



---

## Evaluation Strategy

---

Distribution: Public

---

### MedIEQ

Quality Labeling of Medical Web content using Multilingual  
Information Extraction

National Centre for Scientific Research "Demokritos"  
Teknillinen Korkeakoulu – Helsinki University of Technology  
Universidad Nacional de Educacion a Distancia  
Col.legi Oficial de Metges de Barcelona  
Zentralstelle der deutschen Ärzteschaft zur Qualitätssicherung in der Medizin  
Vysoka Skola Ekonomicka V Praze  
I-Sieve Technologies Ltd

2005107 **D15**

February 2007

Project ref. no.	2005107
Project acronym	MedIEQ
Project full title	Quality Labeling of Medical Web content using Multilingual Information Extraction

Security (distribution level)	Public
Contractual date of delivery	31 December 2006
Actual date of delivery	16 February 2007
Deliverable number	D15
Deliverable name	Evaluation strategy
Type	Report
Status & version	Final
Number of pages	25
WP contributing to the deliverable	WP3
WP / Task responsible	AQUMED
Other contributors	WMA, NSCR, UNED, UEP, TKK
Author(s)	Dagmar Villarroel Gonzales (AQuMed) Miguel Angel Mayer (WMA) Konstatinos Stamatakis, Vangelis Karkaletsis (NSCR)
EC Project Officer	Artur Furtado
Keywords	evaluation, information retrieval, web content labelling, quality labelling, recall, precision,
Abstract (for dissemination)	<p>This document presents the evaluation strategy for the MedIEQ – System (Assisting Quality Assessment - AQUA). The evaluation consists of two parts. The first part reviews the results of the automated process in three major tasks, locating unlabeled web resources, labeling web resources, and monitoring labeled resources. In addition the usability of the user interface is evaluated.</p> <p>The second part evaluates AQUA in the context of a scenario that integrates AQUA within the usual work of a labeling agency.</p>

## Table of Contents

Executive summary .....	4
1. Introduction .....	5
2. General remarks about the quality labeling process .....	6
3. Evaluation of AQUA (first prototype).....	8
3.1 Objectives and general remarks.....	8
3.2. Evaluation I .....	9
3.2.1 Evaluation I - Identification of unlabeled medical web sites.....	9
3.2.2 Evaluation I - Accuracy of the automated labeling .....	12
3.2.3 Evaluation I - Accuracy of the automated monitoring and alerting system.....	15
3.2.4 Evaluation of the usability of the user interface .....	15
3.3 Evaluation II .....	16
4. Some remarks about the evaluation of the final prototype .....	19
Appendix I.....	20
Appendix II.....	21

## Executive summary

The MedIEQ project aims to develop technology to support labeling agencies in their work of quality labeling of medical web resources.

MedIEQ's integrated system AQUA (Assisting Quality Assessment) will support the following tasks of the quality labeling process:

- crawl the web to identify unlabeled medical websites,
- label the web resources according to a set of machine readable quality criteria and
- monitor labeled web resources, alerting the expert if changes occur against the labeling criteria.

It must be noted that AQUA provides tools for creating and maintaining machine readable labels (see Deliverable D5) which follow a common model (currently RDF-CL) and which may contain different vocabularies of labeling criteria (currently MedIEQ uses an initial vocabulary of 11 labeling criteria) (see Deliverable 4.1).

It is expected that these technologies will impact different areas of the labeling agencies' work, decreasing the manual effort and the time needed for labeling and monitoring medical information in internet, and increasing the effectiveness of the identification of unlabeled medical web sites.

At the time of writing this report, a first version of the labeling expert interface that supports the above three tasks as well as the labels' management, is available.

This deliverable presents the details of the evaluation strategy for AQUA first prototype. The evaluation strategy was designed based on the labeling processes of both participating agencies. It is composed of two parts. The first one (Evaluation I) concerns the separate evaluation of each of the three tasks that are part of the labeling process and the evaluation of the usability of the interface. The second part (Evaluation II) considers a scenario that proposes the initial integration of MedIEQ system into the daily work of the labeling expert.

## 1. Introduction

The MedIEQ's assisting system AQUA (Assisting Quality Assessment) was developed with the purpose of supporting the labeling expert in the review of health-related web resources. Since the labeling of such web resources is carried out mainly by hand, it is expected that by including AQUA in the labeling work the labeling effort per resource will decrease, increasing in this way the number of labeled resources.

The first prototype of AQUA will support the three principal tasks that are part of the labeling process: identification of unlabeled web resources, review of web resources and monitoring of already labeled resources; as well as the first set of quality criteria and two languages, English and Spanish.

This document presents the evaluation strategy for the AQUA first prototype. For the design of this strategy, AQUA characteristics (e.g. supported languages, quality criteria, etc.) as well as the current labeling process of both participating agencies were taken into account.

A brief description of WMA and AQUMED labeling processes, as well as some remarks regarding quality labeling of medical web resources are presented in section 2. The evaluation strategy is described in detail in section 3. It consists of two parts. The first part is the evaluation of each of the three tasks that are part of the labeling process. Additionally, an evaluation of the usability of AQUA interface is also described within part one. The second part of the evaluation will be conducted within a scenario that suggests the integration of AQUA in the daily work of the labeling expert. The performance of the expert using AQUA will be examined in terms of labeling effort and quality of the results. This part of the evaluation is based on a comparison of the expert's performance with and without AQUA support. Finally, section 4 presents few considerations about the evaluation of the final prototype.

