

Χαρακτηρισμός ιστοχώρων με περιεχόμενο σχετικό με θέματα υγείας: τρέχουσες προσεγγίσεις και το έργο MedIEQ



Βαγγέλης Καρκαλέτσης

Ημερίδα «Γλωσσική Τεχνολογία στο Βιοϊατρικό χώρο»

Αθήνα, 10 Δεκεμβρίου 2007



Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών «ΔΗΜΟΚΡΙΤΟΣ»
Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών



Περιεχόμενα

- Ετικέτες χαρακτηρισμού περιεχομένου
- Διαδικασίες χαρακτηρισμού ιατρικού περιεχομένου
- Ανάγκη για νέες τεχνολογίες
- Έργα και πρωτοβουλίες του Εργαστηρίου
 - Το έργο MedIEQ
- Συμπεράσματα



Quality labels / trustmark schemes - I

- Quality labels / trustmark schemes have been established in many parts of the world
 - some are online versions of existing schemes,
 - others have been developed specifically for the web.





Quality labels / trustmark schemes - II

- Inform the user about the quality of data and services provided
 - for example, a label may include an assertion that the labeled web site has a suitable privacy policy, that the publisher is clearly identified, and that it meets legal practice in one or more identified countries.
- Two notable areas of interest for quality labels / trustmarks schemes are
 - those that indicate that health related content has been peer reviewed
 - those designed to give consumers confidence in eCommerce operations, and



Quality labels / trustmark schemes – III

the WMA label for health related web content

- Identity
- Content
- Confidence
- Advertising
- Virtue
- Non-maleficence

Criteria Name	Values	Description	Question
A. Identification			
1.1. Mail address	Present/not present/not checked	Mail of contact in the web easily visible, in the homepage for example	Is the postal address present?
1.1. Valid mail address	Not checked Valid Invalid	checking the answer given by the health information provider	Is the postal address a valid address?
2. Person in charge mail address	Compulsory in the request form	Mail address in the request form	
3. Person in charge	Present/not present/not checked	Person in charge is identified	Is the person or team in charge identified? ¹
4. Health care professional	Present/not present/not checked	The health care professional is identified with name, speciality and position	Is there all the information about the health care professional? ¹
5. Category of web²	Compulsory in the request form	Type of the web in the request form	Type of the web fill in the request form
6. Name of the web	Compulsory in the request form (free text)	Which name is called the web (provider information)	Fill in the name of your web
7. Description of the web	Compulsory in the request form (free text)		Fill in the description of your web
8. Another Seal of quality or third-party program	Present/not present/not checked	Is the site a subscriber of any other seal quality or third-party program?	



Quality labels / trustmark schemes – IV

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



Archivo Edición Ver Favoritos Herramientas Ayuda

Sign in



Web Images Groups News Froogle Maps more »

vaccine OR vaccination

Search

Advanced Search
Preferences

Web

Results 1 - 10 of about 108,000,000 for **vaccine OR vaccination**. (0.39 seconds)

www.sciencedirect.com/science/journal/0264410X

[Similar pages](#)

Vaccine Page, The

The **Vaccine** Page is a serious attempt to collect all useful information about vaccines, organize it meaningfully, present it appropriately for each of ...

vaccines.org/ - 24k - [Cached](#) - [Similar pages](#)

Vaccine

VACCINE is the pre-eminent journal for those interested in vaccines and **vaccination**. It serves as an interface between academics, those in research and ...

www.elsevier.com/locate/vaccine - 67k - 6 May 2006 - [Cached](#) - [Similar pages](#)

National Vaccine Information Center

Has my child had a bad reaction to a **vaccination** before? Does my child have a personal or family history of:.. **vaccine** reactions; convulsions or neurological ...

www.909shot.com/ - 59k - 6 May 2006 - [Cached](#) - [Similar pages](#)

Vaccine Website

THE **VACCINATION** HOAX · The Vaccines · MMR/MR · Contaminants · SV40 · Ingredients · Aluminium · Mercury · Package Inserts · Schedule · **Vaccine** Exemption, USA ...
www.whale.to/vaccines.html - 17k - 6 May 2006 - [Cached](#) - [Similar pages](#)

National Immunization Program (NIP)

