The MedIEQ AQUA system
User Manual

August 2008

Table of Contents

List of Figures

1. Introduction

The number of health information web sites and online services is increasing day by day. It is known that the quality of these websites is very variable and difficult to assess; we can find web sites published by government institutions, consumer and scientific organizations, patients associations, personal sites, health provider institutions, commercial sites, etc. [1]. 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 decisions about their health [2, 3]. 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 [4].

Although there are divergent opinions about the need for accreditation of health web sites and adoption by Internet users [5], different organizations around the world are working on establishing standards of quality in the accreditation of health-related web content [6, 7].

By analyzing the main approaches of medical quality labeling, we have identified the following key tasks, followed entirely or partially by most labeling agencies:

- *Identification of new web resources*: this could happen either by active web searching or by voluntary application from the information provider, i.e. the web site responsible asks actively for a review, usually in order to get an accreditation seal.
- *Labeling of the web resources*: this could be done with the purpose of awarding an accreditation 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 an accreditation seal.

As a result, the AQUA system was designed to support the main tasks of the web content accreditation process, that is:

- Identification of unlabeled resources having health-related content;
- Visit and review of the identified resources;
- Generation of content labels for the reviewed resources;
- Monitoring the labeled resources.

AQUA aims to provide the infrastructure and the means to organize and support various aspects of the daily work of labeling experts by making them computer-assisted. More specifically, AQUA supports labeling experts in:

- *Creating machine readable labels,* by adopting the use of the RDF model [8] for producing machine-readable content labels; at the current stage, the RDF-CL model [9] is used.
- *Automating the accreditation process* by helping in the identification of unlabeled resources, extracting from these resources information relative to specific accreditation criteria, generating content labels from the extracted information and facilitating the monitoring of already labeled resources.

This document provides a step-by-step how-to manual for the basic functionalities supported by the first version of the AQUA system.

2. AQUA User Manual

2.1. Login to AQUA

Step 1: From a browser go to: "http://www.medieq.org/aqua/welcome.seam". This web link directs you to the AQUA login page (see Figure 1)

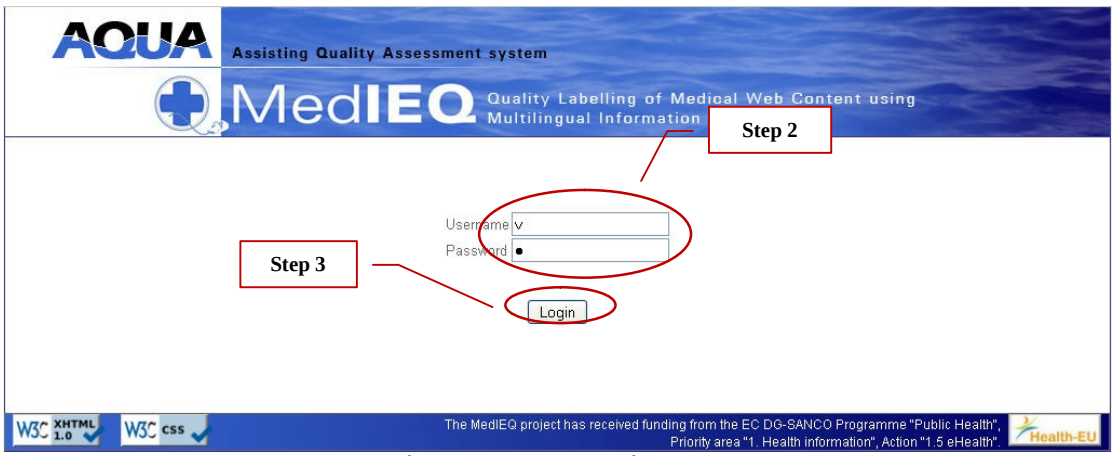

Figure 1. AQUA Login Page

Step 2: Fill your username and password

Step 3: Press the "Login" button. Provided that you have used a correct username and password, you will be directed to AQUA's welcome page (see Figure 2).

Figure 2. AQUA's Welcome Page

2.2. Changing Login Account Information

Step 1: At any time you can modify your account information by following the link "Edit my account" from the "My account" submenu.

Step 2: In your account information form update whatever you want

Step 3: Press the "Save Preferences" button.

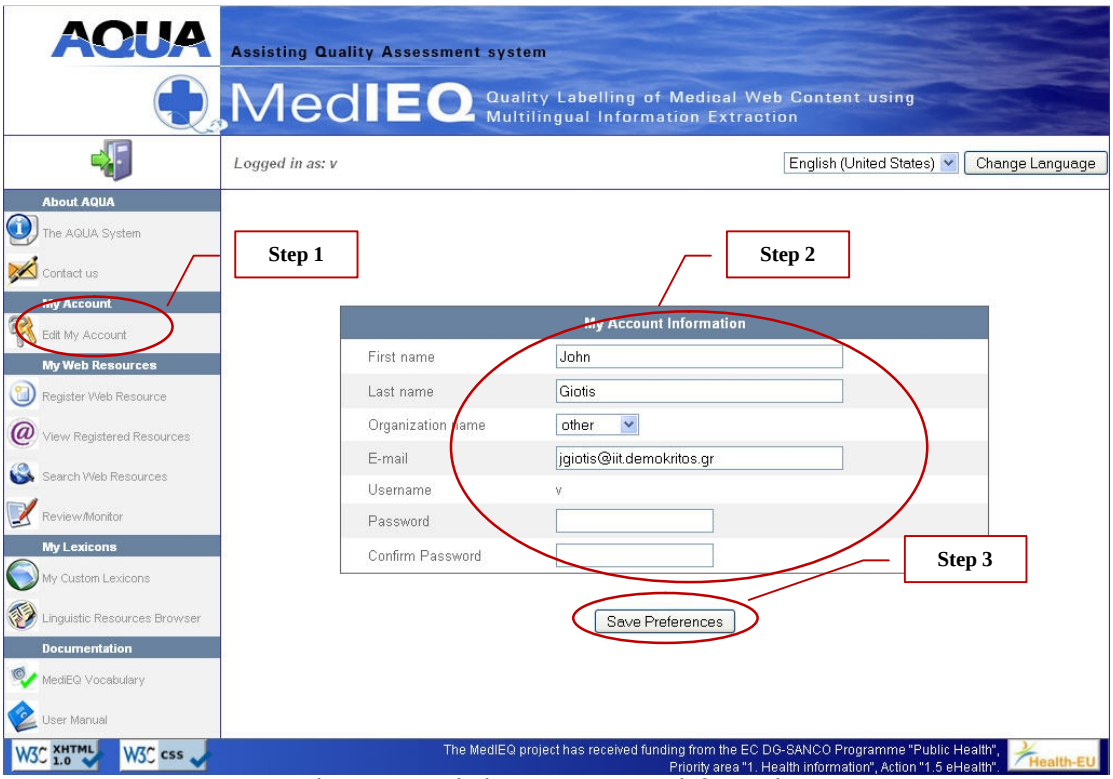

Figure 3. Updating you account information

2.3. Define a Web Resource

Step 1: From the submenu “My Web Resources” on your left, click on “Register Web Resource”. You then see the web form of Figure 4.

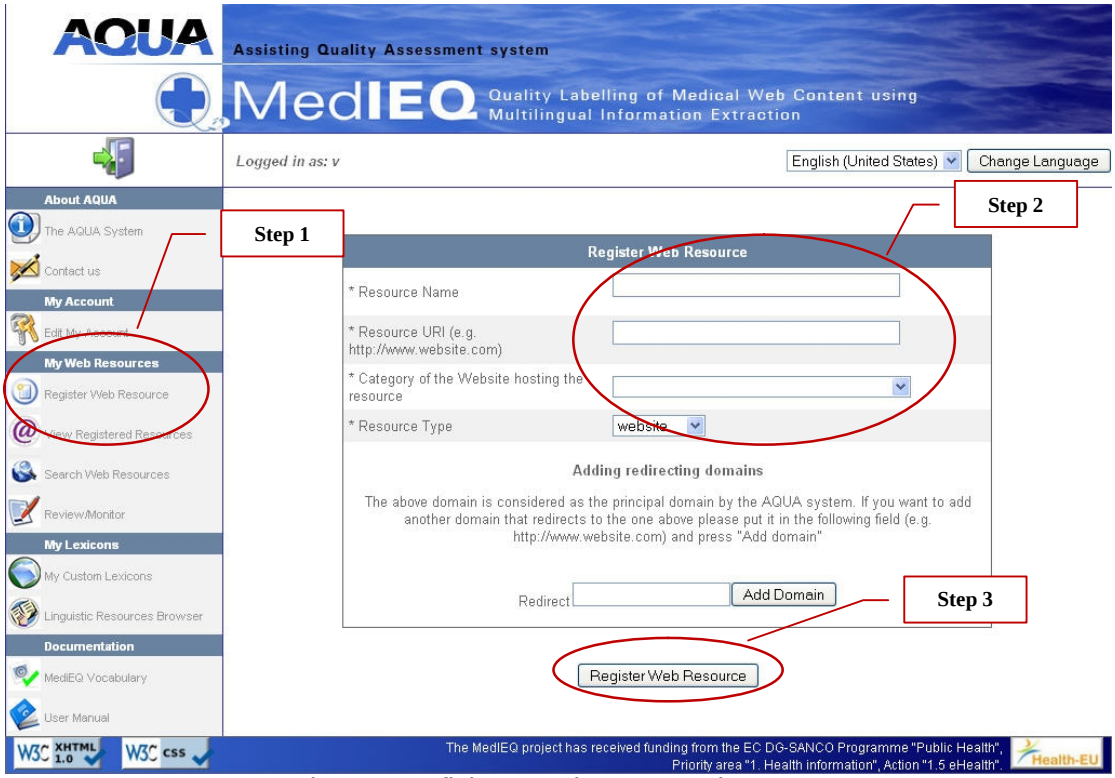

Figure 4. Defining a Web Resource in AQUA

Step 2: Fill the fields of the form with relevant information.

Step 3: Click the “Register Web Resource” button.

Figure 5. Viewing the list of My Web Resources

Step 4: Then, AQUA redirects you to a page displaying the list of web resources added by you or any other member of your organization (see Figure 5).

2.4. Manual Label Creation

Step 1: From the submenu “My Web Resources” on your left, click on “View Registered Resources”. You then see the web form of Figure 6.

Figure 6. Selecting a Web Resource in order to create a Label

Step 2: For the desired Web Resource, press the “Create Label” link. Then, AQUA will redirect you to a web form for defining the Label attributes for the selected Web Resource (see Figure 7 to Figure 9).

Figure 7. Manual Creation of a Label

Step 3: For a desired Label attribute, you have the option to directly define it via the corresponding interface. Alternatively, for the attributes that use a pre-defined vocabulary, you can use the corresponding vocabulary values by pressing the corresponding “proposed values” button and selecting the desired value for the attribute under review.