## 2. General remarks about the quality labeling process

As described in Deliverables 4.1 and D5, there are two different approaches for the active labeling of medical information in internet: filtering portals and third party accreditations. Although these are different approaches with specific goals, they have some general tasks in common.

In general, the active labeling process comprises three tasks that are followed completely or partially by most labeling agencies:

- Identification of new web resources: this could happen in two different ways, either by active searching, usually in internet, or by voluntary application from the information provider, i.e. the web site's responsible asks actively for a review, usually in order to get a quality seal.
- Labeling of the web resources: this could be done with the purpose of awarding a quality 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 proof if a resource still deserves to be awarded a quality seal.

This is only a general description of the basic tasks of the active labeling process. Of course each agency may integrate other steps which are specific and necessary for its work.

Two labeling agencies are participating in MedIEQ: Web Mèdica Acreditada (WMA) and Agency for Quality in Medicine (AQUMED). These two agencies represent both approaches for labeling medical web sites, third party accreditation and filtering portals respectively.

Since the evaluation will be conducted by both agencies and the MedIEQ system will be integrated and adapted to the usual work of WMA and AQUMED, it is necessary to give a short explanation about the current accreditation process followed in each agency.

WMA<sup>1</sup> undertakes a third party accreditation of medical web sites. The complete process consists of four steps:

1. The person in charge of a website sends a (voluntary) request to WMA in order to initiate the process. Using the online application form, the person in charge provides certain information to WMA and has the chance to auto-check the WMA criteria, based on the Code of Conduct and the Ethical Code to express acceptance of these recommendations;
2. The WMA Standing Committee assesses the web site based on the WMA criteria;
3. WMA sends a report to the person in charge, who implements the recommendations;
4. When the recommendations are implemented, it is possible to obtain the seal of approval. In such a case, WMA sends an HTML seal code to be posted on the accredited website. In addition, WMA includes the site's name and URL to the index of accredited web sites and an RDF file is generated.

The labeled web sites are monitored periodically in order to identify if they still satisfy WMA criteria.

On the other hand, AQUMED<sup>2</sup> is a filtering portal. The indexing and labelling process consists of five steps:

---

<sup>1</sup> [http://wma.comb.es/home\\_eng.php](http://wma.comb.es/home_eng.php)

1. New medical patient information: there are two ways through which a new resource can be identified for indexing in AQUMED database. The first one is through internet search and the second one is through a direct request from the information provider. The web sites are selected according to general criteria: content, form and presentation should be serious, authorship, sponsorship and creation/update date should be clear, and only web sites without commercial interest should be indexed.
2. Web site classification: the new indexed web sites are classified in four groups: treatment information, background information, medical associations/scientific organisations and self-help/counselling organisations. Only the treatment information goes to the next step.
3. Evaluation: treatment information is evaluated with DISCERN and Check-In instruments. DISCERN is a known user guidance and Check-In was developed by AQUMED in collaboration with the "Patient Forum of the German Medical Association". Check-In is based on DISCERN and the AGREE instrument for critical evaluation of medical guidelines.
4. Confirmation: the database administrator has to confirm the result of the evaluation. It can be modified, erased or simply confirmed.
5. Feed back to the information provider: AQUMED sends an e-mail with the result of the evaluation in case of treatment information and with the information about the admission in AQUMED database in case of the other categories.

AQUMED database is periodically populated through new internet searches and it is examined for broken links as well. The evaluated web resources are also periodically re-reviewed in order to identify changes against the criteria or other updates.

---

<sup>2</sup> <http://www.patienten-information.de>

### 3. Evaluation of AQUA (first prototype)

#### 3.1 Objectives and general remarks

The general objectives of the evaluation strategy are:

- evaluate the performance of MedIEQ tools to each of the three major labeling tasks described in section 2,
- evaluate the usability of the user interface of AQUA,
- evaluate the labeling performance within a scenario that shows the first step of AQUA integration in the usual work of a labeling agency.

Taking into account these objectives, the evaluation strategy of the MedIEQ system comprises two parts:

- Evaluation I: it evaluates separately each of the tasks described in section 2.
- Evaluation II: it considers for the evaluation a scenario, where AQUA is partially integrated in the day-to-day work of the participating agencies.

The first prototype will support two languages: English and Spanish. So, evaluation of the 1<sup>st</sup> prototype will also be conducted in these languages, English by AQUMED and Spanish by WMA.

Apart from the evaluation of the impact of MedIEQ tools in the labeling performance, we also want to evaluate the following factors:

- Effective extraction from large collections of medical web content. In both parts of evaluation we experiment with a large amount of web resources with respect to the size of resources currently managed by the labeling agencies. During Evaluation II, the processing time required for labeling a specific number of resources with and without MedIEQ tools is calculated.
- Effort required to customize the system into new languages. As noted before, the 1<sup>st</sup> prototype will cover two of the seven languages of the project (English, Spanish). For each of the rest languages, to be covered by the 2<sup>nd</sup> prototype, we will measure the effort required to customize the tools to each language and localize the user interface.
- Implementation of an open architecture. This is a design requirement (see deliverable D12) which guides the AQUA implementation from its beginning. We will evaluate this in practice when new techniques or language-specific modules will be integrated in the system.

