Health Informatics Journal

Editorial

Health Informatics Journal
17(2) 91–94

DOI: 10.1177/1460458211405003
jhi.sagepub.com
SAGE

Special issue on semantic descriptions of medical web resources: Technologies to support their creation, maintenance and access

The internet is designed for direct use by people searching for information through their browsers; accessing health-related information has never been easier. The increasing amount of freely-available, health-related web content creates, on the one hand, excellent conditions for self-education of patients as well as physicians but, on the other hand, however, it entails substantial risks if such information is trusted, irrespective of the low competence, or even bad intentions, of its authors. Furthermore, it is difficult for health information consumers, such as patients and the general public, to assess, by themselves, the quality of the information, as they are not always familiar with medical domains and vocabularies.[1]

Semantic Web technologies have been proposed as a way to address this problem. The Semantic Web is "an extension of the current Web in which information is given a well-defined meaning, better enabling computers and people to work in cooperation".[2] It is based on metadata; that is, on semantic annotations of web content. These metadata can be expressed in different ways using the Resource Description Framework (RDF) language[3] and Web Ontology Language.[4] Using these standardized languages, which express both data and rules for connecting and reasoning about the data, it is possible to create machine-processable descriptions (labels) of web resources.

To support patients and consumers in retrieving suitable information sources, a number of quality labelling initiatives have been developed across Europe.[5, 6] One approach is the "trust mark" method, where a third-party agency ascertains, on a regular basis, whether the quality of the information on the website is acceptable or not. Another approach is when a third-party authority selects, or filters, websites for the public to use.[7] Both approaches can be supported with the use of Semantic Web tools and metadata that facilitate health-related web resources description, classification and retrieval.[8]

The aim of this special issue of *Health Informatics Journal* is to reflect progress in the use of Semantic Web techniques for the description of health-related web resources and to outline the major challenges and future prospects. It builds on the 1st International Workshop on Describing Medical Web Resources (DRMed 2008),[9] held in conjunction with the 21st International Congress of the European Federation for Medical Informatics (MIE 2008).

The four papers selected after a peer review process, cover the major topics targeted by this special issue:

- knowledge representation and management technologies that enable the creation of machine-processable descriptions of health-related web resources, either by experts or end-users;

- knowledge and linguistic resources, along with content collection and information extraction techniques, which facilitate the indexing and retrieval of web resources;
- applications of health-related web content certification, utilizing Semantic Web technologies/tools;
- tools/technologies facilitating end-users accessing, and even rating, the descriptions of health-related web resources.

The special issue starts with the paper by Mayer et al.,[10] which presents an approach for describing health-related web resources by experts or end-users and accessing these descriptions through web browsers and search engines. This approach was implemented in the context of the European projects MedIEQ[11] and Quatro Plus.[12] The paper promotes the use of machine-processable labels adopting the recently announced W3C POWDER (PrOtocol for Web DEscription Resources) recommendation (W3C-POWDER).[13] The AQUA system developed in MedIEQ is a software platform that supports the work of labelling experts by facilitating the creation and monitoring of labels using manual and automatic approaches. Content collection and information extraction techniques are exploited, along with several linguistic and knowledge resources for the seven languages covered. Machine-readable labels can then be exploited by web browsers and search engines, using the client applications developed in the Quatro Plus project. Furthermore, end-users are provided with tools for rating official labels, as well as for creating their own labels in the context of the so-called Quality Social Networks. These end-user ratings and labels can also be viewed in a uniform way by the same client applications.

The article by Gaudinat et al.[14] presents an operational model of medical web content certification established by the Health On the Net (HON) Foundation.[15] This model exploits the RDF language for the representation of the machine-processable descriptions of medical web resources. A manual certification approach is used together with an automatic detection approach for some of the certification criteria. The article also presents an approach for the automatic generation of privacy policies; that is, how a specific website treats confidentiality of data related to individuals, adopting the Platform for Privacy Preferences (P3P) standard.[16] The implemented approach is used within a search engine. Although the authors report an average precision of 76% for the automatic part, they stress that user evaluation remains to be done, and also that the automatic approach can only complement the work of experts in assessing the quality of web resources. The authors also point out the need to involve end-users in the rating of web resources, which provides an additional source of information towards a trustworthy health-related web.

The other two articles in this special issue (Vagelatos et al.,[17] Maragoudakis and Maglogiannis[18]) deal with the issue of developing the knowledge and linguistic resources that are necessary for the indexing, labelling and retrieval of health-related web resources. Whereas the work by Vagelatos et al. exploits textual corpora for the creation of a biomedical ontology, Maragoudakis and Maglogiannis process web image data for the creation of a dermatological ontology.[17, 18] Both approaches are necessary in order to index health-related multimedia web content.