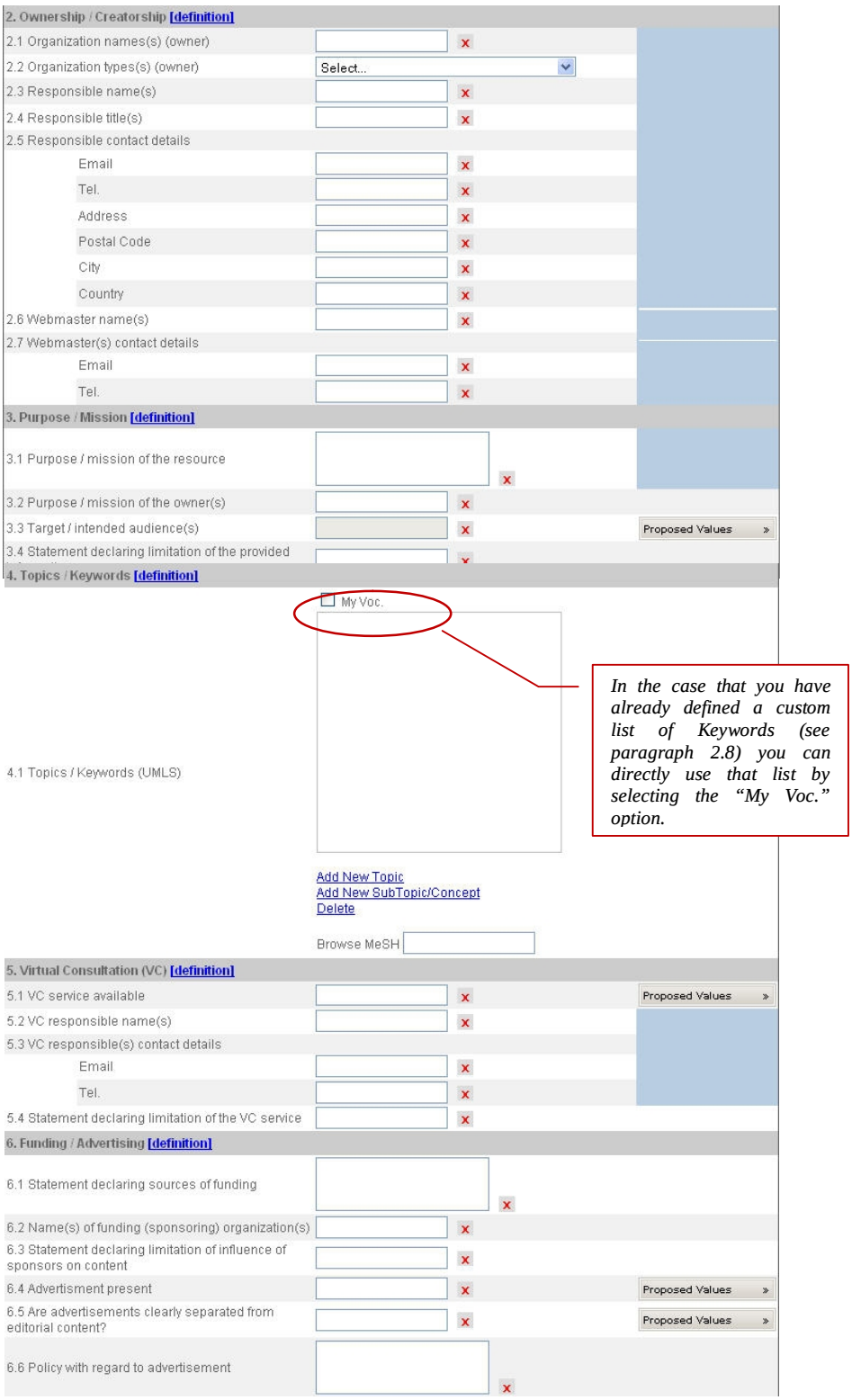

Figure 8. Manual Creation of a Label (cont.)

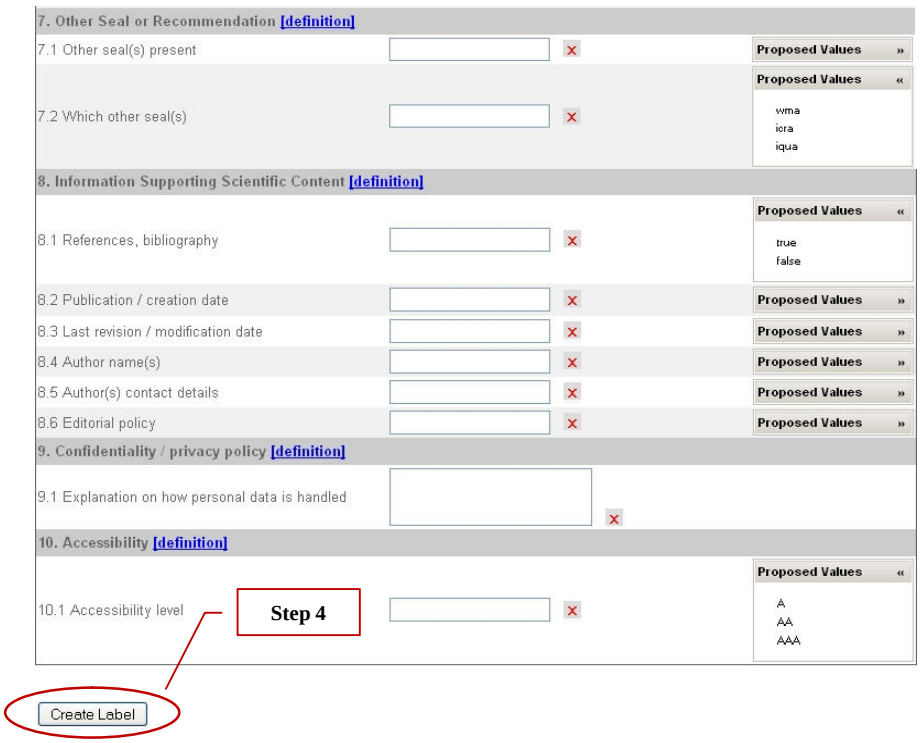

Figure 9. Manual Creation of a Label (cont.)

Step 4: Press the “Create Label” button, in order for the label to be saved.

2.5. Modify/Update an existing Label

Step 1: From the submenu “My Web Resources” on your left, click on “View Registered Resources”. You then see the web form of Figure 10.

Figure 10. Selecting a Web Resource in order to modify/update a Label

Step 2: For the desired Web Resource, press the “Edit Label” link. Then, AQUA will redirect you to a web form for updating the Label attributes for the selected Web Resource (see paragraph 2.4).

2.6. Download a Label in RDF-CL or POWDER Format

Step 1: From the submenu "My web resources" on your left, click on "View Registered Resources". You then see the web form of Figure 11, presenting the list of web resources defined in AQUA for your organization.

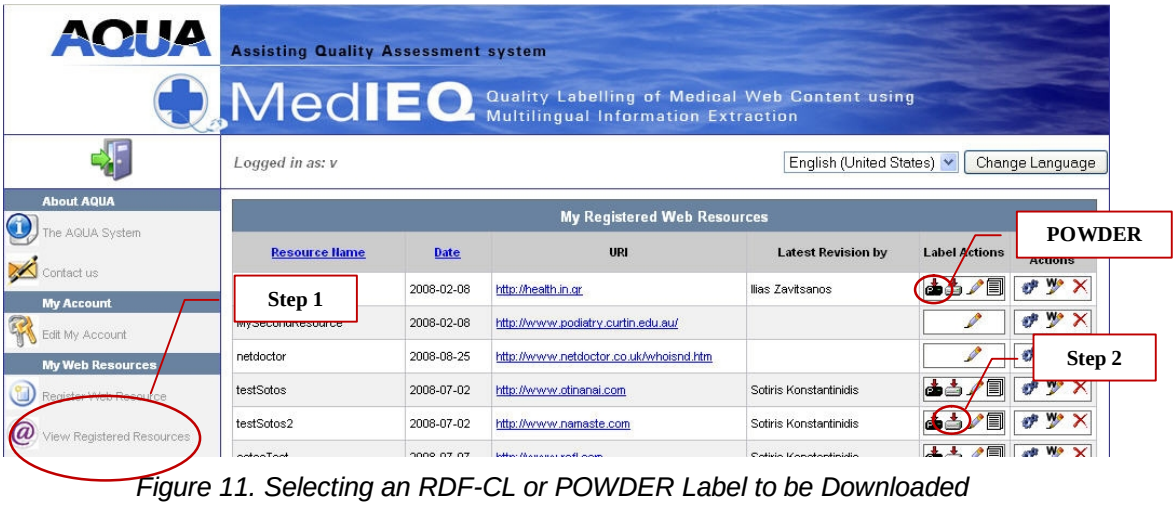

Figure 11. Selecting an RDF-CL or POWDER Label to be Downloaded

Step 2: For the desired resource press the download link. Then, AQUA will prompt you a dialog for defining a local path for the RDF-CL label to be saved (see Figure 12). The same procedure is being followed to download a Label in POWDER format.

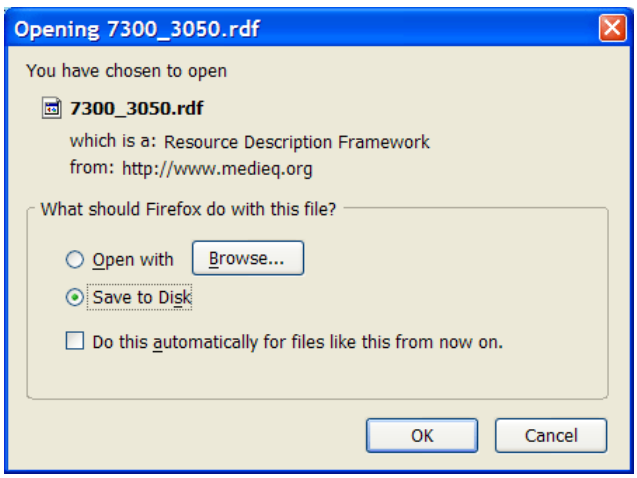

Figure 12. Saving an RDF-CL Label

2.7. Search for Unlabelled Web Resources

2.7.1. Create a Search Task

Step 1: From the submenu “My Web Resources” on your left, click on “Search Web Resources”. You then see the web form presented in Figure 13, from where you can create a new search task.

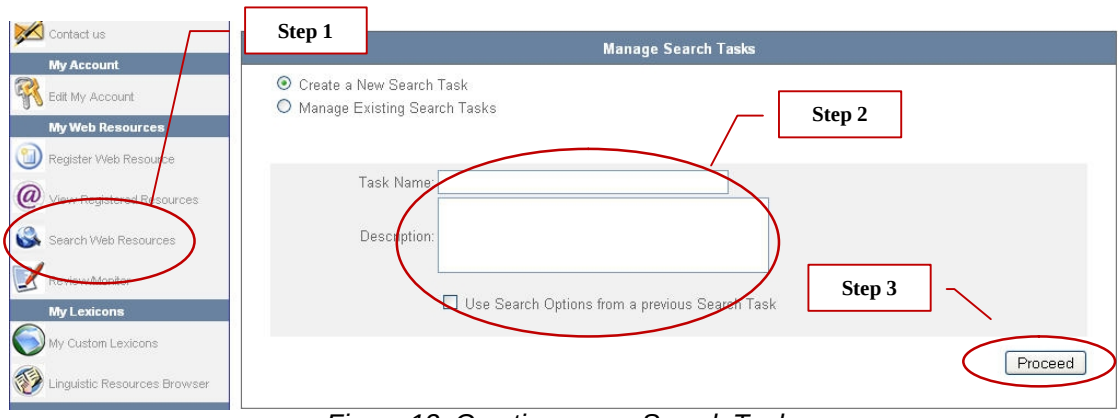

Figure 13. Creating a new Search Task

Step 2: Give a task name and a description, e.g. Task name: “My First Task”

Step 3: Click the “Proceed” button.