In the following sub-sections, the two parts of the evaluation strategy are described in detail.

### 3.2. Evaluation I

The goal of the part of evaluation (Evaluation I) is to measure the performance of MedIEQ system (AQUA) in each of the three major labeling tasks described in section 2. In addition, the usability of the AQUA interface will be examined. More specifically, Evaluation I consists of the four following tasks:

1. Evaluation of the location and identification of unlabeled medical web sites.
2. Evaluation of the accuracy of the automated labeling.
3. Evaluation of the accuracy of the automated monitoring and alerting system.
4. Evaluation of the usability of the user interface

In order to measure AQUA performance (first three tasks), standard evaluation measures like recall and precision will be used:

- Recall: defined as the ratio of correctly extracted items to the total number of items that are expected to be extracted by the labeling expert (gold standard).
- Precision: defined as the ratio of correctly extracted items to the total items extracted by the system.

#### 3.2.1 Evaluation I - Identification of unlabeled medical web sites

As mentioned in section 2, in contrast to AQUMED, WMA receives voluntary requests from the information provider. For this reason, this evaluation task is relevant only for AQUMED and therefore will be conducted only in English.

Evaluating the identification of unlabeled medical web sites aims to examine the performance of the automated search with the focused crawler of AQUA, in terms of precision. Hence, we will perform a manual inspection of a subset of the crawler output in order to examine their relevance.

For the definition of relevance criteria, we have drawn upon the experience of previous works<sup>3 4</sup> adopting the following criteria:

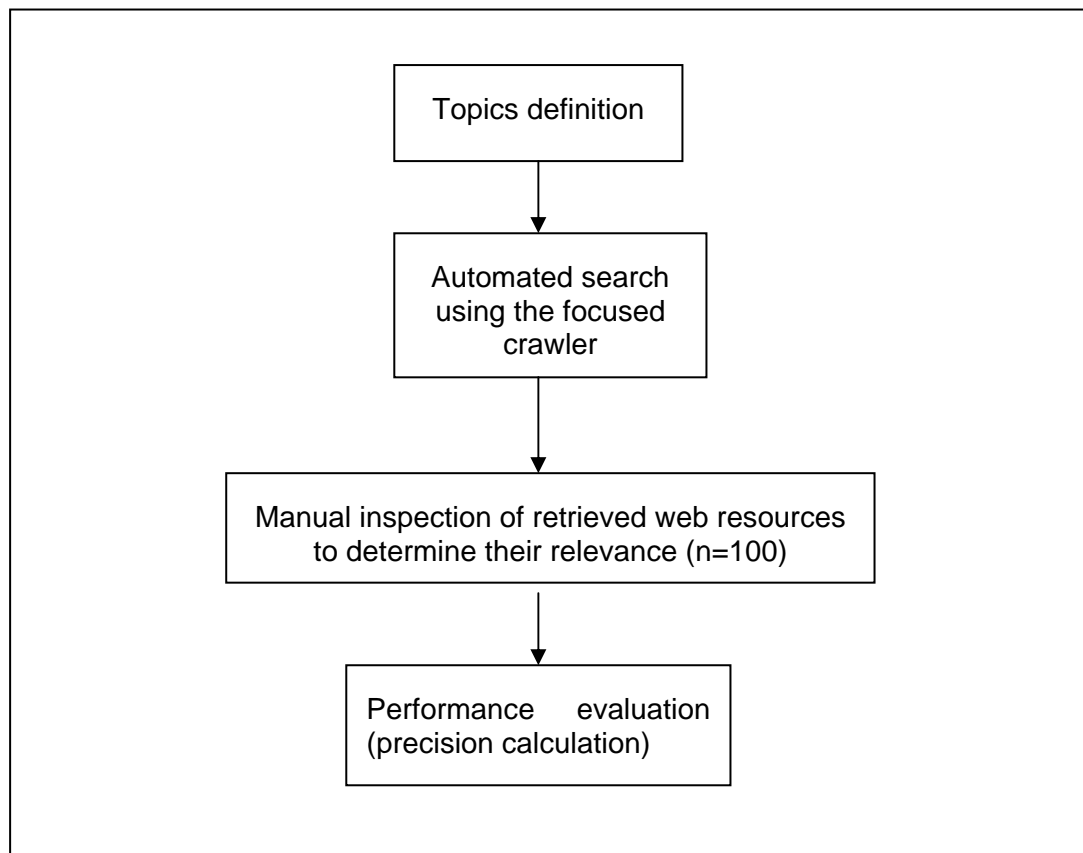
- Relevant: the website deals about the query or it consists of a collection of links about the topic of the query (only if after inspection of two links they are considered relevant)
- Irrelevant: the ones which do not deal about the topic of the query, doublets and inactive links.

