOMED CT in the context of clinical practice; the third presents an ontology to create a question-answering system; the fourth describes a biomonitoring network based on semantic web technologies.

Conclusion: These four articles clearly indicate that the health semantic web has become a part of daily practice of health professionals since 2012. In the review of the second set of 86 articles, the same topics included in the previous IMIA yearbook remain active research fields: Knowledge extraction, automatic indexing, information retrieval, natural language processing, management of health terminologies and ontologies.

Keywords

Semantic web, knowledge management, abstracting and indexing as a topic, vocabulary, controlled, biomonitoring;

Best Paper Selection

For the KRM section of the IMIA Yearbook 2013, four papers were selected based on the following review process [1]: (1) two queries were launched in PubMed and Web of Science as these two bibliographic databases were considered by consensus of IMIA Yearbook section editors to be the most relevant. These two queries are available at the following URL: http://www.chu-rouen.fr/documed/yearbook12/krm.html and in order for the results to be replicated need to be cut and pasted respectively in PubMed and in Web of Science. The main MeSH terms used in the PubMed query were: (“Knowledge”[MH] OR “Knowledge bases”[MH] OR “Knowledge management”[MH] OR “Abstracting and indexing as topic”[MH] OR “Vocabulary, controlled”[MH]). A number of terms were added to cover more articles using [TIAB] (Title or Abstract) metadata. A total of 3,506 unique articles were obtained from these two queries. One section editor (NG) reviewed each article based on the title and abstract and selected 86 papers. The main selection criterion was the adequacy to the theme section. These 86 papers were then reviewed by the three section editors. Fifteen articles were selected by consensus. Then, each of these 15 articles was read and rated by the same three experts. The list was subsequently discussed by all section editors and the two co-editors during a one day conference meeting. The top four articles were finally selected. These four papers are shown in Table 1 and are summarized in the appendix of this synopsis.

Koutkias et al. [2] proposed a contextualized knowledge-based framework to prevent adverse drug events via a specific computer-aided decision support system. This study was developed and granted as part of the PSIP project [3], and funded by the European Union 7th Framework Program. Meizoso García et al. [4] developed and evaluated mappings between clinical archetypes and SNOMED CT in the context of clinical practice. Several Natural Language Processing (NLP) tools were used to implement these mappings. Patrick and Li [5] developed a very original paper on building an ontology to create a clinical question-answering (QA) system based on clinical notes. Teodoro et al. [6] evaluated and have described a biomonitoring network based on semantic web technologies. This ARTEMIS network provides a real-time antimicrobial...
resistance monitoring system, and is already piloted in several hospitals. This work was developed during the DebugIT project [7], and also funded by the European Union 7th Framework Program.

When reviewing the broader second set of 86 selected articles, the same topics included in the previous IMIA yearbook [8] were identified to still be active research topics by the three section editors of the KRM i.e.: knowledge extraction, automatic indexing, information retrieval, natural language processing, management of health terminologies and ontologies using semantic web technologies. A particular hot topic was the mapping methods and techniques between archetypes, interface terminologies as well as reference terminologies and ontologies. These mappings are mandatory basic tools to create evidence-based biomedical informatics. This is particularly true for computer-aided decision systems.

**Conclusion and Outlook**

The four selected articles for the KRM section clearly confirm that the health semantic web is now an integral part of daily practice of health professionals in 2012. Several specialized conferences already exist: Knowledge Representation in Medicine (KR-Med) [9] organized by the AMIA working group Knowledge Representation and Semantics [10], the international conference on biomedical ontology (ICBO) [11], and Semantic Web Applications and Tools for Knowledge Sciences (SWAT4LS) [12]. An interest group has also been created on this topic within W3C: the Semantic Web Health Care and Life Sciences Interest Group [13]. In 2011, a network of excellence was funded by the European Union on the same topic: Semantic Interoperability for Health Network (SemanticHealthNet) [14]. Its objective is to develop a scalable and sustainable pan-European organizational and governance framework for the semantic interoperability of clinical and biomedical knowledge, and to help ensure that electronic health record (EHR) systems are optimized for patient care, public health and clinical research across healthcare systems and institutions [14]. Health terminologies and ontologies (HTO) in health information systems are spreading in English via BioPortal [15] and other languages, in particular in French with more than 30 HTO available via a cross-lingual portal (URL: www.hetop.eu) [16]. Informatics for Integrating Biology & the Bedside (I2B2) [17] with its *omics extension TRANSMART (genomics, proteomics, metabolomics, etc) is an operational platform to enable clinical researchers to perform information retrieval in EHRs. I2B2/Transmart is becoming a “de facto” standard used by a considerable number of university hospitals around the world to practice translational medicine by health professionals.