2.7.2. Define the Search Options

Step 1: After creating a new Search Task, AQUA redirects you to another form (see Figure 14), where you can setup the options of your search task.

Step 2: Here you see 3 tabs: “Search engines”, “Web directories” and “Black & white lists”. In the “Search engines” tab at the area “Queries”, add some keywords about a topic of your interest. You can put for example the keywords shown in Figure 14: “myocardial infarction, heart infarction, hearth attack, coronary syndrome”. Note that by putting these keywords in different lines we indicate the system to send separate queries to the search engines (each line corresponds to another query).

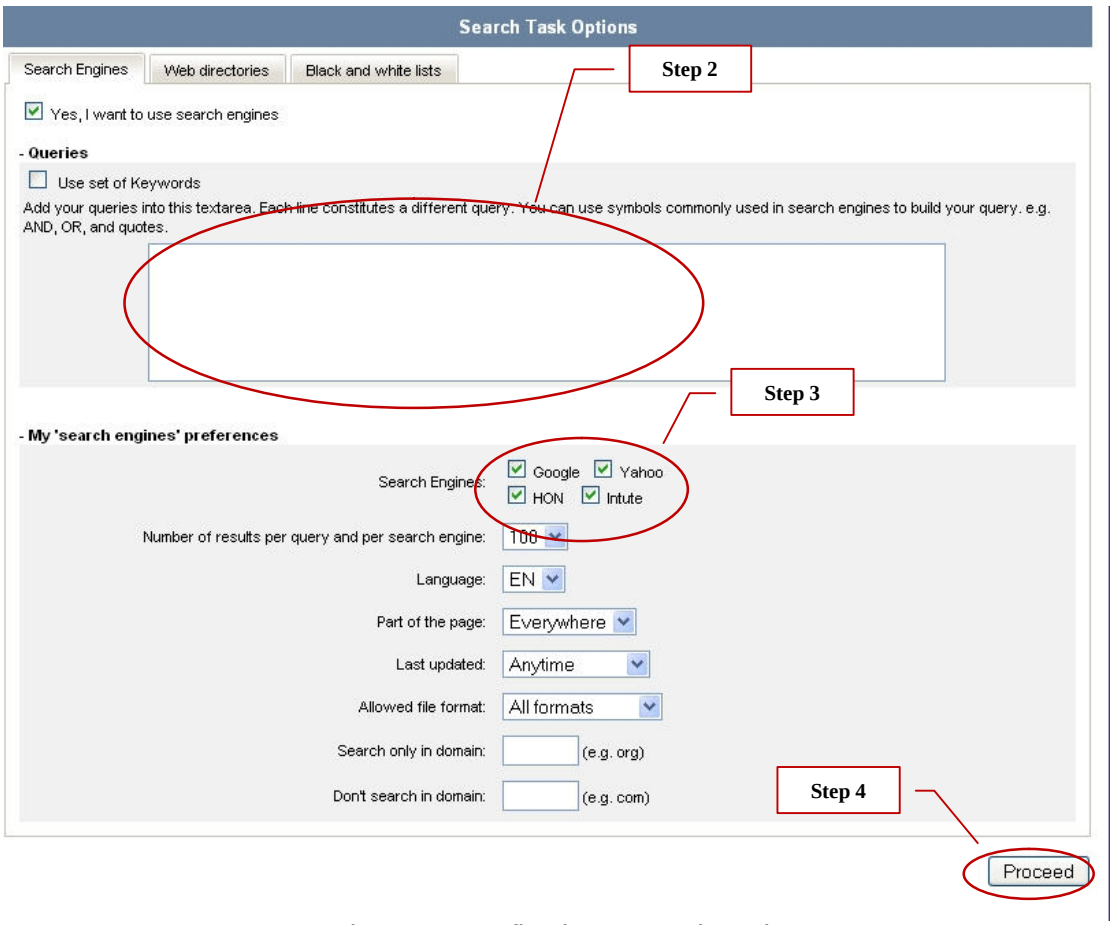

Figure 14. Configuring a Search Task

Step 3: Then, define the settings for the Search Engines to be used, such as the desired engines, the number of results per query etc.

Step 4: Press the "Proceed" button.

TIP

For refined searches using Web Directories and/or Black & White Lists, please refer to Appendixes B1 and B2 respectively.

2.7.3. Scheduling a Search Task

Step 1: After defining the Search Options, AQUA redirects you to the Search Task Scheduling interface (see Figure 15). From here you can

a. Ask AQUA to run your search task immediately, by checking the "Run task now" checkbox, and
b. Ask AQUA to send you an alerting e-mail once the task is completed, by checking the "Send me an e-mail when task finishes" checkbox.

Step 2: Click the “Finish” button.

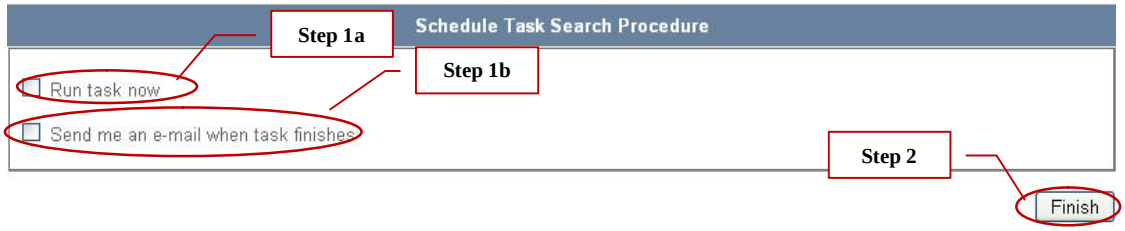

Figure 15. Scheduling a Search Task

Why Search Tasks?

Searching for new web resources may take some time (if, for example, the input keywords are numerous). Also, a user may desire to initialize several search tasks, running simultaneously and in parallel. In this context, search for new web resources is performed in search tasks. The user creates and configures search tasks, each task corresponding to a different search (e.g. one task searching for resources containing information related to diabetes mellitus, a second task discovering content referring to Alzheimer’s disease, etc.).

2.7.4. View Search Results

Step 1: [OPTIONAL] In the case that you have enabled the email notification option of the Search Task Scheduler, after executing a Search Task, you should check your mailbox: once your search task is completed, AQUA sends an alert message with subject “Search task ..... completed”.

Step 2: Go to the submenu My Web Resources > Search Web Resources. You then see the list of all the search tasks that you or any other member of your organization has created (see Figure 16).

Step 3: Select the “Manage existing tasks” option.

Step 4: By pressing the “Results” link next to the desired Search Task name, you see the results as in Figure 17.

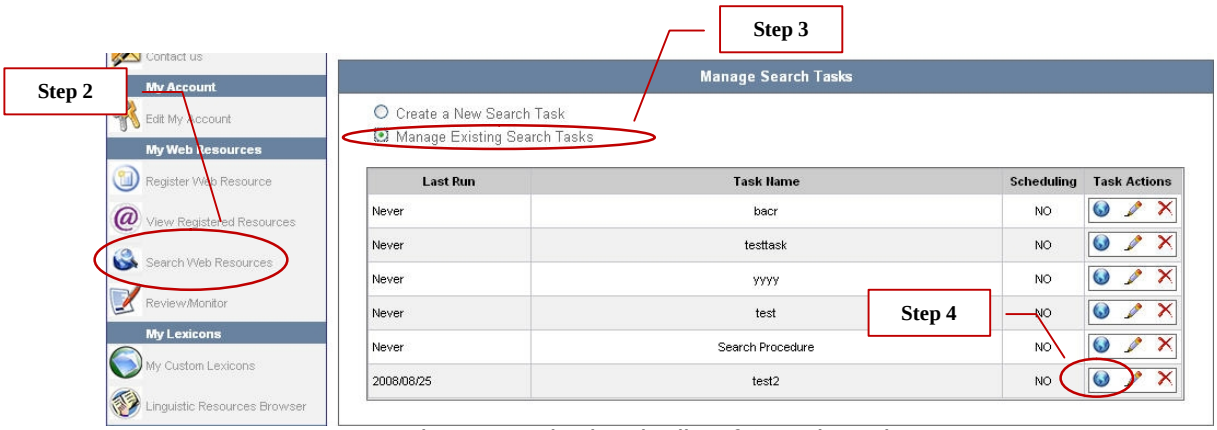

Figure 16. Viewing the list of Search Tasks

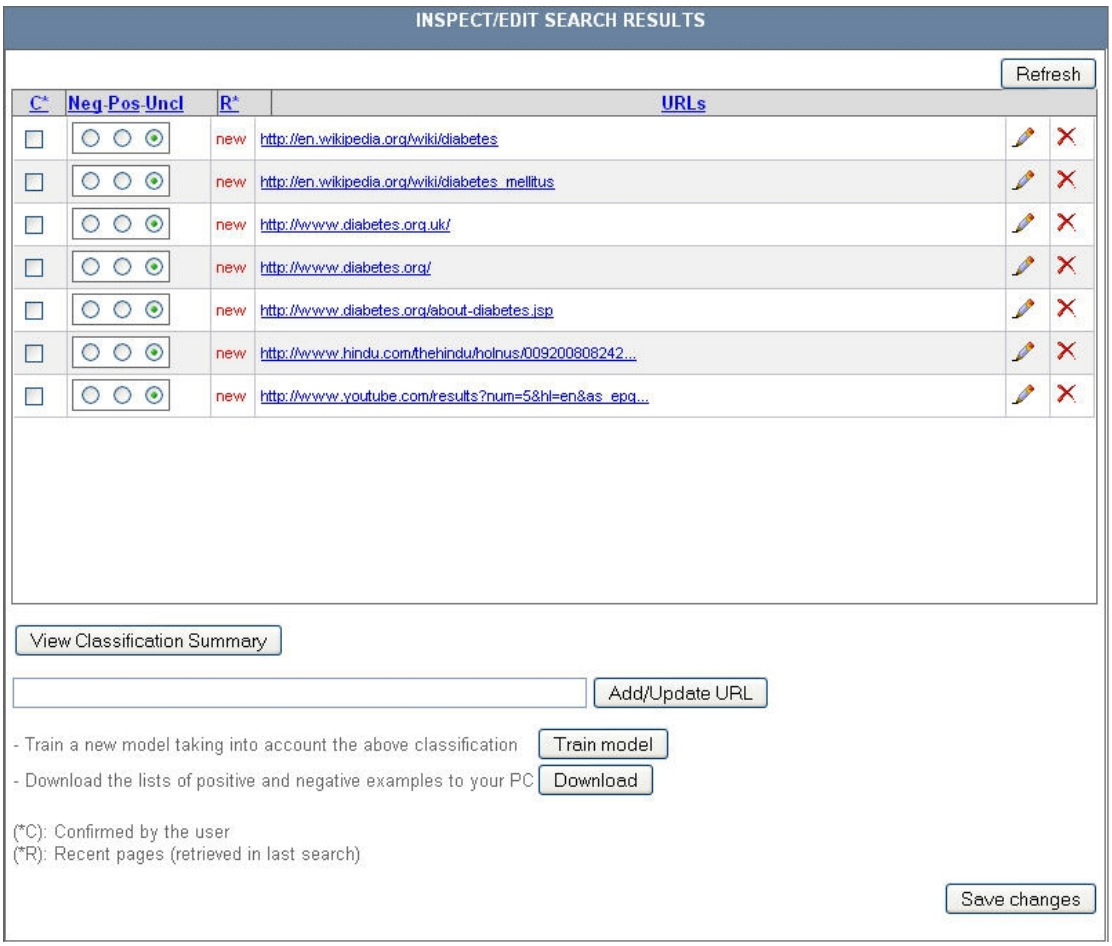

Figure 17. Viewing the Results of a Search Task

2.8. *Create a Custom Search Vocabulary*

2.8.1. Create a Linguistic Resource

Step 1: Go to: “My Lexicons” on your left, and click on “My Custom Lexicons”. You then see the form of Figure 18