---

<sup>3</sup> Clarke S.J., Willett P. Estimating the recall performance of Web search engines. *Aslib Proceedings*, 1997; 49(7):184-189

<sup>4</sup> Olvera M.D. Rendimiento de los sistemas de recuperación de información en la World Wide Web: revisión metodológica. *Rev. Esp. Cient.*, 200; 23(1):63-77

The evaluation of the location of unlabeled medical web sites consists of four steps as illustrated in Figure 1.



**Figure 1.** Evaluation of the identification of unlabeled medical websites

1. Definition of the search topics: the focused crawler supports two types of start points, keywords and web directories (see Deliverables D6, D7.1). The start point is determined by the user. For the evaluation, the search will be conducted using both start points. Regarding the topic selection, some surveys showed that the most searched topics are common medical conditions<sup>5</sup> like cancer, heart diseases and children's health<sup>6</sup>. Based on this, we selected the two most common search topics:
  - a. Ischaemic/ischemic heart disease, which is also the most common cause of death in Europe and
  - b. Cancer; because this theme is very extensive, we decided to limit it to breast cancer. It has also an epidemiological meaning as the most common cancer in women population.

For the search, several keywords as well as web directories for each of the selected topics are proposed, a list of which is available in Appendix I. We believe that medical

<sup>5</sup> Diaz J.A., Griffith R.A., et al. Patient's Use of the Internet for Medical Information. *J Gen Intern Med* 2002;17:180-185

<sup>6</sup> Cline R.J.W., Haynes K.M. Consumers health information seeking on the internet: the state of the art. *Health Edu Res* 2001; 16(6):671 - 692

information for patient should use an easily understandable vocabulary. Nevertheless, medical terms need to be explained also, in order for patients to know the correct denominations. This is a prerequisite for a truly supportive information source. Taking this into account, the keywords were selected, in the majority of cases, with help of MESH terms. This approach allowed us to find medical resources that are addressed at professionals as well as a large number of resources addressed to patients.

2. Automated search for unlabeled medical resources using the crawler for the specific topic definition.
3. Manual inspection of the retrieved resources in order to determine their relevance according to the relevance criteria mentioned above. As the output is usually big, only the first 100 retrieved resources will be manually reviewed.
4. The performance of the crawler will be measured in terms of precision (ratio of relevant web sites returned by the crawler to all sites returned by the crawler)

Since the specific evaluation task concerns only the AQUMED labeling process, it is interesting for AQUMED to know the performance of the crawler when AQUMED inclusion and exclusion criteria<sup>7</sup> are used in addition to the general criteria described above. According to these criteria:

- Relevant are considered the Web sites that belong to one of the following categories:
  - Self-help groups
  - Patient organization
  - Medical / professional associations
  - Governmental organizations
  - Information websites e.g. about treatment or diagnostic procedures or background information
  - Personal reports or patients experiences
  - Link collections
- Irrelevant are considered the following types of Websites:
  - Clinics or medical practices (if they do not offer medical information),
  - Health care insurance companies (if they do not offer medical information),
  - Laboratory and pharmaceutical companies,
- Information about alternative or complementary therapies and
- Chat rooms or bulleting board postings

In order to evaluate the system performance in this special case, we will follow the same procedure described above, i.e. performing a manual inspection of a subset of the crawler output.

---

<sup>7</sup> <http://www.patienten-information.de/content/informationsqualitaet/auswahl>

### 3.2.2 Evaluation I - Accuracy of the automated labeling

The second task of Evaluation I, aims to examine the accuracy of the automated labeling with MedIEQ system.

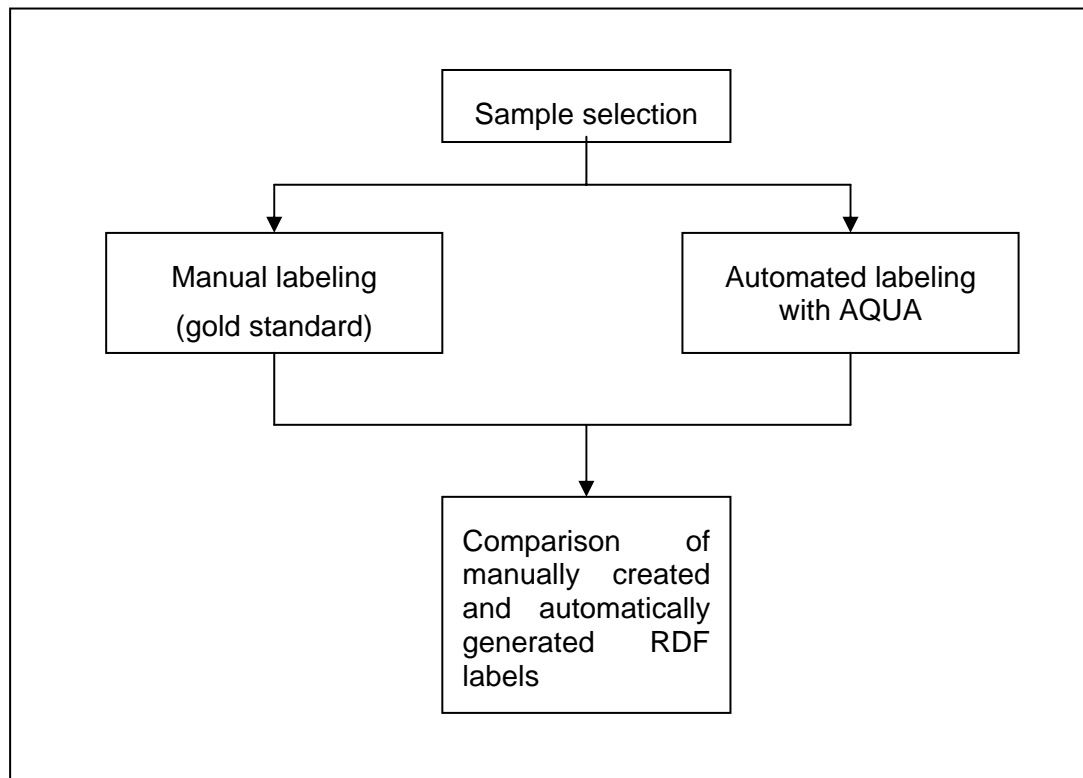
AQUA first prototype will support the first set of criteria described in Deliverable D4.1. A list of the criteria is shown in table 1.