Recommended **Vaccination**/Immunization Schedules | News & Events | Frequently Asked For Items | Diseases | Vaccines | Resources | **Vaccine** Safety | Why ...
www.cdc.gov/NIP/ - 71k - [Cached](#) - [Similar pages](#)

Vaccination News

Your source for daily news, reporting all sides of the **vaccination** controversy.
www.vaccinationnews.com/ - 2k - [Cached](#) - [Similar pages](#)

Vaccine - Wikipedia, the free encyclopedia



Internet

Inicio



11:52



Archivo Edición Ver Favoritos Herramientas Ayuda



Dirección <http://www.909shot.com/>

Ir BullsEye

HOME ABOUT US REPORT A PROBLEM GET INFORMED E-NEWS MEDIA SIGN-IN RESOURCE CENTER MEMBERSHIP DONATION LAWYER REFERRAL CONTACT US



STAY INFORMED
FREE E-NEWS

MAKE A DIFFERENCE
BECOME A MEMBER

VACCINE NEWS OF THE WEEK

HEALTH TALK CALENDAR

ISSUES

VACCINATION DECISIONS

FREEDOM IN VACCINATION & INFORMED CONSENT

STATE INFORMATION

TRACKING SYSTEMS & PRIVACY

HOMELAND SECURITY

VACCINE ADVERSE EVENTS REPORTING

WELCOME !

Vaccines or immunizations are recommended for every child born in the United States. Vaccinations shouldn't hurt a child but sometimes they do. Before your child takes the risk, find out what it is.



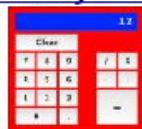
Click on photo to read these children's stories

The National Vaccine Information Center (NVIC) is a national, non-profit educational organization founded in 1982. It is the oldest and largest national organization advocating reformation of the mass vaccination system.

(More about NVIC)

CURRENT PROJECTS

The Mercury Calculator



How much Mercury has your child received



Protesting Forced Anthrax Vaccination

ACTION ALERT & ADVOCACY

NVIC CONFERENCES

BARBARA LOE FISHER SPEAKS OUT

ARTICLES AND INTERVIEWS

MOTHERING MAGAZINE
In the Wake of Vaccines Sept/Oct 2004

THE BRIAN LEHRER SHOW
Public Health vs Parents' Fears 10/9/03

RFDFA.GSDAILY.COM

Inicio



Bandeja de ...

Microsoft Po...

http://www....

National Vac...

ES



8:57



Existing labeling processes - I

- Organisations around the world are working on establishing quality standards for health-related web resources
 - European Commission
 - ✓ eEurope 2002: *Quality criteria for health related web sites*
 - American Medical Association
 - ✓ *Guidelines for medical health information sites on the Internet*
 - Internet Healthcare Coalition
 - ✓ eHealth Code of Ethics
 -



Existing labeling processes - II

- Quality standards initiatives are not enough
 - Self-adherence to codes of conduct or ethics, nothing more than 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 - III

Codes of Conduct: are defined as sets of quality criteria that provide a list of recommendations for the development and content of websites

Codes for health related content

Quality Label (logo): is displayed on screen and represents a commitment by a provider to implement or adhere to a code of conduct

Internet Health Coalition
American Medical Association (AMA)
e-Europe 2002

Quality Label (logo)

User Guidance: enables users to check if a site complies with certain standards by accessing a series of questions from a displayed logo

Internet Medical Association Mark (JIMA Mark)

DISCERN
NetScoring

Filtering Tools: applied manually or automatically, accept or reject web resources
- resources are selected for their quality and relevance to a particular audience

AQUMED, OMNI
CISMeF

Third Parties certification: quality and accreditation labels are awarded by a third party to inform consumers that a site provides information meeting current standards for content and form

QUATRO
Health on the Net Code
MedCIRCLE
Web Mèdica Acreditada (WMA)
URAC



Existing labeling processes – IV

- Problems due to
 - High costs to offer the service
 - Huge amount of information to assess (too many web resources)
 - Content changes rapidly
 - Broken links to accredited web resources
 - Non standardised rating criteria
 - Dishonest use of the label



Need for new labeling technologies – I

- Most of the work in labeling processes is currently performed manually
 - A site may have hundreds of pages (static/dynamic ones) or other resources (.doc, .rtf, .pdf, images, ...)
 - Probably all or most of the resources have to be checked
 - A single label may be used for the whole site or different labels may be used for different parts of the site