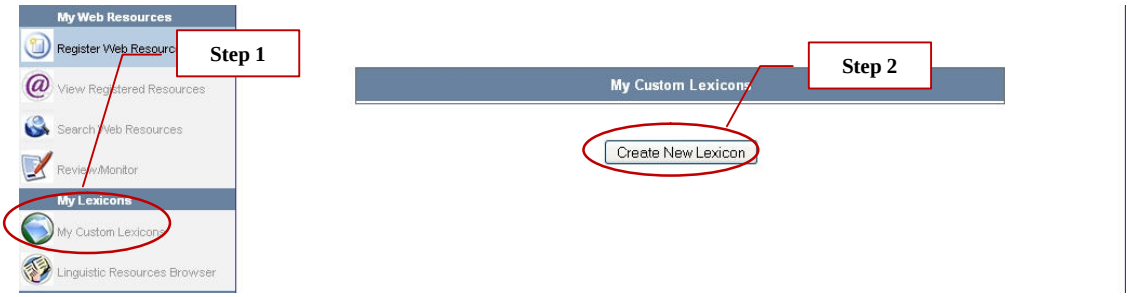

Figure 18. Create a Custom Search Vocabulary

Step 2: Click the “Create New Lexicon” link. Then, AQUA presents the form of Figure 19, where you can define a new Linguistic Resource (set of keywords)

Figure 19. Defining a new Linguistic Resource

Step 3: Define a Name and [OPTIONAL] provide a description for this new Lexicon.

Step 4: Add manually concepts and/or sub-hierarchies of concepts to the selected Lexicon.

Step 5: Click the "Create New Lexicon" button. Then, AQUA will redirect you to a new web form presenting the list of resources that you have created (Figure 20).

2.8.2. Adding Concepts in a Linguistic Resource

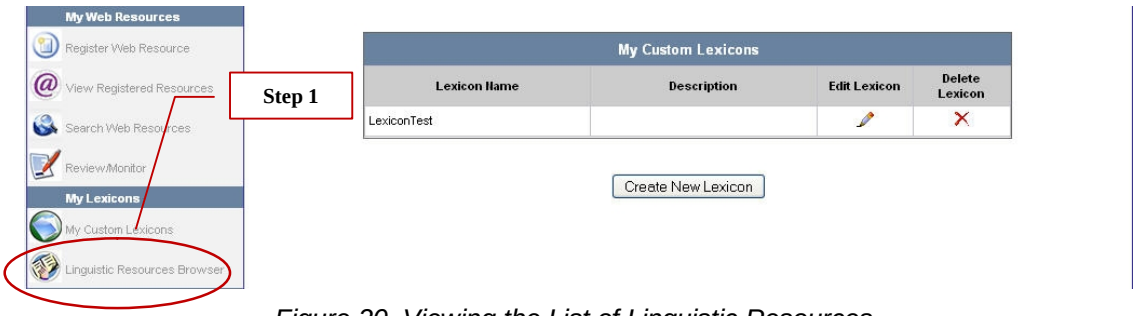

Figure 20. Viewing the List of Linguistic Resources

Step 1: Go to: "My Lexicon" on your left, and click on "Linguistic Resources Browser". You then see the form of Figure 21.

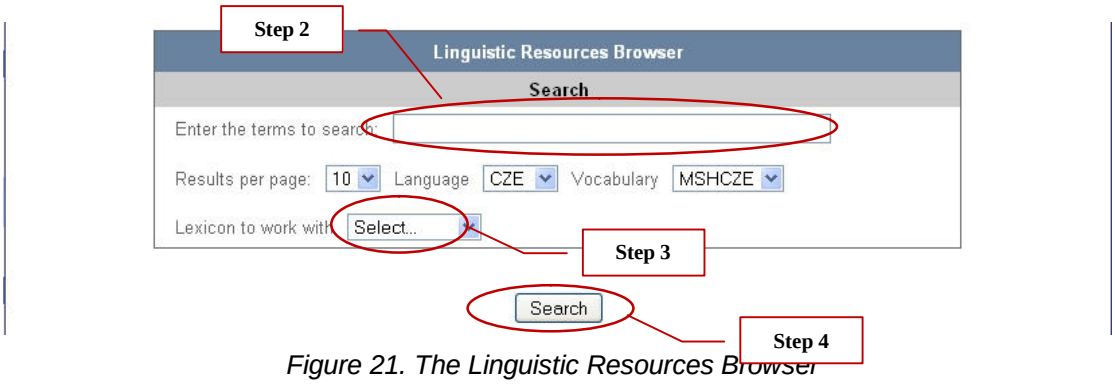

Figure 21. The Linguistic Resources Browser

Step 2: Configure the Terms Search properties of the Linguistic Resources Browser, by defining a term to be found, the desired search results per page, the language of the desired terms, as well as the vocabulary to be used for searching terms.

Step 3: Select the desired Resource to work with.

Step 4: Press the "Search" button. The term search results are presented in the lower part of the Linguistic Resources Browser, as shown in Figure 22.

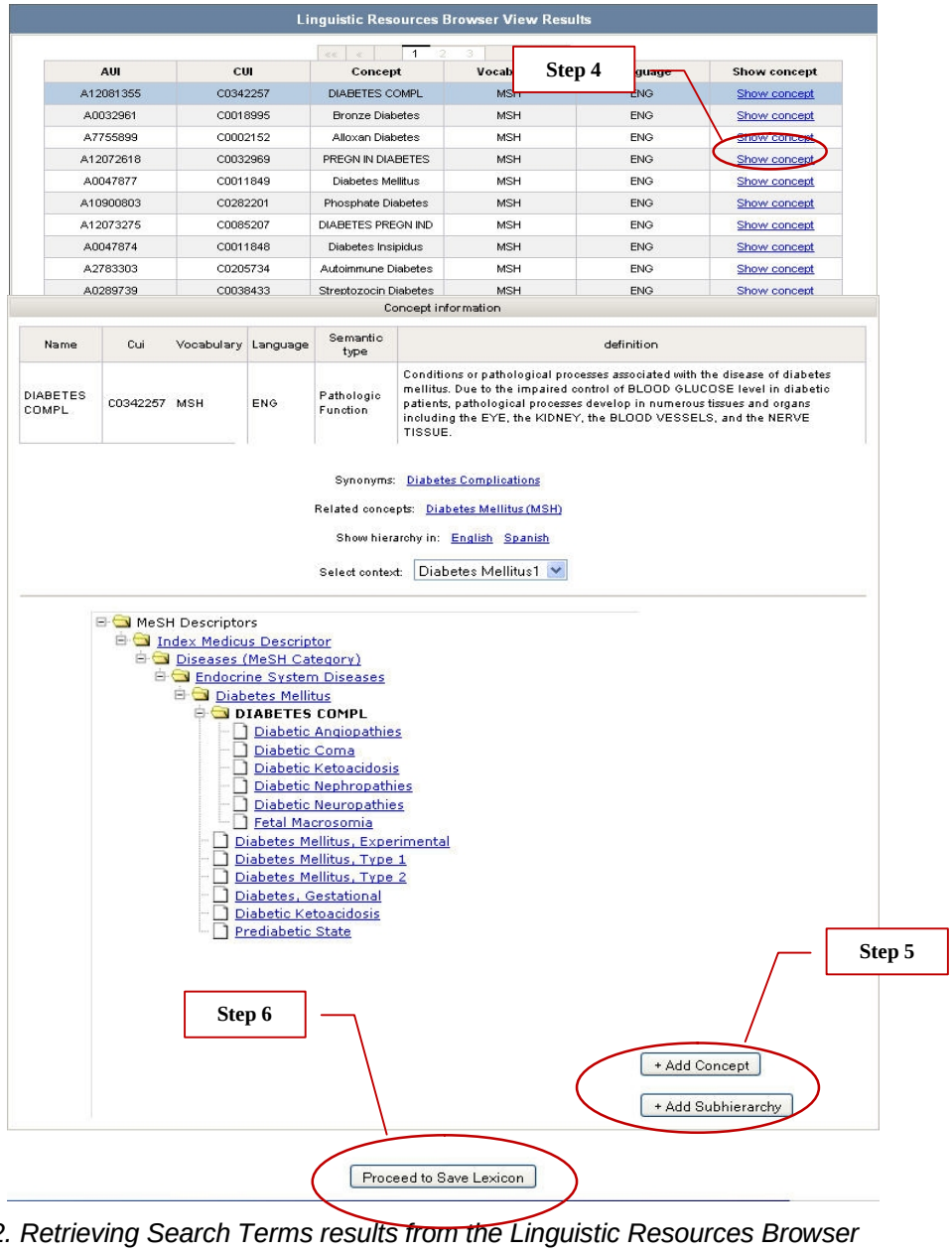

Figure 22. Retrieving Search Terms results from the Linguistic Resources Browser

Step 4: Select the desired Concept from the list of found ones.

Step 5: Add concepts and/or sub-hierarchies of concepts to the selected Lexicon Resource.

Step 6: Press the “Proceed to Save Lexicon” button. You can see the form of Figure 23.

Figure 23. Review before saving the current Linguistic Resource

Step 7: Before saving the Lexicon Resource you can manually add concepts and/or sub-hierarchies of concepts to the selected Lexicon Resource.

Step 8: Press the “Update Lexicon” button in order to save the Lexicon Resource.

2.9. *Semi-automatic Label Creation*

2.9.1. Create a Review/ Monitor Task

Step 1: From the submenu “My Web Resources”, click on “Review/Monitor”. You then see the web form of Figure 24.