Criteria		Value		
1	Title	String		
2.	Resource URI	URI		
3.	Responsible / Author	String		
4.	Contact details of the responsible / Author	E-mail address Postal address Telephone number		
5.	Last update	Date		
6.	Topic / Keywords	MESH – UMSL		
7.	Resource language(s)	String for language and language code		
8.	Target audience	Groups: - Professional - Non professional: adult, children		
9.	Advertisement	Boolean		
10.	Quality seal or third party program	<table border="1"> <tr> <td> Seals:  - WMA  - HON Code  - pWMC  - URAC  - Health TRUSTe  - Afgis </td> <td> Virtual seals:  - AQuMed  - OMNI  - WHO Vaccines Safety </td> </tr> </table>	Seals: - WMA - HON Code - pWMC - URAC - Health TRUSTe - Afgis	Virtual seals: - AQuMed - OMNI - WHO Vaccines Safety
Seals: - WMA - HON Code - pWMC - URAC - Health TRUSTe - Afgis	Virtual seals: - AQuMed - OMNI - WHO Vaccines Safety			
11.	Virtual consultation	Boolean		

**Table 1.** 1<sup>st</sup> version of the MedIEQ vocabulary of quality criteria

For the evaluation of the accuracy of the automated labeling, a sample of unlabeled websites will be labeled by AQUA. These automatically created RDF labels will be compared with manually created (i.e. by labeling experts) RDF labels (gold standard). This approach will allow us to calculate the evaluation measures of recall and precision.

The second task of Evaluation I will be conducted in English and Spanish by AQUMED and WMA respectively.



**Figure 2.** Evaluation of the accuracy of the automated labeling

The following steps describe in detail the evaluation of the automated labeling (see figure 2):

1. **Sample selection:** we will collect a sample of unlabeled websites in both languages (n=100). The web site selection should consider the special features and requirements of each participating agency, e.g. already available resources, internal classification system etc. Therefore the samples in Spanish and English will have different characteristics.
  - a. **WMA sample (Spanish):** for the sample selection we take into account that WMA has approximately 500 web sites in Spanish in its database; furthermore it receives new applications constantly. The Spanish sample (n=100) will consist of 50% of web sites that are saved in WMA database and have already obtained WMA label or are in process to get it, and the other 50% will be new applications received during the period between January – April 2007. In case that the new applications are not enough, the sample will be completed with websites from the database.
  - b. **AQUMED sample (English):** this sample (n=100) has to be completely collected because AQUMED only reviews resources in German. For the web site selection, we will take the following features into account:
    - i. AQUMED internal terms<sup>7</sup>

- ii. Although AQUMED reviews only websites which address the lay population, and especially patients, all target groups have to be represented for the evaluation. In order to meet this requirement, a half of the sample will address patients (non-professionals) and the other half professionals
- iii. AQUMED focuses on information about medical treatments or diagnostic procedures. For this reason, only resources with this kind of information will be considered for the evaluation.
- iv. The first prototype will support only information in .html format.

The sample will be collected from following databases and general search machines:

- Organising Medical Networked Information (OMNI)<sup>8</sup> (for professionals)
- Database from Karolinska Institute Library, Stockholm, Sweden<sup>9</sup>, (for professionals and patients)
- Database of US Department of Health and Human Services<sup>10</sup> (for patients)
- NOAH: New York Online Access to Health<sup>11</sup> (for patients)
- Consumer and Patients Health Information Section (CAPHIS)<sup>12</sup> (for patients)
- Google, Yahoo and Altavista: some resources have to be collected using general search engines, in order to get also websites that are not evaluated or indexed in a portal.

2. Automated review of the sample using AQUA
3. Manual review of the sample (gold standard creation): there are two different approaches for this step that result from the differences between the samples. In case of WMA sample (Spanish), we already have several of the RDF labels (at least the 50%), because the sample consist of web sites, that were already labeled. These will form the gold standard. In the case of AQUMED sample (English), this has to be labeled in order to get the gold standard. The selected sample will be reviewed by two experts independently. Possible divergences between the experts will be unified after a discussion. The results of the manual labeling and the post-processing will form the gold standard.
4. Finally the automatically generated RDF labels will be compared with the gold standard. Standard evaluation scores, such as recall and precision will be calculated.