**Acknowledgments**

The authors thank Martina Hutter for her consistent support and Richard Medeiros, Rouen University Hospital Medical Editor, for editing this manuscript.

**References**


**Table 1** Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2010 in the sections ‘Knowledge Representation and Management. The articles are listed in alphabetical order of the first author’s surname.

<table>
<thead>
<tr>
<th>Section</th>
<th>Knowledge Representation and Management</th>
</tr>
</thead>
</table>
rule-based adverse drug event signals, all

dynamic knowledge sources in the form of
The framework integrates diverse and dy-
implemented by adopting Computer In-
architectural framework was specifi c and
model had been defi ned. The entire ar-
elaborated in the scope of this work, for
formalization of the knowledge sources
involving the systematic analysis and
European (PSIP) project. The methodol-
drug event prevention via Clinical Decision
This article presents a sustainable knowl-
sustainable knowledge-based framework
opment of a uniform, contextualized and
Knowledge Representation and Management1

Knowledge engineering for adverse drug
event prevention: on the design and develop-
ment of a uniform, contextualized and
knowledge-based framework
J Biomed Inform 2012 Jun;45(3):495-506

This article presents a sustainable knowledge-
based framework to support adverse drug
event prevention via Clinical Decision
Support Systems. The framework was de-
developed during the Patient Safety Through
Intelligent Procedures in Medication European (PSIP) project. The methodol-
yology involved the systematic analysis and
formalization of the knowledge sources elaborated in the scope of this work, for
which an application-specifi c knowledge
model had been defi ned. The entire ar-
chitectural framework was specifi ed and
implemented by adopting Computer In-
terpretable Guidelines as the knowledge
ing engineering formalism for its construc-
tion. The framework integrates diverse and dy-
namic knowledge sources in the form of
rule-based adverse drug event signals, all
under a uniform Knowledge Base structure, according to the defi ned knowledge model.
Equally important, it employs the means to
contextualize the encapsulated knowledge,
in order to provide appropriate support
considering the specifi c local environment
(hospital, medical department, language,
etc.), as well as the mechanisms for knowl-
edge querying, inference, sharing, and
management.

Therefore, in this paper, the authors
insist on the results obtained: implementa-
ion performance and validation aspects
that highlight its applicability and virtue in
medication safety.

MeizosoGarcía M, Iglesias Allones JL,
Martínez Hernández D, Taboada Iglesias MJ
Semantic similarity-based alignment be-
tween clinical archetypes and SNOMED CT:
an application to observations
Int J Med Inform 2012 Aug;81(8):566-78

Clinical data models, such as OpenEHR
Archetypes, deifi ne data structures that
are agreed upon by experts to ensure the
accuracy of health information. In addition,
they provide an option to normalize clinical
data by means of binding terms used in the
model defi nition in standard medical vocab-
ularies. Nevertheless, the effort needed to
establish the association between archetype
terms and standard terminology concepts is
considerable. Therefore, the purpose of this
study was to provide an automated approach
to bind OpenEHR archetypes terms to the
external terminology SNOMED CT, with
the capability of performing alignment at
a semantic level.

The authors propose to use lexical tech-
niques and external terminological tools in
combination with context-based techniques,
which use information about structural
and semantic proximity to identify simi-
larities between terms in order to establish
alignments between them. The proposed
approach exploits both the structural context
of archetypes and the terminology con-
text, where concepts are logically defi ned
through the relationships (hierarchical and
defi nitional) to other concepts. A set of 25
ARCHETYPE archetypes with 477 bound
terms was used to test the method. Of these,
342 terms (74.6%) were linked with 96.1%
precision, 71.7% recall and 1.23 SNOMED
CT concepts on average for each mapping.
It has been detected that about one third of
the archetype clinical information is grouped
logically. Context-based techniques take
advantage of this to increase the recall and
to validate a 30.4% of the bindings produced
by lexical techniques.