Need for new labeling technologies – II

- Access of the end-users to labeled resources must be improved
 - If labels could be recognised by web browsers and search engines this would motivate content providers to label their resources
- Enable end-users to express opinions on labels or even label web resources
 - User-generated labels and opinions expressed about existing labels, without attribution, to be provided into an open source of data that anyone can access, through social networking



Need for new labeling technologies – III

- Enable the automation of the labeling operators' work; this involves
 - Technology for creating machine processable labels
 - Technology for maintaining the labels
 - Technology for locating unlabeled web resources



Need for new labeling technologies – IV

- Enable the access of the end-user to the label and its content; this involves
 - Technology for locating the label inside a web resource, reading and validating label's content
 - Technology for presenting the label's content and validation results to the end user
- Enable user generated labels and expressing opinions on labels; this requires
 - User-friendly label and opinion generators
 - Social networking functionalities



Technology for creating machine processable labels

- Establishing common schemas and vocabularies exploiting semantic web technologies (RDF, OWL)
 - RDF/XML files make simple statements:
 - ✓ Advertising is present here
 - ✓ There is a service of virtual consultation for professionals
 - ✓ The intended audience is health professionals
 - ✓ It's in the MeSH "quality of health care" category
 - Existing RDF vocabularies can be re-used
- Having the languages for representing labels' data is not enough
 - Developing label generators with user-friendly interfaces is also needed



Technology for maintaining the labels

- Monitoring the label's content
 - When a label is generated, the following data may be stored in the labeling authority's data base:
 - ✓ Issue date, expiration date, label's content hash, ...
 - The data stored about the label can be used by a tool to examine
 - ✓ the label against the expiration date (if date has passed, alert the authority)
 - ✓ the label's integrity
- Monitoring the label's content against the content of the labeled resource using content analysis technologies
 - spidering technology that enables navigating the monitored site to locate resources related to the labeling criteria
 - information extraction technology to extract from the located resources the data corresponding to the labeling criteria



Technology for locating unlabeled web resources

- Use of focused crawling technology to locate unlabeled domain specific web resources exploiting
 - existing search engines
 - machine learning techniques
 - domain specific resources



Technologies for improving accessibility to the labels

- Locating the label of a web resource, reading and validating its content
 - parse the resource's content to locate a label
 - if a label exists, identify the labeling authority, validate the label against the data in the authority's data base
- Presenting the label's content and validation results to the end user
 - enabling existing web browsers and search engines
 - ✓ to communicate with web services that are able to locate and validate labels in the retrieved resources, as well as
 - ✓ to present the label's data and validation results in a format understandable by the end-user



QUATRO project

- The EC-funded project “Quality Assurance and Content Description – QUATRO” (Safer Internet Programme) developed
 - A machine processable RDF schema for content labeling (RDF-CL)
 - QUATRO proxy server (QUAPRO)
 - A browser extension, the Metadata Visualizer (ViQ)
 - A search engine wrapper displaying annotated search results that link to the corresponding labels, the Label Display Interface (LADI)



QUATRO project: ViQ

The screenshot shows a Mozilla Firefox window titled "ViQ Test Pages: We have labels! - Mozilla Firefox". The address bar displays the URL <http://localhost/www/quattro/viq/testpages/2/>. The main content area shows a list of labels:

- ICRA (red icon)
- IQUA (yellow icon)
- WMA (green icon)

A tooltip for the IQUA label states: "This label is not valid". Below the labels, the text "rest Pages" is visible. A red box highlights the "Case 2" section, which contains the following text:

Case 2

The following labels are associated with this Web page:

- an **ICRA** label (this label is **invalid**)
- an **IQUA** label (this label is **valid**)
- a **WMA** label (the validity of this label could not be evaluated—its validity is **unknown**)



QUATRO project: LADI

WebForm1 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites PageRank ABC Check AutoLink Options

Address http://www.quatro-project.org/search.aspx Go Links

Google Search PageRank ABC Check AutoLink Options

Alexa Search Info 987,349

Amazon.com

Search

Results 1 - 10 of about 1340000 for quattro [definition].(0.648993 seconds)