---

<sup>8</sup> <http://www.intute.ac.uk/healthandlifesciences/medicine/>

<sup>9</sup> <http://micf.mic.ki.se/Diseases/>

<sup>10</sup> <http://www.healthfinder.gov/>

<sup>11</sup> <http://www.noah-health.org/>

<sup>12</sup> <http://caphis.mlanet.org/consumer/index.html>

### 3.2.3 Evaluation I - Accuracy of the automated monitoring and alerting system

The third task in Evaluation I will measure the system's performance when monitoring already labeled websites. The same samples that were reviewed during the second task of Evaluation I (Evaluation of the accuracy of automated labeling) will be re-reviewed (monitored) nine months, at the latest, after the first review. The second review will focus on the identification of changes or updates according to the labeling criteria.

In a similar way to the process followed in the 2<sup>nd</sup> task of Evaluation I, the websites sample will be reviewed by two independent experts and by AQUA. After the labeling experts discuss the discrepancies and unify them, the results of the manual and the automated re-review will be compared. The gold standard is the manual review. Calculation of evaluation scores such as recall and precision will be performed.

The evaluation of monitoring performance will be conducted in English by AQUMED and in Spanish by WMA.

### 3.2.4 Evaluation of the usability of the user interface

Since AQUA will support the work of the labeling expert, it is necessary that labeling experts evaluate the usability of the system. In general, the user of AQUA (and also a candidate to participate in the usability evaluation) is a person with experience in quality labeling of medical web sites, who dominates at least one of the supported languages (English and Spanish), and who didn't have any previous knowledge of AQUA use.

For the evaluation, the users will be asked to carry out the five different tasks described in Appendix II. After that they will have to fill a questionnaire that was designed to collect information about their experiences with AQUA, possible updates, recommendations, etc. (see the tasks description and the questionnaire in Appendix II).

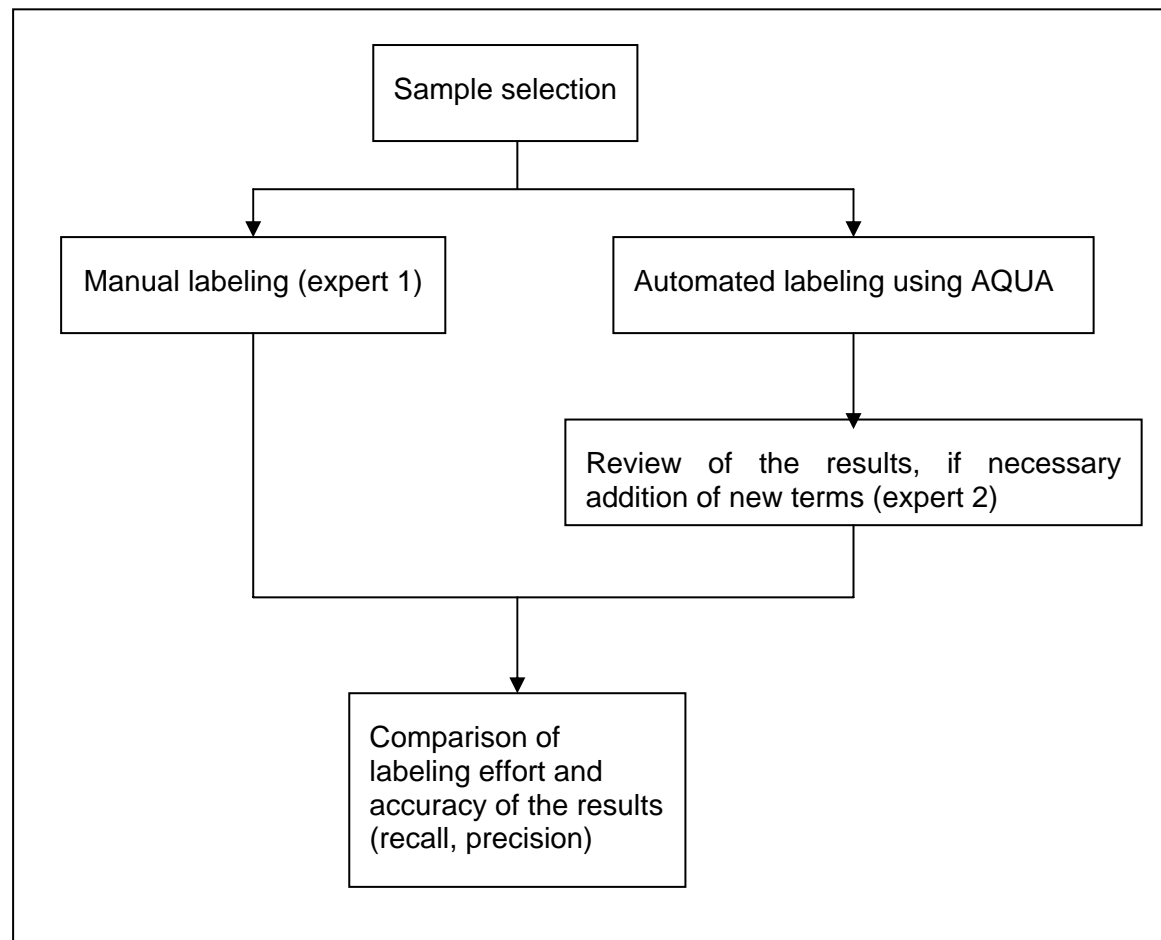
### 3.3 Evaluation II

As it was mentioned in the introduction, the main goal of MedIEQ is to develop a system that supports the work of the labeling expert, decreasing the manual effort. Therefore, it is expected that AQUA will be incorporated within the usual work of the labeling agencies.