Vagelatos et al. present the design and implementation of a collection of tools and resources in the context of the Greek research project IATROLEXI.[17, 19] This led to the creation of the first biomedical ontology for the Greek language, adopting the Unified Medical Language System® (UMLS) Semantic Network (SN) as an initial top-level ontology, mapping it onto Greek and enriching it with Greek biomedical terms, processing existing dictionaries and glossaries, as well as extracting additional terms from a large corpus of biomedical texts. The final version of the IATROLEXI ontology involves 14,000 concepts represented by 16,500 terms. Although not as large as other vocabularies of major European languages, the authors consider this a significant

resource as it enriches Medical Subject Headings® (MeSH®) with terms found in Greek dictionaries and corpora. As the authors note, there are many possible applications of the developed infrastructure in the biomedical field, such as text mining, information integration and indexing of medical texts. It is good to note, for the purposes of this issue, that IATROLEXI ontology has already been exploited by the MedIEQ project for content collection and information extraction from Greek Web resources.[10]

The paper by Maragoudakis and Maglogiannis describes the process of creating a medical ontology (a dermatological ontology on skin lesions) but, instead of using textual content, it exploits web image data.[18] The ontology created is based on features of skin lesion images and is enhanced with new features from relevant images found on the web. The aim is to use the enhanced ontology for better classification and retrieval of skin lesion-related web resources. The ontology was implemented in the Web Ontology language, OWL, in order to establish a uniform syntax that can be accompanied by a strong inference engine. The authors present experiments on two image data sets: the first collected from two hospitals; and the second by two image search engines. As the authors note, the results show the ability of their approach to generate new ontologies from web image data.

To conclude, Semantic Web applications hold great potential and offer tremendous opportunities that should be developed further in the near future. The next generation of the web will exploit tools and resources such as the ones described in this special issue, thereby enabling the description of medical web resources with metadata that facilitate their retrieval, as well as with content labels by experts, and with opinions and comments by end-users.

Acknowledgments

We would like to thank the Editor of *Health Informatics Journal*, Rob Procter, for giving us the opportunity to prepare this special issue and for his assistance with its preparation. We would also like to thank the authors of all the papers submitted, and the following reviewers that helped us significantly with this initiative: Enrique Amigo, Universidad Nacional de Education a Distancia, Spain; Hercules Dalianis, Royal Institute of Technology and Stockholm University, Sweden; Irini Fundulaki, FORTH, Greece; Heng Huang, University of Texas at Arlington, USA; Dimitrios Kokkinakis, Goteborg University, Sweden; Stasinos Konstantopoulos, NCSR "DEMOKRITOS", Greece; Anastasia Krithara, NCSR "DEMOKRITOS", Greece; Fillia Makedon, University of Texas at Arlington, USA; Georgios Petasis, NCSR "DEMOKRITOS", Greece; Matti Polla, Helsinki University of Technology, Finland; Georgios Vouros, University of the Aegean, Greece; Pierre Zweigenbaum, LIMSI - CNRS, France;

Vangelis Karkaletsis
National Centre for Scientific Research "DEMOKRITOS" (NCSR), Greece
Miguel A. Mayer
Research Programme on Biomedical Informatics (GRIB), IMIM-Universitat Pompeu Fabra, Spain
Pythagoras Karampiperis
National Centre for Scientific Research "DEMOKRITOS" (NCSR), Greece

References

1. Soualmia LF, Darmoni SJ, Douyère M and Thirion B. Modelisation of Consumer Health Information in a Quality-Controled gateway. In: Baud R et al. (ed.). *The New Navigators: from Professionals to Patients*. Proc of MIE2003, 2003, pp. 701–706.
2. Berners-Lee T, Hendler J and Lassilla O. The semantic web. *Sci Am* 2001; 284: 34–43.
3. http://www.w3.org/TR/rdf-schema
4. http://www.w3.org/TR/owl-features

5. Nabarette H, Romaneix F, Boyer C, Darmoni SJ, Rémy PL and Caniard E. Certification des sites dédiés à la santé en France, Presse Med, August 2009.
6. Wilson P. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the Internet. *BMJ* 2002; 321: 598–602.
7. Rice RE and Katz JE. *The Internet and Health communication. Experiences and Expectations*. London: Sage Publications, Inc. 2001.
8. Karkaletsis V, Stamatakis K, Karampiperis P, Labský M, Ruzicka M, Svátek V, et al. Management of Medical Website Quality Labels via Web Mining. In: Berka P, Rauch J, Abdelkader Zighed D (eds) *Data Mining and Medical Knowledge Management: Cases and Applications*. IGI Global Inc., 2009, pp. 206–226.
9. http://www.medieq.org/drmed2008.
10. Mayer MA, Karkaletsis V, Villaroel D, Leis A, Karampiperis P, Kukurikos A, et al. Applying semantic web technologies to improve the retrieval, credibility and use of health websites. *Health Informatics Journal* 2011; 17 (2): 95–115.
11. http://www.medieq.org
12. http://www.quatro-project.org
13. http://www.w3.org/2007/powder
14. Gaudinat A, Cruchet S, Boyer C and Chrawdhry P. Enriching the trustworthiness of health-related web pages. *Health Informatics Journal* 2011; 17 (2): 116–126.
15. http://www.hon.ch
16. http://www.w3.org/P3P
17. Vagelatos A, Mantzari E, Pantazara M, Tsalidis Ch and Kalamara Ch. Developing tools and resources for the biomedical domain of the Greek language. *Health Informatics Journal* 2011; 17 (2): 127–139.
18. Maragoudakis M and Maglogiannis I. A medical ontology for intelligent web-based skin lesions image retrieval. *Health Informatics Journal* 2011; 17 (2): 140–158.
19. http://www.iatrolexi.gr