

The Case of MedIEQ

V. Karkaletsis^a, P. Karampiperis^a

^a*Software & Knowledge Engineering Lab, Institute of Informatics and Telecommunications,
National Center for Scientific Research (NCSR) "Demokritos",
P. Grigoriou & Neapoleos, 15310 Aghia Paraskevi Attikis, Greece
{vangelis, pythk}@iit.demokritos.gr*

Extended Abstract

The number of health information websites and online services is increasing day by day. We can find websites published by government institutions, consumer and scientific organizations, patients associations, personal sites, health provider institutions, commercial sites, etc. On the other hand, patients continue to find new ways of reaching health information and more than four out of ten health information seekers say the material they find affects their health decisions (Eysenbach, 2000; Diaz et.al., 2002). However, it is difficult for health information consumers, such as the patients and the general public, to assess by themselves the quality of the information because they are not always familiar with the medical domains and vocabularies (Soualmia et.al., 2003).

Although there are different opinions about the need for certification of health websites and adoption by Internet users, different organizations around the world are working on establishing standards of quality in the certification of health-related web content (Winker et.al., 2000; Kohler et.al., 2002; Curro et.al., 2004; Mayer et.al., 2005). The European Council supported an initiative within eEurope 2002 to develop a core set of "Quality Criteria for Health Related Websites" (EC, 2002). The specific aim was to specify a commonly agreed set of simple quality criteria on which Member States, as well as public and private bodies, may build upon for developing mechanisms to help improving the quality of the content provided by health-related websites. These criteria should be applied in addition to relevant Community law. As a result, a core set of quality criteria was established. These criteria may be used as a basis in the development of user guides, voluntary codes of conduct, trust marks, certification systems, or any other initiative adopted by relevant parties, at European, national, regional or organizational level.

Currently, two major approaches exist concerning the labeling of health information in the internet: a) *filtering portals* (organizing resources in health topics and providing opinions from specialists on their content) and b) *third-party certification* (issuing certification trustmarks or seals once the content conforms to certain principles). In general, and in both approaches, the labeling process comprises three tasks that are carried out entirely or partially by most labeling agencies:

- *Identification* of new web resources: this could happen either by active web searching or on the request of the information provider, i.e. the website responsible actively asks for the review in order to get a certification seal.
- *Labeling* of the web resources: this could be done with the purpose of awarding a certification seal or in order to classify and index the web resources in a filtering portal.
- *Re-reviewing* or *monitoring* the labeled web resources: this step is necessary to identify changes or updates in the resources as well as broken links, and to verify if a resource still deserves to be awarded the certification seal.

This is the general case; eventually, any particular agency can integrate additional steps which may be necessary in its work. The Agency for Quality in Medicine – AQuMed (<http://www.aeqq.de>) and Web Mèdica Acreditada - WMA (<http://wma.comb.es>), are representative examples of the two approaches mentioned above: AQuMed maintains a filtering portal while WMA acts as a third-party certification agency.

Medical quality labeling mechanisms, applied in either approach, could be improved if equipped with semantic web technologies that enable the creation of machine-processable labels as well as the automation of the labeling process. Among the key ingredients for the latter are *web crawling* techniques that allow for retrieval of new unlabelled web resources, or *web spidering and extraction* techniques that facilitate the characterization of retrieved resources and the continuous monitoring of labeled resources alerting the labeling agency in case some changes occur against the labeling criteria.

MedIEQ "Quality Labelling of Medical Web content using Multilingual Information Extraction" (<http://www.medieq.org>) is a technology project employing semantic web technologies for the description of web resources, and content analysis technologies for collecting domain-specific web resources and extracting information from them. These technologies can also be applied in other domains. What we aim in this project is to examine their applicability in a specific area with special importance, that of assisting the labelling of health related web resources. MedIEQ involves labeling experts from WMA from Spain and AQuMed from Germany, and cooperates, via its advisory committee, with experts from Health on the Net Foundation (HON), Catalog and Index of French-language Health Internet resources (CISMeF), the Family Online Safety Institute (FOSI).

In the context of MedIEQ, we are developing machine readable domain specific labels, based on the RDF-CL model proposed by the EC-funded QUATRO project. The vocabulary used in these labels is based on the criteria that are currently used by the participating labelling agencies WMA and AQuMed, the eEurope criteria guidelines and a labelling agency of reference as HON. Our aim is not to suggest this vocabulary as "the only one to use" but to show instead the value of machine readable labels. So, *the emphasis is on the technology and not on the terms (criteria) included in the vocabulary*. The terms selected capture important aspects of health related content and form the case study for MedIEQ technology partners.

In the MedIEQ project we decided to develop a representative collection of *labeling criteria*, which would reflect the needs of the *labeling agencies* involved in the project consortium and at the same time provide an adequate proof of concept for our general methodology for computer-assisted labeling. It is important to stress that the methodology and software tools are to a large degree independent of the concrete criteria and thus could be easily adapted to different criteria used by various agencies. Such adaptation is also eased by the fact that the criteria specification was also influenced by the analysis of criteria used by other organizations such as HON, and thus has significant overlap with them.

To make content labels machine-readable the use of the RDF model is adopted. At the current stage, the RDF-CL model is used. The RDF-CL model was issued by the EC-funded project Quality Assistance and Content Description (QUATRO) (<http://www.quatro-project.org>); it is currently being refined by the W3C Protocol for Web Description Resources (POWDER) working group (W3C, 2007). POWDER is expected to be completed before the end of the MedIEQ project and the plan is to use it in the final version of the MedIEQ prototype.

References

- Curro V, Buonomo PS, Onesimo R, de RP, Vituzzi A, di Tanna GL, D'Atri A. A quality evaluation methodology of health web-pages for non-professionals. *Med Inform Internet Med* 29(2) (2004), 95-107.
- Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical Information. *J Gen Intern Med* 17(3) (2002), 180-5.
- EC, European Commission. eEurope 2002: Quality Criteria for Health related Websites, 2002. Available Online at: http://europa.eu.int/information_society/eeurope/ehealth/doc/communication_acte_en_fin.pdf.
- Eysenbach G. Consumer health informatics. *BMJ* 320 (4) (2000), 1713-16.
- HON, Health on the Net Foundation, Analysis of 9th HON Survey of Health and Medical Internet Users Winter 2004-2005, 2005. Available Online at: <http://www.hon.ch/Survey/Survey2005/res.html>
- Kohler C, Darmoni SD, Mayer MA, Roth-Berghofer T, Fiene M, Eysenbach G. MedCIRCLE – The Collaboration for Internet Rating, Certification, Labelling, and Evaluation of Health Information. Technology and Health Care, Special Issue: Quality e-Health. *Technol Health Care* 10(6) (2002), 515.
- Mayer MA, Leis A, Sarrias R, Ruíz P. Web Médica Acreditada Guidelines: reliability and quality of health information on Spanish-Language websites. In: Engelbrecht R et al. (ed.). *Connecting Medical Informatics and Bioinformatics*. Proc of MIE2005 (2005), 1287-92.
- Soualmia LF, Darmoni SJ, Douyère M, Thirion B. Modelisation of Consumer Health Information in a Quality-Controlled gateway. In: Baud R et al. (ed.). *The New Navigators: from Professionals to Patients*. Proc of MIE2003 (2003), 701-706.
- W3C, Protocol for Web Description Resources (POWDER), 2007. Available Online at: <http://www.w3.org/2007/powder/>
- Winker MA, Flanagan A, Chi-Lum B. Guidelines for Medical and Health Information Sites on the Internet: principles governing AMA websites. *American Medical Association. JAMA* 283 (12) (2000), 1600-1606.