Evaluation I measures the performance of AQUA (3 first tasks) and the usability of the interface. However, this type of evaluation cannot measure the effect of MedIEQ technology in the daily work of the labeling expert. For this reason, it is necessary to design a second evaluation strategy that considers a scenario where AQUA becomes a part of the labeling expert's work, in order to examine how the use of AQUA affects the expert's performance.

Evaluation II involves a comparison between an expert that works using MedIEQ tools and another expert, who makes his work only manually (see figure 3). For the comparison, the following aspects will be taken into account:

- Processing time: one challenge of the project is to provide tools with a high performance in terms of precision, without increasing too much the processing time. For this reason, the processing time required to extract the data will also be measured.
- Labeling effort: this indicator allows us to acquire information about the influence of AQUA in the work of the expert in terms of time. Labeling effort is defined as the time required to label a web site or a group of web sites. It will be measured by comparing the time required to label a group of web sites with and without AQUA support.
- Precision of the extracted data: this indicator gives us information about the influence of MedIEQ tools in the work of the expert in terms of quality. It will be measured by comparing the extracted data of the expert using AQUA to the extracted data of the expert, who works only manually.



**Figure 3.** Evaluation II steps

Evaluation II will be carried out in five steps:

1. Sample selection: the requirements and internal specifications of each agency will be considered for the selection of the web sites. Sample size for each language = 100
  - a. WMA sample (Spanish): the Spanish sample for Evaluation II will be collected from WMA database, i.e. it consists of web sites which already got WMA trust mark or are in the process to get it.
  - b. AQUMED sample (English): the sample collection will follow the same requirements described in section 3.2.2.
2. Manual labeling of the selected sample conducted by expert 1.
3. Automated labeling of the selected sample using AQUA.
4. The results of the automated labeling will be thoroughly proofed by expert 2. Every extracted term will be approved or discarded. Non extracted terms should be added, if necessary.
5. Comparison and calculation of the indicators mentioned above.

The evaluation will be conducted in Spanish and in English, by WMA and AQUMED respectively.

In order to increase the representativeness of the evaluation results, we will examine the possibility to perform Evaluation II in two rounds. The first round will be conducted as described above; at the second round the experts will change the roles, i.e. the one who labeled with support of AQUA at round one will label only manually at the second round and vice versa. The decision to perform the second round will be taken during the evaluation, because it depends on the obtained results, the available time and resources as well as the necessary effort.

## 4. Some remarks about the evaluation of the final prototype

The final prototype will support, additionally to English and Spanish, the other five languages of the project: Czech, German, Greek, Finnish and Catalan, as well as other types of resources (.pdf and .doc).

All these new characteristics should be taken into account in the selection of the evaluation samples and the final definition of the evaluation strategy.

The evaluation of the final prototype will be based on the presented strategy which may be modified according to the experiences with the evaluation of the first prototype and new requirements that may arise.

For the execution of the evaluation in all the languages it should be considered the involvement of other labeling experts in order to obtain adequate standards for measuring the performance of AQUA in all languages.

## Appendix I

### List of the proposed input for the crawler

#### 1. Ischaemic/ischemic heart disease

##### a. Keywords

Myocardial infarction – heart infarction – hearth attack

(Acute) Coronary syndrome – coronary artery disease

Angina pectoris – chest pain

Myocardial revascularization – coronary bypass

##### b. Web directories

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cardiovascular\\_Disorders/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cardiovascular_Disorders/)

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cardiovascular\\_Disorders/Heart\\_Disease/Angina\\_Pectoris/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cardiovascular_Disorders/Heart_Disease/Angina_Pectoris/)

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cardiovascular\\_Disorders/Heart\\_Disease/Heart\\_Attack/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cardiovascular_Disorders/Heart_Disease/Heart_Attack/)

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cardiovascular\\_Disorders/Heart\\_Disease/Coronary\\_Artery\\_Disease/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cardiovascular_Disorders/Heart_Disease/Coronary_Artery_Disease/)

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Heart\\_Attack/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Heart_Attack/)

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Angina\\_Pectoris/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Angina_Pectoris/)

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Coronary\\_Artery\\_Disease/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Coronary_Artery_Disease/)

#### 2. Breast Cancer

##### a. Keywords

Breast cancer – cancer of breast

Neoplasm breast – breast tumor

Mammary carcinoma – mammary neoplasm

Carcinoma, ductal, breast

Phyllodes tumor

Treatment – breast cancer

Lumpectomy – segmetal mastectomy

Mastectomy

Breast conserving surgery

Radical mastectomy

##### b. Web directories

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Breast\\_Cancer/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Breast_Cancer/)

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Breast\\_Cancer/Mastectomy/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Breast_Cancer/Mastectomy/)