Step 2: Select the “Create a new review task” option

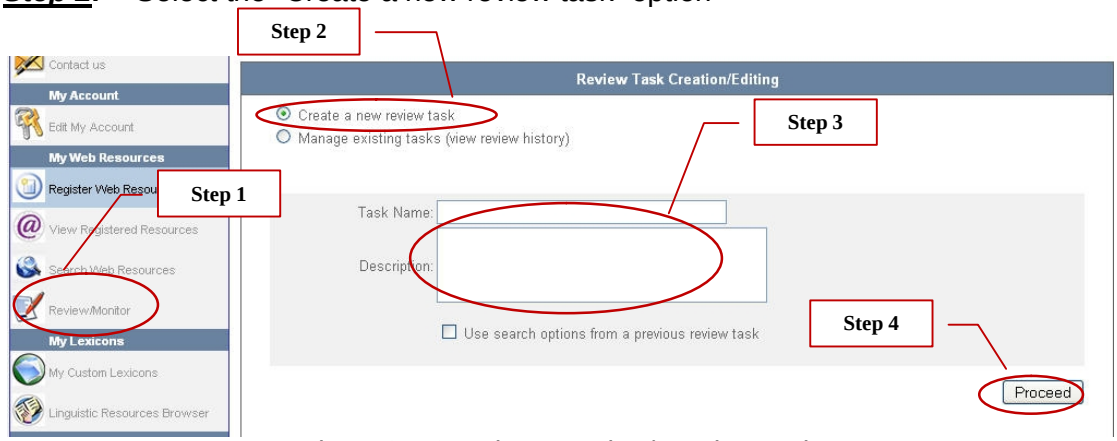

Figure 24. Creating a Review/ Monitor Task

Step 3: Define a Task Name and [OPTIONAL] provide a description for this new Task.

Step 4: Click the “Proceed” button.

2.9.2. Configuring a Review/ Monitor Task

Step 1: After creating a new Review/ Monitor Task, AQUA redirects you to another form (see Figure 25), where you can setup the options of your review/monitor task.

Step 2: Define the desired URL under review in the corresponding text field

Step 3: Press the “Add/Update URL” button. The defined URL now appears in the area above the button.

Step 4: Check the box next to the URL indicating which web resources AQUA will Spider.

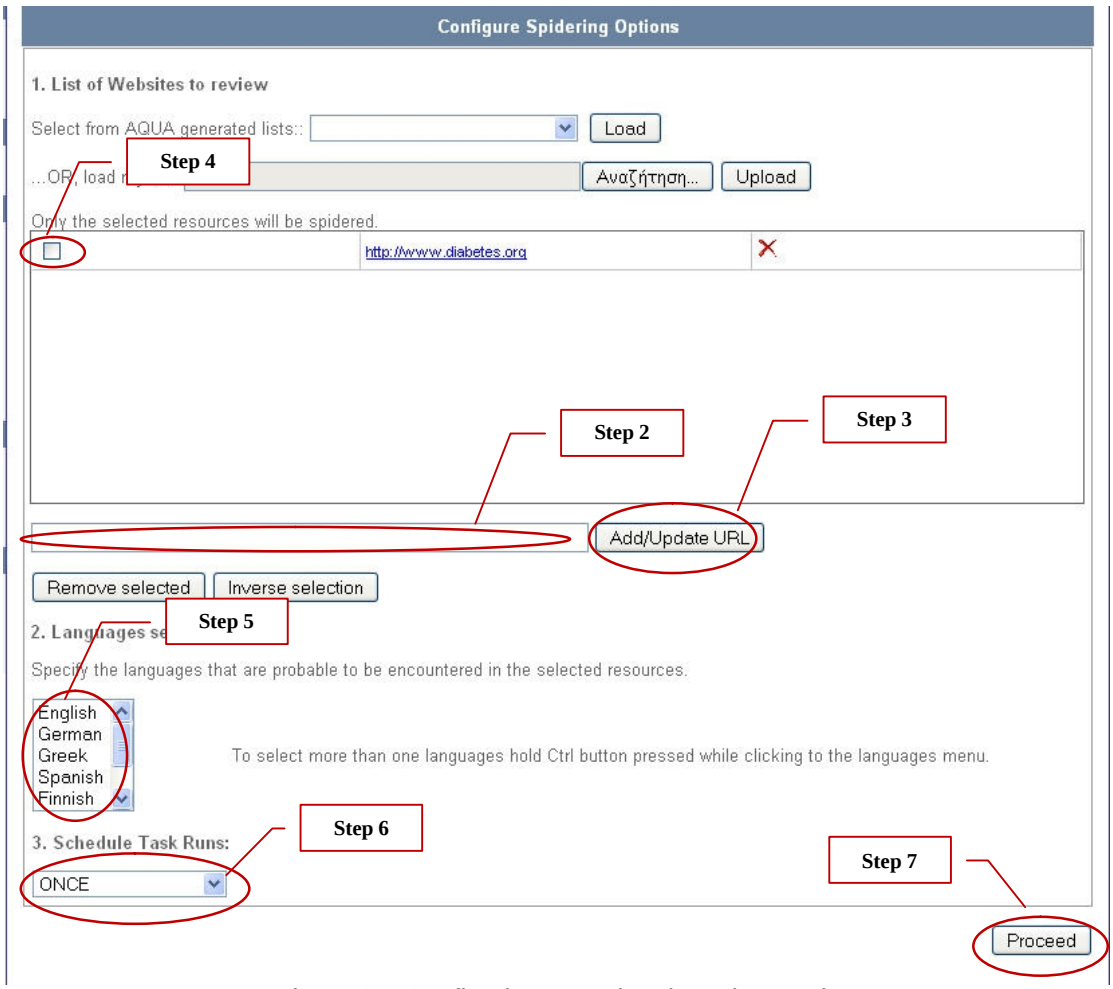

Figure 25. Configuring a Review / Monitor Task

Step 5: Select the language to which the defined Web Site (URL) provides content

Step 6: Select the desired frequency the task should ran.

Step 6: Press the “Proceed” button

2.9.3. Scheduling a Review/Monitor Task

Step 1: After configuring the Review/Monitor Task, AQUA redirects you to the Review/Monitoring Task Scheduling interface (see Figure 26). From here you can

a. Ask AQUA to run the review/monitor task immediately, by checking the “Run task now” checkbox, and
b. Ask AQUA to send you an alerting e-mail once the task is completed, by checking the “Send me an e-mail when task finishes” checkbox.

Step 2: Click the “Finish” button.

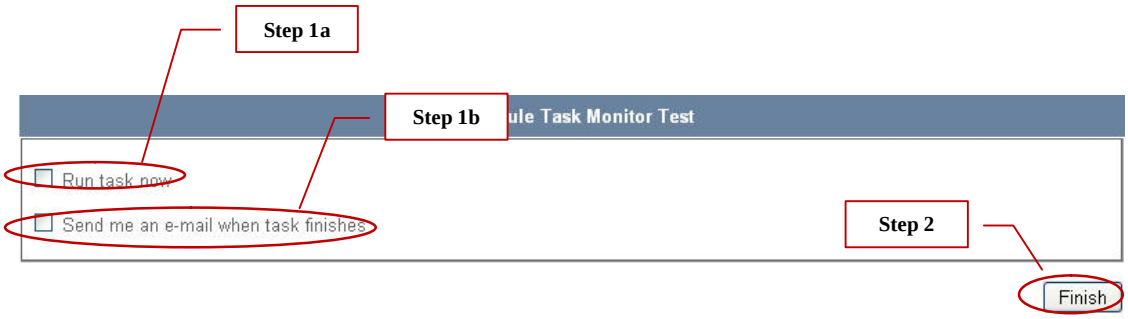

Figure 26. Scheduling a Review/Monitor Task

Why Review/ Monitor Tasks?

Reviewing a web resource may take some time (this time being proportional to the number of web pages/documents a resource may offer). Also, a user may desire to initialize several review/monitor tasks, running simultaneously and in parallel. In this context, reviewing web resources is performed in review/monitor tasks.

2.9.4. Using the AQUA proposed values for creating a Label

Step 1: [OPTIONAL] In the case that you have enabled the email notification option of the Review/Monitor Task Scheduler, after executing a Review/Monitor Task, you should check your mailbox: once your search task is completed, AQUA sends an alert message with subject "Review/Monitor task ..... completed".

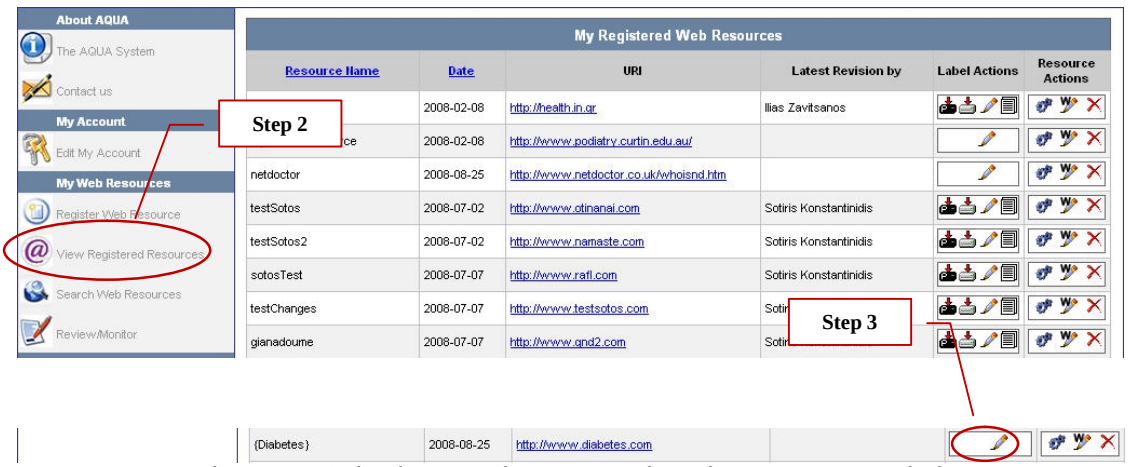

Figure 27. Selecting a Web Resource in order to create a Label

Step 2: Go to the submenu "My Web Resources" >"View Registered Resources". You then see the list of all the web resources that you or any other member of your organization has created. Now in this list a new web resource has been added with the same name as the corresponding review/monitor task (see Figure 27).

Step 3: For this Web Resource, press the "Create Label" link. Then, AQUA will redirect you to a web form for defining the Label attributes of the selected Web Resource (see Figure 28 to Figure 30).