This research demonstrates the interest in
establishing the modeling of data structures
on ontologies. It also shows that it is possible
to semi-automatically map archetype terms
to a standard terminology with a high level of
precision and recall, with the help of appro-
propriate contextual and semantic information
from both models. Thus, as in other papers
of this selection, contextualization is the
key word in the implementation of semantic
technologies in information systems.

Patrick J, Li M
An ontology for clinical questions about the
contents of patient notes
J Biomed Inform 2012 Apr;45(2):292-306

The considerable amount of information
available in electronic patient records make
it an attractive resource for answering a
variety of questions that users may have.
Current information retrieval techniques
have proven quite successful at locating
patient records that might be relevant to a
user’s query. Many studies have been com-
pleted on question classifi cation in the open
domain, however only a limited amount
of work has focused on the medical fi eld.
Also, to the best of our knowledge, most of
these medical question classifi cations were
designed for literature based question and
answering systems.

The authors describe a four stage ap-
proach. Firstly, a relatively large set of
clinical questions was collected from staff
in an Intensive Care Unit. Then, a clinical
question on taxonomy was designed for
question and answering purposes. Subse-
quently an annotation guideline was created
and used to annotate the question set. Fi-
nally, a multilayer classifi cation model was
built to classify the clinical questions. In this
paper, a comprehensive study on clinical
questions was completed. A major outcome

1 The complete papers can be accessed in
the Yearbook’s full electronic version,
provided that the article is freely
accessible of that your institution has
access to the respective journal.
of this work is the multilayer classification model. It serves as a basic component of a patient’s record based on a clinical question and answering system. Also, the question collections can be reused by the research community to improve the efficiency of their own question and answering systems.

On the basis that there will always be free text available in the Electronic Health Record (EHR) system, it will subsequently be necessary to develop a natural language processing system of information retrieval in the EHR. The question of how to do this, with a Q/A system or a semantic-based information retrieval system will arise and depend on the type of GUI institutions want to set up, as well as the context.

Teodoro D, Pasche E, Gobeill J, Emonet S, Ruch P, Lovis C
Building a transnational biosurveillance network using semantic web technologies: requirements, design, and preliminary evaluation
J Med Internet Res 2012 May 29;14(3):e73

Antimicrobial resistance has reached globally alarming levels and is becoming a major public health threat. Lack of effective antimicrobial resistance surveillance systems has been identified as one of the causes of increasing resistance, due to the time lag between new resistances and alerts to care providers. Several initiatives to track drug resistance evolution have been developed. However, no effective real-time and source-independent antimicrobial resistance monitoring system is available publicly. In order to propose an effective architecture that can provide real-time and source-independent antimicrobial resistance monitoring, the authors investigated the use of a Semantic Web-based model to foster integration and interoperability of inter-institutional and cross-border microbiology laboratory databases. The architecture uses an ontology-driven approach to promote the integration of a network of sentinel hospitals or laboratories. Local databases are wrapped into semantic data repositories that automatically expose local computing-formalized laboratory information on the Web. A central source mediator, based on local reasoning, coordinates the access to the semantic end points. On the user side, a user-friendly Web interface provides easy access and graphical visualization to the integrated views.

The developed prototype, the Antimicrobial Resistance Trend Monitoring System has a strong positive correlation with the European Antimicrobial Resistance Surveillance Network. The distributed monitoring architecture developed during the DebugIT European project between 2008 and 2012 can be used to build transnational antimicrobial resistance surveillance networks, and results have indicated that the Semantic Web-based approach has provided an effective and reliable solution for development of eHealth architectures.

As for the Koutkias’ paper, we have observed that ontology-based architecture allows proposals of systems, which integrate heterogeneous data. It in fact also offers services in the field of public health.