[http://uk.dir.yahoo.com/Health/Diseases\\_and\\_Conditions/Cancers/](http://uk.dir.yahoo.com/Health/Diseases_and_Conditions/Cancers/)

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cancer/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cancer/)

[http://directory.google.com/Top/Health/Conditions\\_and\\_Diseases/Cancer/Breast/](http://directory.google.com/Top/Health/Conditions_and_Diseases/Cancer/Breast/)

[http://directory.google.com/Top/Health/Women%27s\\_Health/Conditions\\_and\\_Diseases/Breast/](http://directory.google.com/Top/Health/Women%27s_Health/Conditions_and_Diseases/Breast/)

## Appendix II

### Usability evaluation of AQUA

AQUA is accessible from any web browser. AQUA's home page is at:

<http://www.medieq.org/aqua/aqua/welcome.seam>.

Please follow the instructions in order to perform the evaluation tasks. After having completed all tasks, you are kindly requested to fill the questionnaire that follows.

#### TASK 1: MANUAL LABELLING or reviewing web sites using AQUA

- Click on "Label management" from the menu on your left.
- Follow the link "Create manually a new RDF label".

#### TASK 2: COMPUTER-ASSISTED LABELLING – (A)

- Click on "Task management" from the menu on your left.
- Select three (3) web sites of the following list and ask AQUA to review them.
  - <http://www.spine-health.com/>
  - <http://health.yahoo.com/ency/healthwise/te6320spec>
  - <http://www.medicinenet.com/tuberculosis/article.htm>
  - <http://www.patient.co.uk/showdoc/23068792>
  - <http://drgreene.mediwire.com/main/Default.aspx?P=Content&ArticleID=128561>
  - <http://www.ncemi.org/cse/cse0805.htm>
  - <http://www.podiatry.curtin.edu.au/encyclopedia/gangrene/gangrene4.html>
  - [http://www.who.int/reproductive-health/impac/Procedures/Correcting\\_P91\\_P94.html](http://www.who.int/reproductive-health/impac/Procedures/Correcting_P91_P94.html)
  - <http://www.alz.org/>
  - <http://www.womenshealth.gov/faq/stdgonor.htm>

#### TASK 3: COMPUTER-ASSISTED LABELLING – (B)

Click on "Label management" and try to locate and edit one of the labels you created, e.g. verify existing data and modify some of them (e.g. the resource's last update).

#### TASK 4: MONITORING

- Click on "Task management".
- Create a new monitoring task and define the characteristics (name, schedule, etc) according to your preferences.
- Add the web sites you labelled within task 2 to this new monitoring task.

#### TASK 5: ALERTING

- Click on "Alert management" and try to create a new alerting profile.

## Questionnaire for evaluating the usability of AQUA

Below is the questionnaire for the scheduled (as described in this document) usability evaluation of AQUA. This is the currently suggested version of the usability questionnaire; it is possible that, by the time of the evaluation, some of its questions may have changed.

### **Part I: About your health quality expertise (questions 1-4)**

1. Which of the following describes you best?
  - a. Medical doctor
  - b. Medical student (postgraduate)
  - c. Medical student (undergraduate)
  - d. I have no medical expertise
  
2. Are you now or were you in the past employed as a quality expert for health-related web content by a quality labelling agency (LA)?
  - a. I am currently employed by a LA
  - b. I was employed in the past by a LA
  - c. I have no experience as a quality expert

*(In case of a. or b. in 2)*

3. How many (in average) websites or online documents do you (or did you) review per month?
  - a. 1 - 5
  - b. 6 - 15
  - c. more than 16

*(In case of a. or b. in 2)*

4. In which languages are the websites or the online documents that you are (or were) given to review?
  - a. En
  - b. Es
  - c. De
  - d. Gr
  - e. Fi
  - f. Cz
  - g. Ca
  - h. Fr

Is any of the above your mother tongue? If yes, please specify .....



**Part III: General remarks on AQUA usability (10-16)**

10. On a scale from 1 to 5, where 1 is “very useful” and 5 “completely not useful”, rate the idea of AQUA alerts.

1	2	3	4	5	Don't know
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. On a scale from 1 to 5, where 1 is “very easy” and 5 “very difficult”, rate how easy was for you to use AQUA (navigation menus, functionalities, etc.).

1	2	3	4	5	Don't know
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

12. On a scale from 1 to 5, where 1 is “very clear” and 5 “very confusing”, rate the language and terms used in AQUA.

1	2	3	4	5	Don't know
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13. On a scale from 1 to 5, where 1 is “excellent” and 5 “poor”, rate your overall impression from AQUA after having used it.

1	2	3	4	5	Don't know
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

14. Do you think there of some other functionality that should be included in AQUA?

<input type="checkbox"/> yes	<input type="checkbox"/> no
------------------------------	-----------------------------

If yes, please specify:

---



---



---



---

15. Do you think there of some change in the AQUA user interfaces that would improve their usability?

<input type="checkbox"/> yes	<input type="checkbox"/> no
------------------------------	-----------------------------

If yes, please specify:

---



---



---



---

16. Is there anything else you would like to add or comment?

---

---

---

---

**Thank you very much!!**