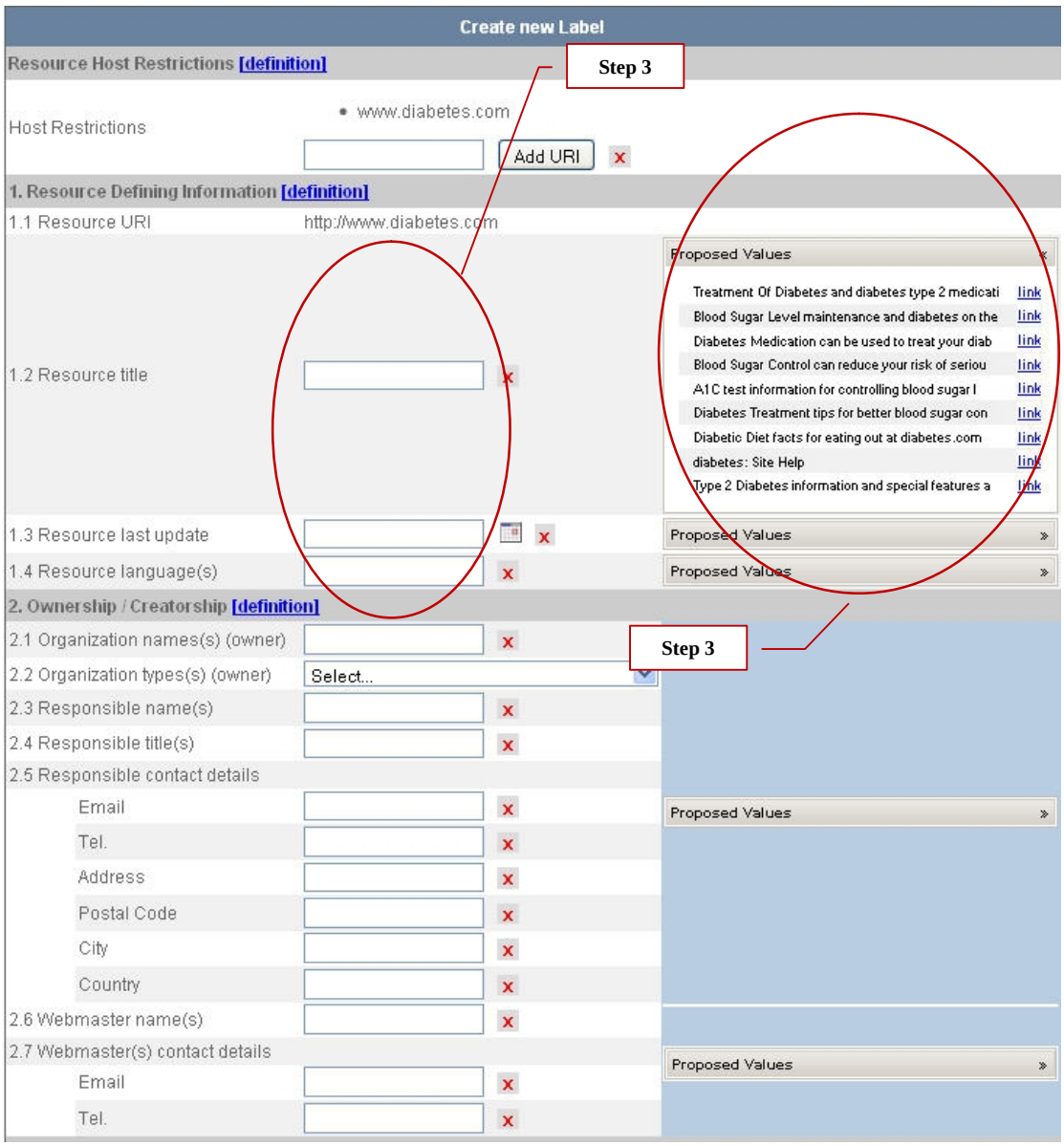

Figure 28. Semi-automatic Creation of a Label

Step 3: For a desired Label attribute, you have the option to directly define it via the corresponding interface. Alternatively, for the attributes that AQUA has found relevant values (from the corresponding review/monitor task), you can use these values by pressing the corresponding "proposed values" button and selecting the desired value for the attribute under review. Each proposed by AQUA value is supported by relevant hyperlinks. By pressing a "link" the corresponding web page opens as a separate window in your default browser.

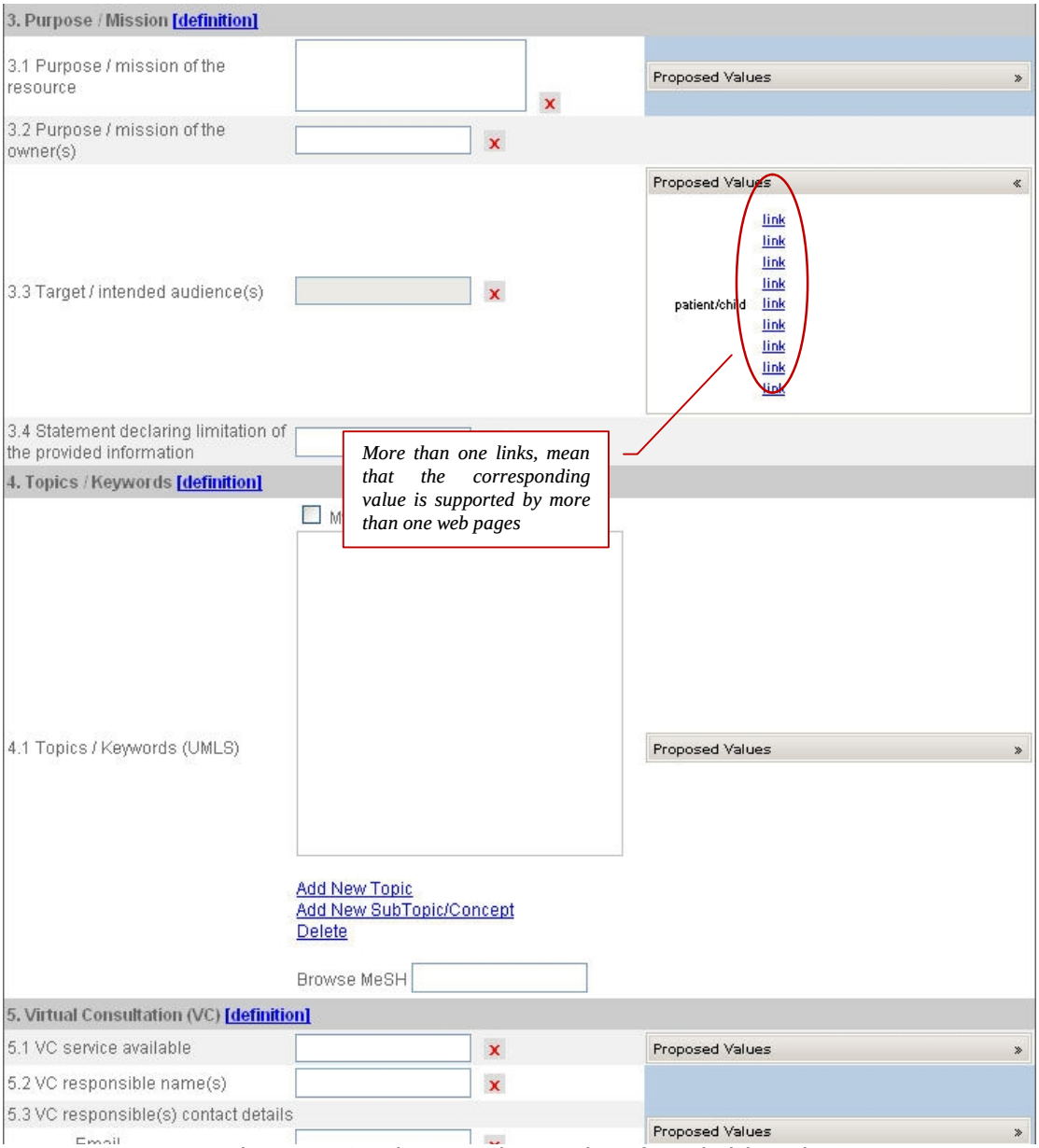

Figure 29. Semi-automatic Creation of a Label (cont.)

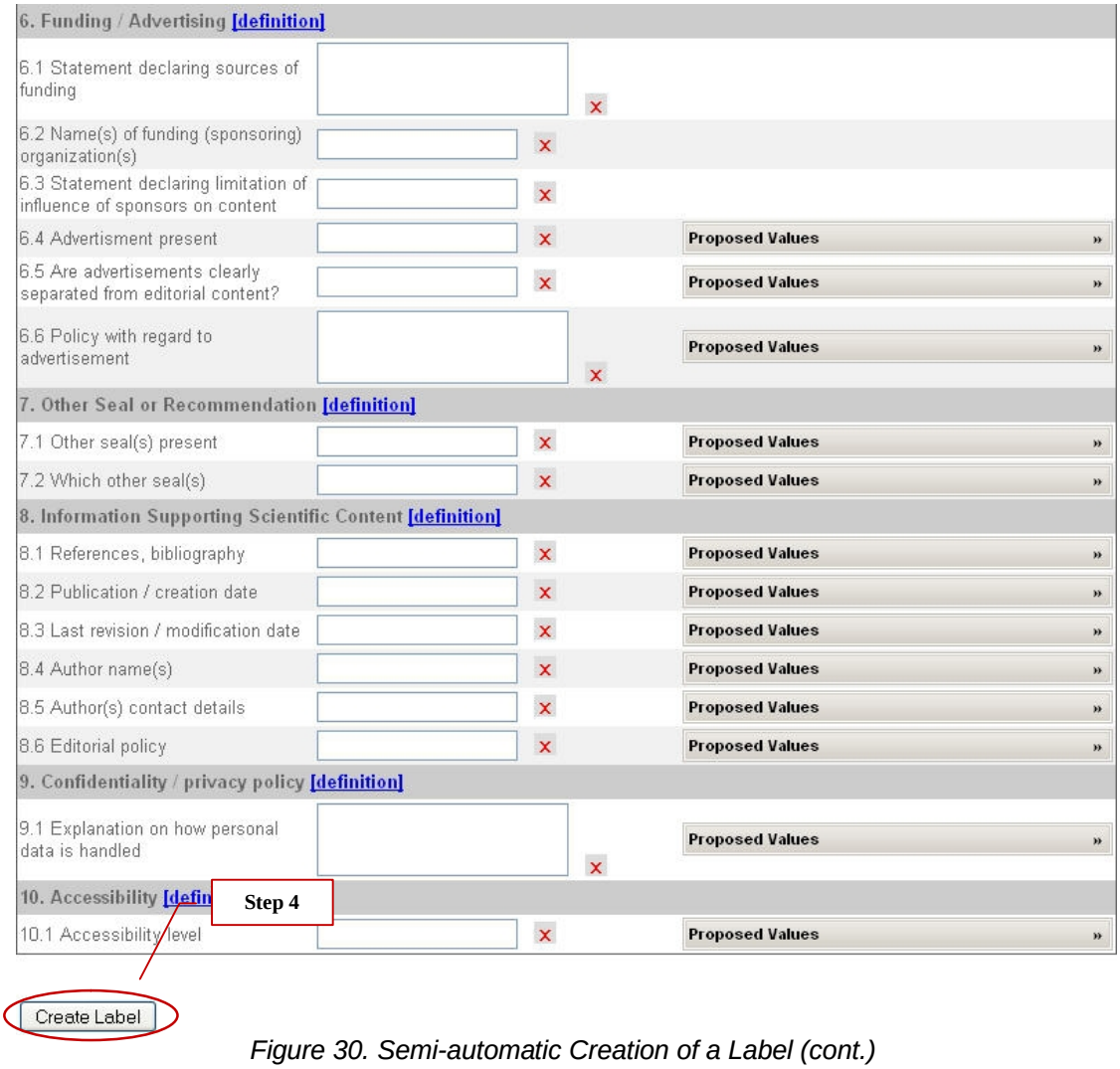

Figure 30. Semi-automatic Creation of a Label (cont.)

Step 4: Press the "Create Label" button, in order for the label to be saved.