Web

▼ Quattro :: Your Key Technology
For more than a decade **Quattro** has...
Quattro has focused solutions for Edu...
<http://www.quattro.com/>

Suzi Quattro.com
Biography, current projects, news, im...
<http://www.suziquattro.com/>

<http://www.quatroidinternational.com/>

▼ SUZI QUATRO
SUZI QUATRO WWW. Dont forget to...
3rd June 1950 in Detroit USA, after to...
<http://www.ozemail.com.au/~suziq/>

Quattro Four Retail Pty. Ltd
Quattro is a supply chain services network that electronically links the four major business categories that underpin the convenience retail market Retailers ...
<http://www.quattro.com.au/>

Information from QUAPRO about http://www.quattro.com/

ICRA IQUA WMA

Authority: The Internet Content Rating Association
Issued: 13/10/2005
Valid: No
Summary: No nudity, No sexual material, ...

Done Internet

Done Internet

Χαρακτηρισμός ιστοχώρων με περιεχόμενο σχετικό με θέματα υγείας

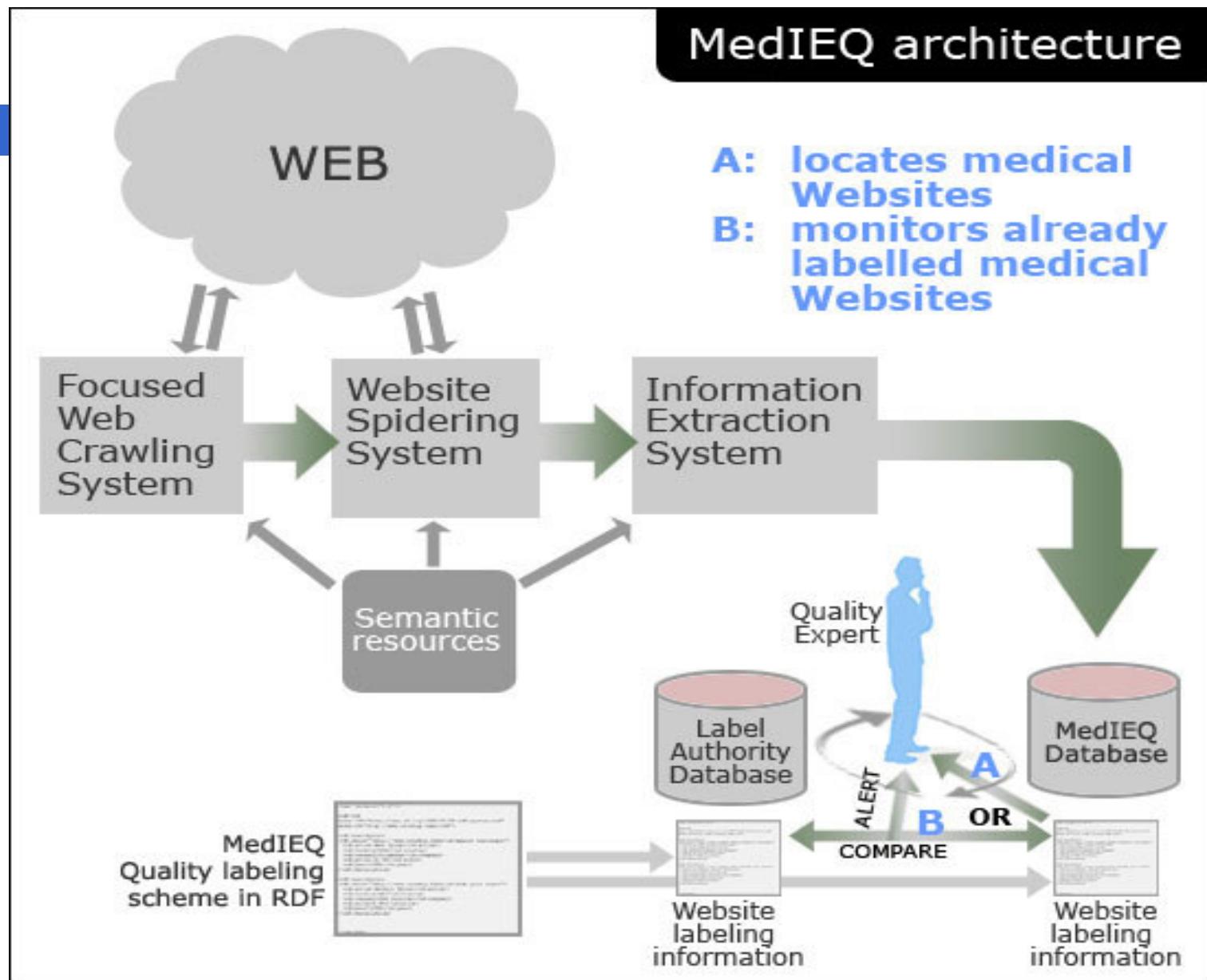
Ημερίδα έργου «ΙΑΤΡΟΛΕΞΗ», 10-12-2007



MedIEQ Project

- 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 objectives
 - develop a scheme for the quality labelling of health related web content and provide the tools supporting the creation, maintenance and access of labelling data according to this scheme;
 - specify a methodology for the content analysis of medical web sites according to the MedIEQ scheme and develop the tools that will implement it;
 - integrate these technologies into a prototype labelling system;
 - demonstrate the resulting prototype in 7 different languages (Spanish, Catalan, German, English, Greek, Czech, and Finnish) and two labelling applications (third party accreditation, classification)

MedIEQ architecture





Label Management (LAM) in AQUA

The Label Restrictions section holds a list of URLs that this label is restricted for

The “Create/Update Label” button and the “Clear Label” link

MedIEQ
Assisting Quality Assessment system

Quality Labelling of Medical Web Content using Multilingual Information Extraction

My account
+ Edit my account
+ Logout

Quality labelling
+ My web resources
+ Task management
+ Search
+ Review/Monitor
+ Form a corpus
+ The quality criteria

The AQUA system
+ About AQUA
+ Contact the system administrators

Create new Label

Label Metadata

*Organization Name:
The labelling organization name

*Mailbox address:
e-mail format : a valid email in the form user@host.com

*Homepage:
Homepage format : a valid URL in the form http://www.website.com

*Issued date: 08/23/2007
Date format : MM/DD/YYYY

*Authority for:
A valid URI for the namespace of the RDF schema that the authority is responsible for

Label Restrictions

Host Edit Host Restriction
1 http://www.ahd.com
[Add New URI] [Remove Selected URI]
[Remove All URIs]

Labelling Criteria

Property	Value	URL found
Advertisement Present	<input type="text"/>	<input type="text"/> Proposed Values*
Target Audience	<input type="text"/>	<input type="text"/> Proposed Values*
Virtual Consultation	<input type="text"/>	<input type="text"/> Proposed Values*
Other Seal	<input type="text"/>	<input type="text"/> Proposed Values*
Resource Language	<input type="text"/>	<input type="text"/> Proposed Values*
Last Update	<input type="text"/>	<input type="text"/>

Save Label

[Create Label] [Clear Label]

The Label Metadata section contains information about the organization that issues the label

The Labelling Criteria section contains a form with the labelling properties



MedIEQ Labeling criteria

Resource Defining Information

- 1.1 Resource URI
- 1.2 Resource title
- 1.3 Resource last update
- 1.4 Resource language(s)

Ownership / Creatorship

- 2.1 Organization name(s) (owner)
- 2.2 Organization type(s) (owner)
- 2.3 Responsible name(s)
- 2.4 Responsible title(s)
- 2.5 Responsible(s) contact details
- 2.6 Webmaster name(s)
- 2.7 Webmaster(s) contact details

Purpose / Mission

- 3.1 Purpose / mission of the resource provided
- 3.2 Purpose / mission of the owner(s) provided
- 3.3 Target / intended audience(s)
- 3.4 Statement declaring limitation of the provided information

Topics / Keywords

- 4.1 Topics / Keywords (UMLS)

Virtual Consultation (VC)

- 5.1 VC service available
- 5.2 VC responsible name(s)
- 5.3 VC responsible(s) contact details
- 5.4 Statement declaring limitation of the VC service

Funding / Advertising

- 6.1 Statement declaring sources of funding (sponsors, grants, advertisers, etc.)
- 6.2 Name(s) of funding (sponsoring) organization(s)
- 6.3 Statement declaring limitation of influence of sponsors on content
- 6.4 Advertising present
- 6.5 Are advertisements clearly separated from editorial content?
- 6.6 Policy with regard to advertisement

Other Seal or Recommendation

- 7.1 Other seal(s) present
- 7.2 Which other seal(s)?

