

The case of MedIEQ



Vangelis Karkaletsis

DrMED@MIE-2008,
Goteborg, 27 May 2008



Institute of Informatics and Telecommunications – NCSR “Demokritos”

 Health and Consumer Protection
Directorate - General



Some facts about health related web content



- § The number of health web sites and on-line services is increasing day by day
- § 70-80% of Internet users seeks health information for them or for their relatives
- § More than 4 out of 10 health information seekers say the material they find affect their health decisions



Existing labeling processes - I



- § Organisations around the world are working on establishing quality standards for health-related web resources
- § Quality standards initiatives are not enough
 - Self-adherence to codes of conduct or ethics, a claim with little enforceability
- § Necessary the establishment of labeling mechanisms
 - by third party accreditation
 - by creating portals where web sites are organized and characterized against certain labeling criteria
- § Such initiatives already exist ...



Existing labeling processes – II



§ Problems due to

- High costs to offer the service
- Huge amount of information to assess
- Content changes rapidly
- Broken links to accredited web resources
- Dishonest use of the label

§ Most of the work in labeling processes is currently performed manually

§ Access of the end-users to labeled resources must be improved

§ Enable end-users to express opinions on labels or even label web resources



Need for new labeling technologies



- § Enable the automation of the labeling authorities' work; this involves
 - Technology for creating machine processable labels
 - ü Establishing common schemas and vocabularies exploiting semantic web technologies (RDF, OWL)
 - ü Developing label generators with user-friendly interfaces
 - Technology for maintaining the labels
 - ü Monitoring the label's content against the content of the labeled resource using content analysis technologies
 - Technology for locating unlabeled web resources



MedIEQ in Brief - I



- § MedIEQ “Quality Labeling of Medical Web content using Multilingual Information Extraction”
 - EC-funded project, DG SANCO “Health & Consumer protection”, Directorate C “Public Health and Risk Assessment”
- § MedIEQ is a technology project employing semantic web technologies for the description of web resources, content analysis technologies for collecting domain-specific web resources and extracting information from them.
- § What we aim is
 - to examine their applicability in a specific area with special importance, that of assisting the labelling of health related web resources,
 - to provide a set of tools to support the work of experts in labelling authorities (LAs).



§ For this purpose

- we develop machine readable domain specific labels
 - ü based on the RDF-CL model proposed by the QUATRO project and at a later stage the model to be proposed by the relevant W3C POWDER working group
- we employ techniques alleviating the mundane tasks of the labeling process
 - ü helping in the identification of unlabeled resources, extracting from these resources information relative to specific labeling criteria, facilitating the monitoring of already labeled resources
- we created a vocabulary re-using existing terms, important for health-related content labels
 - ü the goal is not to suggest this vocabulary as the one to use but to highlight the value of machine readable labels.



Achievements so far - I



§ 1st version of the MedIEQ Integrated Prototype (AQUA)

- a S/W platform for supporting labeling experts
- implements an open architecture facilitating the addition of new components
- focus on the provided infrastructure and the customisation to new languages and labeling criteria

§ Final version of the MedIEQ labelling schema

- covers the full set of the labelling criteria proposed implementing the adopted RDF-CL model



Achievements so far - II



- § **Final version of the Web Content Collection toolkit**
 - provides tools for web crawling and for spidering of web sites
- § **1st version of the Information Extraction toolkit**
- § **Final version of the Multilingual Resources Management toolkit**
 - provides tools for accessing, editing, and exploiting UMLS resources in the project languages
- § **Evaluation of the 1st version of AQUA was conducted (in English and Spanish)**



Next steps



∅ Final version of AQUA:

- ü integrate the final versions of all sub-systems,
- ü support additional labeling criteria,
- ü cover seven European languages,
- ü support POWDER labels

∅ Evaluation of the 2nd version of AQUA

- ü scope of the first evaluation was the performance evaluation on supporting the labeling process (i.e. identification of new unlabeled web resources and labeling of web resources), as well as, the usability evaluation of AQUA interface
- ü the evaluation of the 2nd version of AQUA focuses on the labeling performance on a real integration scenario in the usual work of a labeling organization.