NOTE

AQUA proposes automatically values, for the following MedIEQ criteria:

1.2 Resource title
1.3 Resource Last update
1.4 Resource Language(s)
2.1 Organization Name(s) (Owner)
2.3 Responsible Name(s)
2.4 Responsible Title(s)
2.5 Responsible Contact Details
3.3 Target/Intended Audience(s)
4.1 Topics/Keywords (UMLS)
5.1 VC Service Available
6.4 Advertisement present
7.2 Which other seal(s)

Appendix A. The MedIEQ Labelling Criteria

1. Resource Defining Information	
Descriptor	1. Resource defining information
Attributes	1.1 Resource URI 1.2 Resource title 1.3 Resource last update 1.4 Resource language(s)
Definition	Includes information identifying/describing the resource. Concerning the resource URI: a) whether the resource's URI is valid or not and b) in case it redirects to external domains, are these domains between those specified when the resource was added? The rest is information like the resource's last update, its title and the language(s) in which content is provided.

2. Ownership / Creatorship	
Descriptor	2. Ownership / Creatorship
Attributes	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
Definition	The user should know who is behind the resource in order to judge by himself the credibility of the provided information. Therefore, information like the name(s) of the organization(s) providing the information and the type of this(these) organization(s) should be available. At the same time, the name(s), title(s) (e.g. MD, PhD, Dr, etc.) and contact details of website responsible(s), to contact in case of questions on health related issues, as well as the name(s) and contact details of the webmaster(s) should be available.
Examples	Organization type(s) may be one or more of the following: - Government Organization - Healthcare service provider - Media and publishers - Pharmaceutical company / retailer - Universities / research institutions - Scientific or professional organizations - Patient organizations / self-support groups - Private individual - Other

3. Purpose / mission	
Descriptor	3. Purpose / mission
Attributes	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
Definition	It has to be clear for the user which is the goal and motivation of the provided information and for what kind of users it was created e.g. adults, children, people with diabetes, etc.
Examples	*Regarding 3.4*: Should be something like “the health information is provided to support and not to replace the relationship that exists between visitors and their health professionals”.

4. Topics / Keywords	
Descriptor	4. Topics / Keywords
Attributes	4.1 Topics / Keywords (UMLS)
Definition	Mapping of the resource’s contents to concepts from the UMLS Metathesaurus.
Examples	*Porfiria Veteada C0162532* /*category: *Disease or Syndrome/ /*Description: *An autosomal dominant porphyria that is due to a deficiency of protoporphyrinogen oxidase (EC 1.3.3.4) in the LIVER, the seventh enzyme in the 8-enzyme biosynthetic pathway of HEME. Clinical features include both neurological symptoms and cutaneous lesions. Patients excrete increased levels of porphyrin precursors, COPROPORPHYRINS and protoporphyrinogen/ /*has parent*: _Porfirias Hepáticas|_/ /*Can be qualified by: *_blood, __cerebrospinal fluid,_chemically induced_.../ /*siblings:* C_oproporfiria Hereditaria,_ _Porfiria Intermitente Aguda_.../

5. Virtual consultation	
Descriptor	5. Virtual consultation
Attributes	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
Definition	Is there an online virtual consultation (VC) service available? A VC service is an online service allowing the user to ask questions and/or

	send/upload information on health related issues asking for advice. A VC service may have one of the following forms: a. discussion forum, b. chat, c. VC e-mail, d. VC request form. The name(s) and details of the person(s) responsible(s) for this service should also be clearly mentioned. Moreover, a declaration that VC is only a supporting means that cannot replace a personal consultation with a physician should be provided.
Examples	*Regarding 5.4*: Should be something like "the VC service can not replace a personal consultation with a physician" or "internet based advice, whether personalised or not, cannot replace a face to face consultation with a healthcare practitioner".

6. Funding / Advertising	
Descriptor	6. Funding / Advertising
Attributes	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
Definition	Health web resources should disclose possible conflicts of interest. For this reason it is important to know how and by whom a web resource is funded. If there are any sponsors, it has to be clear who they are. Furthermore, it should be stated that sponsors do not have any influence on the content. Additionally, it has to be known whether the web resource hosts or not advertising material in whatever format. In case that happens, such material should be clearly distinguished from informative material. Furthermore, information on resource's policy with regard to advertising must be easily accessible and clear.
Examples	*Regarding 6.1*: This must be something like "site sponsors are xx, yy, zz" or "About our site sponsor: …" or "this site is kindly sponsored by …" or "we are a non profit organization supported by individuals, foundations, and corporations" or "this web site is sponsored by donations", etc. *Regarding 6.3*: Something like "sponsorship will not be accepted in situations where the fact of the sponsorship would raise an inference of influence on editorial content or decision-making, or of xx endorsement of the sponsor or its products and services". *Regarding 6.6*: Something like: "*… guidelines have been established by our site to govern various aspects of Advertising … including banner, button, and contextual Advertising, Sponsorship Messages, and Promotions*".

7. Other Seal or Recommendation	
Descriptor	7. Other seal or Recommendation
Attributes	7.1 Other seal(s) present 7.2 Which other seal(s)?
Definition	Are there other seals identified in the resource? Indicates that the resource already conforms to other, known quality criteria. Identifiers for other seals: a) Real seals: WMA, HONcode, pWMC, URAC, eHealth TRUST-E, AFGIS, b) Filtering health portals (a resource is recommended by): AQUMED, Intute, WHO ("Vaccine Safety Net")

8. Information Supporting Scientific Content	
Descriptor	8. Information Supporting Scientific Content
Attributes	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
Definition	Regarding the provided specialized health information (scientific parts of the resource) it is relevant to know if it is based on scientific books, medical journal articles, etc. For this, scientific articles or documents should include a references or bibliography section. Additionally, it is important to know if such information is up-to-date (publication and last modification dates are required) and who is the author of such content (author(s) name(s) and contact details are required for pages/documents providing scientific information).
Examples	*Regarding 8.6* This must be something like “Information provided by this resource (or hosted in this site) • is selected upon the following procedure ... which includes health professionals/specialists etc.", or • is controlled/reviewed by medical doctors", or • is given only by medically trained and qualified professionals unless a clear statement is made that a piece of advice offered is from a non-medically qualified individual or organisation."

9. Confidentiality / privacy policy	
Descriptor	9. Confidentiality / privacy policy
Attributes	9.1 Explanation on how personal data (visitor coordinates, e-mail messages, etc.) is handled
Definition	Internet users are much concerned about protection of their privacy and personal data. For this reason the resource should provide a confidentiality/privacy policy ensuring that personal data (visitor coordinates, e-mail messages, etc.) is safely handled, describing how

	these data are handled.
Examples	*Regarding 9.1*: To declare e.g. that “any personal data used by xx, including any information collected through this website will be treated as strictly confidential …” or “the use of this data is strictly for statistical purposes …etc.” or “organization xx pledges always to respect the privacy and anonymity of its users, including survey participants” or “the information provided to us voluntarily by our visitors in their communications with us is securely stored and not shared with any third party” or clearly explain “With whom the information may be shared” or to state “the kind of security procedures that are in place to protect the loss, misuse or alteration of information” or “What choices are available to users regarding collection, use and distribution of the information” or “use personally identifiable information for internal purposes” or explain how they are using cookies or if they back up the information collected or “maintain a very strict privacy policy”, etc.

10. Accessibility	
Descriptor	10. Accessibility
Attributes	10.1 Accessibility level
Definition	The resource is examined upon various accessibility criteria and information on its accessibility level (whether the resource is of level A, AA or AAA) is deduced.

Appendix B. The AQUA Crawler Configuration

B1. Configuring "Web Directories"

In the tab "Web directories" of the AQUA Crawler you put the Web directory URLs of your choice. The Crawler will parse their listings and collect all contained URLs. For example, if you want to search for diabetes organizations you put "http://dmoz.org/Health/Conditions_and_Diseases/Endocrine_Disorders/Pancreas/Diabetes/Organizations/" in the relevant area. Note, however, that there is a specific syntax to follow when putting Web directory URLs.

There are four different, possible syntax formats:

- [L] <start URL> <strings to exclude>
- [L1] <start URL> <URL regular expression> <group> <strings to exclude>
- [S] <start URL> <strings to exclude>
- [S1] <start URL> <URL regular expression> <group> <strings to exclude>

(All fields are separated by a single space character)

e.g.

- [L] http://search.centrum.cz/s-1303000-hubnuti search centrum
- [S] http://search.centrum.cz/s-5629004-zdravotnicka-zarizeni search centrum
- [S1] http://katalog.quick.cz/k.aspx?id=661 &url=(((http|https)://|www)[a-z0-9\-\._]+/?[a-z0-9_\.\-\?\+/~=&#;,%]*[a-z0-9/]{1})(&katID=661)? 1 katalog quick

In the first format "L" means that the start URL given represents a leaf page, which means that the crawler should not search deeper in this page. <start URL> is the URL provided by the user where the crawler will start searching from. In <strings to exclude> field we specify strings that should not be contained in the final list of URL that will be returned. For example if we search in dmoz directory we know that URL containing the string dmoz.org should not be contained in the list as they are not really URLs to external pages that we need. They actually are links to other pages in dmoz domain. The crawler uses the following regular expression to extract URLs for HTML source: href="?(((http|https)://|www)[a-z0-9\-\._]+/?[a-z0-9_\.\-\?\+/~=&#;,%]*[a-z0-9/]{1})["><]

In the second format ("L1" type) user provides the crawler with a regular expression that will match the URLs he wants to extract. This could help if the user knows that a specific web directory uses a standard pattern to include external URLs in his HTML

code. The <group> field indicates which group of regular expression the user wants to keep. This group is supposed to be the part of the regular expression that matches the URL.

The third and the forth formats ("S" and "S1" types) are similar to the first and the second with the difference that they represent subtrees rather than leaf web pages. This means that if an extracted URL is sub-URL (lower in hierarchy) of the initial given URL then the crawler will follow this URL in order to go deeper and extract more URLs.

B2. Configuring “Black &White Lists”

In the tab “Black & white lists” of the AQUA Crawler you can put either URLs that you want to be filtered out from Crawler’s results (black list) or URLs that you would absolutely like to see between the results (white list).

B3. Get better results by training the Classification Model

By using the AQUA Crawler, a user is able to make searches for health-related Web Sites based on a user-defined set of keywords. These results are presented in the form depicted in Figure 31. By clicking on a URL from the received URLs list, the corresponding web page opens in a new browser window.

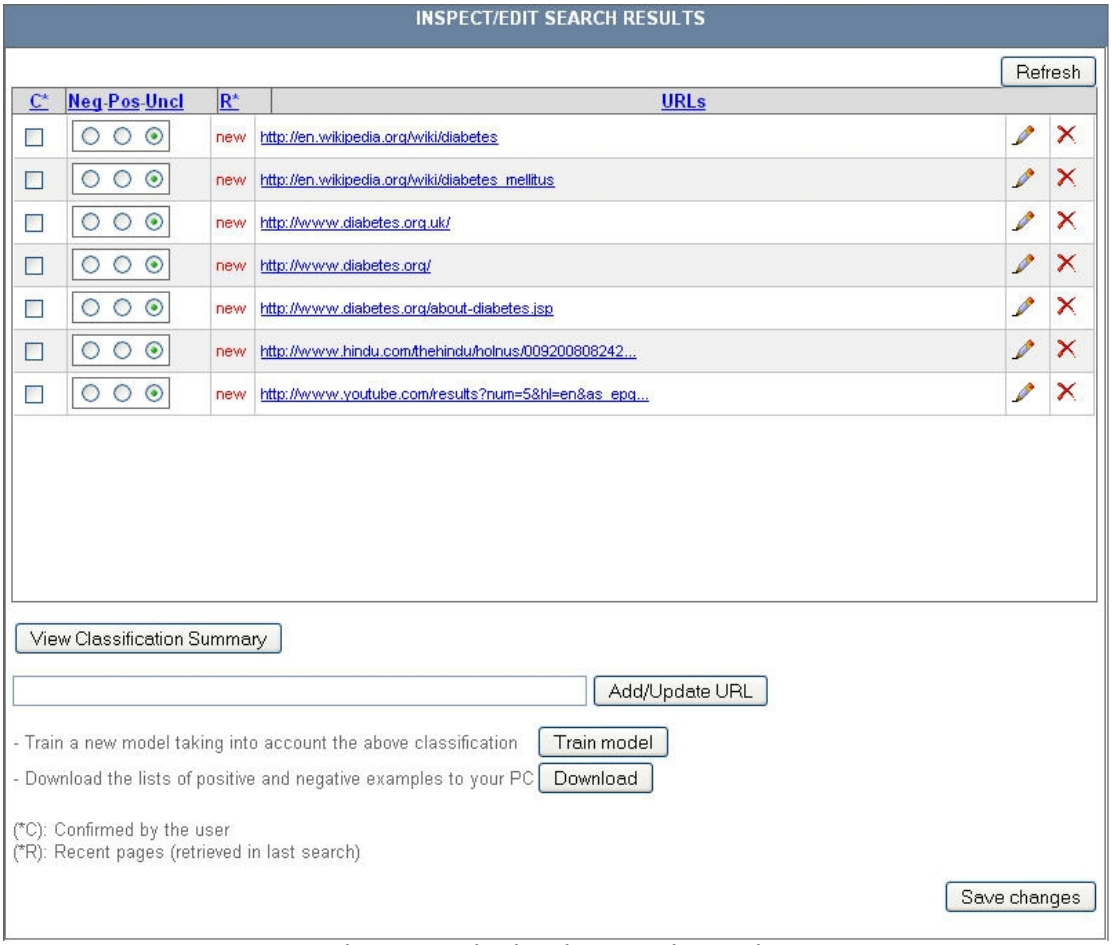

Figure 31. Viewing the Search Results

At this point, the user has the ability to decide whether this specific URL is relevant for his/her search or not. The user interface is designed to help in such classification: the user can use the radio buttons next to every URL; selecting the Pos (=Positive) value means that the corresponding URL is relevant while by selecting the Neg (=Negative) value signifies that it is not. Leaving the Uncl (=Unclassified) value selected means that the user is not certain about a URL's relevance or that the URL has not been checked yet.

Once the user has manually classified as Pos or Neg a good part of the first results (e.g. at least 20 URLs) returned by this search task, the user can train a new

classification module by clicking on the button “Train model”. This model will then be used for the automatic classification of the results returned in next search iterations of this same task (note that search tasks can run as many times as required). At the end of a search iteration, and if a trained classification model is available, the new results automatically get a classification score (pos, neg or uncl) decided by the model.

The user can continue by manually verifying the automatically classified URLs, as well as, by checking the unclassified ones and then re-train the classification model by pressing again the “Train model” button. Thus, the model becomes better and better after every new search iteration and every subsequent re-training. In that way, searching is expected to become more and more focused.

Finally, the user at any point can download the list of URLs being classified by pressing the “Download” button. To do so, the user has to specify (see Figure 32) the path in his/her local disk where the URL lists will be saved.

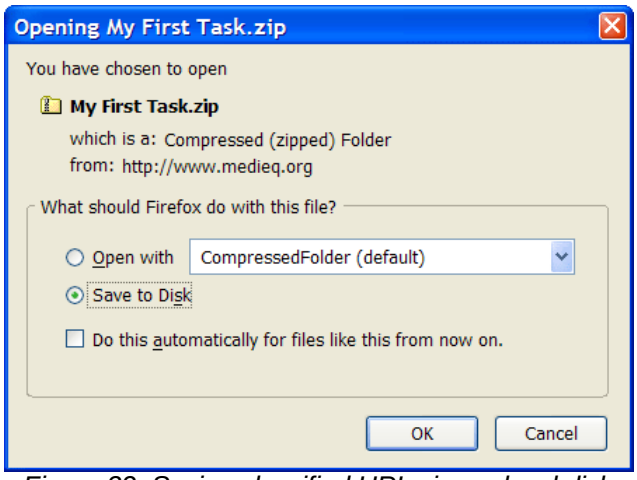

Figure 32. Saving classified URLs in my local disk

Appendix C. Sample Data for Testing AQUA

C1. Search Task Samples

The following table presents a set of keywords to be used for Search Tasks.

Topic	**Ischaemic/ Ischemic Heart Disease**
Keywords	• Myocardial infarction OR heart infarction OR hearth attack • (Acute) Coronary syndrome OR coronary artery disease • Chest pain • Coronary bypass
Topic	**Breast Cancer**
Keywords	• Breast cancer OR cancer of breast • Breast tumor • Treatment AND breast cancer • Mastectomy AND breast cancer • Mastectomy

C2. Review/Monitor Samples

The following table presents a set of URLs to be used for Review/Monitor Tasks.

URLs	
http://www.bacr.org.uk/	http://www.wrongdiagnosis.com/
http://www.jcmh.com/	http://www.drgreene.org/
http://www.alexandermd.com/	http://pathweb.uchc.edu/
http://www.skinovations.org/	http://www.aacap.org/
http://mcdhospital.org/	http://www.aafp.org/
http://www.spine-health.com/	http://www.helpguide.org/
http://www.medicinenet.com/	http://www.rcpsych.ac.uk/
http://www.ncemi.org/	http://www.xps.org/
http://www.podiatry.curtin.edu.au/	http://www.cdc.gov/
http://www.patient.co.uk/	http://www.eatright.org/
http://www.vasculitisfoundation.org/	http://gamma.wustl.edu/
http://www.clevelandclinic.org/	http://www.niams.nih.gov/
http://www.uhrad.com/	http://www.guideline.gov/
http://www.britishlivertrust.org.uk/	http://www.who.int/
http://www.allayurveda.com/	http://www.urologychannel.com/
http://www.curezone.com/	http://www.ncemi.org/
http://healthlink.mcw.edu/	http://www.nlm.nih.gov/

Appendix D. Brief Technical Details of AQUA

The main characteristics of AQUA implementation include: a) open architecture, b) accepted standards adopted in its design and deployment, c) character of large-scale, enterprise-level web application, and d) internationalization support.

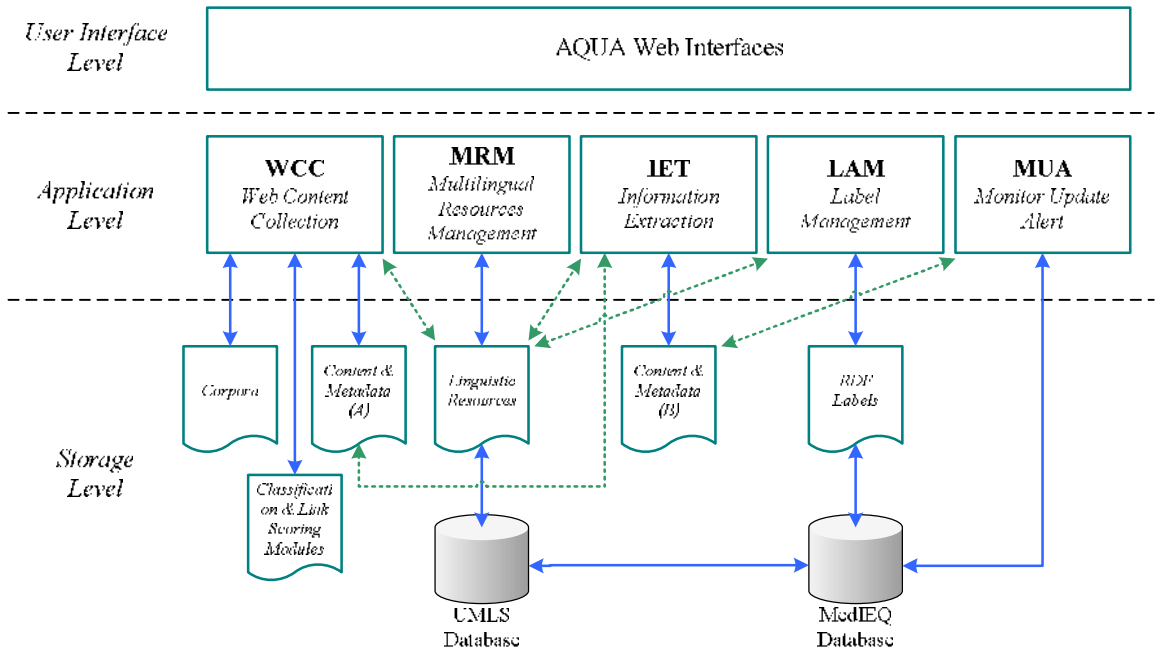

Figure 33. Architecture of the AQUA system

AQUA incorporates several subsystems (see the application level in Figure 33) and functionalities for the labeling expert:

- The *Web Content Collection* (WCC) component identifies, classifies and collects online content relative to the criteria proposed by the labeling agencies participating in the project.
- The *Information Extraction Toolkit* (IET) analyses the web content collected by WCC and extracts attributes for MedIEQ-compatible content labels.
- The *Label Management* (LAM) component generates, validates, modifies and compares the content labels based on the schema proposed by MedIEQ.
- The *Multilingual Resources Management* (MRM) subsystem gives access to health-related multilingual resources; input from such resources is needed in specific parts of the WCC, IET and LAM toolkits.
- Finally, the *Monitor-Update-Alert* (MUA) tool handles auxiliary but important jobs like the configuration of monitoring tasks, the MedIEQ database updates,

or the alerts to labeling experts when important differences occur during the monitoring of existing content labels.

Figure 33 shows all the possible data flows in AQUA (dashed arrows):

a) From WCC to IET: pages collected by WCC, once undergone a first-level extraction by WCC (extraction of metadata 1), are then forwarded to IET for further processing (extraction of metadata 2);

b) From IET to MUA: MUA takes all metadata collected by both WCC and IET and updates the MedIEQ database;

c) From MRM to WCC, IET, LAM: custom vocabularies generated by the MedIEQ users through MRM interface, can be accessed from other toolkits (WCC, IET, LAM), where the user may need them.

References

[1] Mayer MA, Leis A, Sarrias R, Ruíz P. Web Mèdica Acreditada Guidelines: realiability 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.

[2] Eysenbach G. Consumer health informatics. BMJ 320 (4) (2000), 1713-16.

[3] 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.

[4] Soualmia LF, Darmoni SJ, Douyère M, 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), 701-706.

[5] 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

[6] Kohler C, Darmoni SD, Mayer MA, Roth-Berghofer T, Fiene M, Eysenbach G. MedCIRCLE – The Collaboration for Internet Rating, Certification, Labeling, and Evaluation of Health Information. Technology and Health Care, Special Issue: Quality e-Health. Technol Health Care 10(6) (2002), 515.

[7] Curro V, Buonuomo 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.

[8] http:// www.w3.org/TR/rdf-schema/

[9] http://www.w3.org/ 2004/12/q/doc/content-labels-schema.htm