Information Supporting Scientific Content

- 8.1 References, bibliography (with links to literature)
- 8.2 Publication / creation date
- 8.3 Last revision / modification date
- 8.4 Author name(s)
- 8.5 Author(s) contact details
- 8.6 Editorial policy

Confidentiality / Privacy Policy

- 9.1 Explanation on how personal data (visitor coordinates, e-mail messages, etc.) is handled

Accessibility

- 10.1 Accessibility level



QUATRO Plus

- EC-funded under the Safer Internet Plus Programme
- Objectives
 - Prove the benefits and potential of machine-readable labels to a variety of commercial interests across Europe.
 - Allow end users to record and share their agreement or disagreement with a label's description.
 - Allow end users to create their own labels to share directly with their friends and associates.
 - Aggregate all available user-generated labels and opinions expressed about existing labels, without attribution, into an open source of data that anyone can access.
 - Provide simple to use tools for label creation.
 - Provide a flexible and scalable infrastructure that makes labels readily accessible and that authenticates the data.



Access Q+ Label Management Environment

The Family Online Safety Institute: Protecting your online future - Windows Internet Explorer

http://www.fosi.org/

Create new Label

Label Metadata

*Organization Name:

The labelling organization name

*Mailbox address:

e-mail format : a valid email in the form user@host.com

*Homepage:

Homepage format : a valid URL in the form http://www.website.com

*Issued date: 08/23/2007

Date format : MM/DD/YYYY

*Authority for:

A valid URI for the namespace of the RDF schema that the authority is responsible for

Label Restrictions

Host Edit Host Restriction

1 http://www.ahd.com

Add New URI Remove Selected URI Remove All URIs

recommended tags

del.icio.us

url <http://www.ahd.com>

description The Family

notes

tags

save

popular tags

security online safety

Done

Powered by KCOM group

Internet 100%



QUATRO, MedIEQ, QUATRO Plus, WCL and POWDER

- QUATRO answered the call to create a pan-European system of interoperable quality labels
 - FOSI, IQUA and COMB publish their labels in a common machine-readable format
 - Users see the labels and the result of the authentication request in their browser extension or directly on their search results.
- MedIEQ (DG-SANCO, eHealth) supports the automation of quality labeling process, by labeling experts, in health related web content
- QUATRO Plus will allow end users to record and share their agreement or disagreement with a label's description.
- W3C Content Label Incubator Activity (WCL) allowed QUATRO labeling platform to come under W3C spotlight
- WCL led to the POWDER Working Group (Protocol for Web Description Resources) that will result to a new W3C standard for content labels



Concluding ...

- Establishment of content labels in practice cannot be enforced by measures
- If health-related content providers realize that
 - content labels and expressions of users opinions can be added easily to their content
 - labeling authorities and experts are equipped with technology that facilitates the monitoring of the provided web content against the labeling criteria
 - end-users will be able to express opinions on labels or create their own labels
 - search engines and browsers can inform users on the existence of content labels and their features
- they will adopt machine readable content labeling technology
 - leading to the increase of labeled sites
 - improving in turn the quality health knowledge disseminated through the Web



Concluding ...

- Labeling of web content is an application area for
 - content analysis
 - ✓ crawling, spidering, information extraction
 - knowledge management
 - ✓ ontologies, semantic web
 - intelligent interfaces
 - ✓ browsers, search engines, social networks



Useful Links

QUATRO site

<http://www.quattro-project.org>

MedIEQ site

<http://www.medieq.org>

POWDER-WG

<http://www.w3.org/2007/powder/>

Χαρακτηρισμός ιστοχώρων με περιεχόμενο σχετικό με θέματα υγείας: τρέχουσες προσεγγίσεις και το έργο MedIEQ



Βαγγέλης Καρκαλέτσης

Ημερίδα «Γλωσσική Τεχνολογία στο Βιοϊατρικό χώρο»

Αθήνα, 10 Δεκεμβρίου 2007



Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών «ΔΗΜΟΚΡΙΤΟΣ»
Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών