



European digital content for the global networks

EDC-22071 KOM2002/27938

## Groupware for Distributed Content Production

KOM2002

### First Evaluation Report

<b>Security (distribution level)</b>	Public
<b>Contractual date of delivery</b>	31.10.2003
<b>Actual date of delivery</b>	18.11.2003
<b>Deliverable number</b>	D 7.2
<b>Deliverable name</b>	First Evaluation Report
<b>Type</b>	R
<b>Status &amp; version</b>	Final report
<b>Number of pages</b>	
<b>WP contributing to the deliverable</b>	WP7
<b>WP / Task responsible</b>	University of Ioannina
<b>Other contributors</b>	
<b>Author(s)</b>	Evagelia Gerasi; Petros Skapinakis; Venetsanos Mavreas
<b>EC Project Officer</b>	Kimmo Rossi
<b>Keywords</b>	Content; mental health; quality of information; discern
<b>Abstract (for dissemination)</b>	This report aims to show the first results of the evaluation which has started after the site has gone public and will end with the termination of the project. It reports the progress made on each separate domain of this evaluation (as was planned in deliverable D7.1), tries to analyse the data collected and to explain why in some domains we had no adequate data to proceed.

# 1. Table of contents

<b>1. Table of contents.....</b>	<b>2</b>
<b>2. Executive summary .....</b>	<b>4</b>
<b>3. Introduction – Background.....</b>	<b>5</b>
<b>4. Evaluation of quality .....</b>	<b>8</b>
According to the Deliverable D7.1, evaluating the project results requires acting in two different directions: First, to evaluate the usefulness and convenience of the groupware used in order to develop the site and second, to evaluate the quality and value of the web site developed.....	8
<b>4.1 EVALUATION OF THE GROUPWARE .....</b>	<b>8</b>
In order to collect the views of all partners for the usefulness of the groupware in a representative way, we have developed and circulated to them via e-mail a simple questionnaire, a copy of which is presented below, followed by the answers of the partners after each question.....	8
<b>Before the project started did you have any previous experience of working with computers and for how long?.....</b>	<b>8</b>
<b>Did you understand quickly and easily how the groupware works?.....</b>	<b>9</b>
<b>Did you find the training provided adequate to use the groupware? Was the technical department helpful enough? .....</b>	<b>9</b>
<b>Do you regularly use the groupware for writing new FAQs?.....</b>	<b>9</b>
<b>How well does the system support the writing of new FAQs? .....</b>	<b>10</b>
<b>What do you think is unnecessarily complex in the support for writing new FAQs? .....</b>	<b>10</b>
<b>Do you regularly use the groupware for answering the ask-the-expert questions and how well do you think the system supports this function? .....</b>	<b>11</b>
<b>What do you think is unnecessarily complex in the support for answering the ask-the-expert questions?.....</b>	<b>11</b>
<b>Have you used the groupware for reviewing FAQs written by other partners or modifying your own FAQs? How well do you think the system supports this function? .....</b>	<b>11</b>
<b>Do you regularly use the “News” page? Do you think it is useful? Do you think it should be improved? .....</b>	<b>12</b>
<b>Have you used the support for sending notifications by e-mail? How well do you think this function works? .....</b>	<b>12</b>
<b>Please give your opinion on the translation system inside the groupware. ....</b>	<b>13</b>

<b>Do you believe machine translation has reached a point where it could be useful to similar projects? .....</b>	<b>14</b>
<b>Would you like some additional personal settings to adjust the groupware to your needs?.....</b>	<b>15</b>
<b>How complicated did you find the groupware to be, compared to working by meetings? Please state below 3 advantages and 3 disadvantages of this way of international collaboration.....</b>	<b>16</b>
<b>Did you find that the chat sessions between the partners inside the groupware worked for the benefit of the project and gave solutions to problems? .....</b>	<b>17</b>
<b>How often was the intervention of the technical department required while you were using the groupware? .....</b>	<b>17</b>
<b>When problems occurred how quick was the response of the technical department? .....</b>	<b>17</b>
<b>What was your overall opinion of the groupware and its contribution to the project? ..</b>	<b>17</b>
<b>What would you change in the groupware for it to become “very good”? .....</b>	<b>18</b>
4.2 EVALUATION OF THE WEB CONTENT .....	19
a. External evaluation: .....	19
b. Evaluation with the help of control users: .....	20
c. Evaluation with the help of real users (ongoing evaluation):.....	21
4.3 EVALUATION OF THE SITE IN OTHER LANGUAGES .....	22
<b>5. Evaluation of the usage .....</b>	<b>23</b>
5.1 Methodology .....	23
<b>6. Conclusions and Timeplan .....</b>	<b>25</b>
<b>Medcircle <a href="http://www.medcircle.org">www.medcircle.org</a> .....</b>	<b>26</b>
<b>APPENDIX.....</b>	<b>27</b>
PART I:.....	27
PART II: .....	31
PART III: .....	37
Test of the natural-language question-answering system.....	37
User satisfaction questionnaire.....	38
Controlled user testing.....	38

## 2. Executive summary

This report, which is part of the seventh work package, is mostly based on the previous deliverable D7.1 (Evaluation Plan) and aims to show the first results of the ongoing evaluation which has started after the site became public and will end with the termination of the project. It is a report of the progress achieved so far in each separate domain of the evaluation plan.

Until now the main domains of the evaluation where we had exploitable results were the evaluation of the groupware and the evaluation of the usage. We are still expecting results from the questionnaire we had inserted into the site for the users to fill in and also an external evaluation from independent experts based on the Discern instrument. Also desirable would be a separate evaluation of the site's different linguistic versions.

The results from the groupware evaluation show that the use of the specific software proved very useful to the purposes of the project despite some practical problems which appeared, especially during the acquaintance with the groupware.

From what we gather of the usage statistics our site seems to have far more visitors than expected. The impression we get from the same statistics is that a main reason for this activity must be the good quality of answers our site provides to common human problems (most popular pages were those about sexual disorders, depression, sleeping problems etc).

### 3. Introduction – Background

It is a common knowledge that the use of Internet has grown wide within the last years, especially in the European countries and the so-called western world. People of all ages and especially young children and adolescents appear to be searching for health information through the Internet daily.

Quite prevalent in this query are problems of mental health. It has been reported, for example, that more Internet users search the Web for information on depression than any other health condition. Furthermore a search in the “google” (www.google.com), one of the most popular search engines, has shown that the term depression yields 5.160.000 different pages, far more than other common conditions such as heart disease (2.910.000 pages), cholesterol (2.370.000 pages), headache (1.780.000) or influenza (870.000 pages).

The Wordtracker data base [11], which stores two million queries made to search engines by Internet users, registers 11354 queries with the term depression, far more than heart disease (2370 queries), cholesterol (4839 queries), headache (1164 queries) and influenza (865 queries).

This is not surprising given the high level of disability associated with depression in the community and the fact that the Web provides a convenient and anonymous means of obtaining information about mental health problems, which is very important considering the stigma mental illness carries in our society.

Given this information, in KOM2002 project we aimed to develop a website providing high quality information on various mental health problems. In particular, we aimed to produce information with the following characteristics:

- It should be accurate and evidence-based.

- It should be easily accessible from citizens across Europe and in particular in several languages.
- It should be easily searchable and relevant.
- It should give the user the opportunity to interact with specialists at his or her own language.

We have now reached a point, where our site has been created and gone public, filled with much information on mental health subjects for those interested. So now an evaluation is demanded.

Evaluation is a systematic investigation, a process (not a single act, but an ongoing procedure) which will help us estimate the usefulness of the project. Information will be gathered to determine what is working and why, and whether the project is meeting its goals and objectives according to the proposed time plan. This information mainly aims in constantly developing and improving the site. Evaluation in our case is important not only to the developers of this site but also and above all to the public which is in need of our services. Patients tend to seek for current and well-provided information and our aim is to fulfill these expectations.

Furthermore this project gave the opportunity to people from different countries and with different degrees of familiarity to the use of computers to work together on a specific groupware. Given that most of the partners were medical experts, with no previous knowledge of the groupware, an evaluation of the groupware and its usefulness will help us determine whether this or a similar software could be used in similar projects.

This report, which is part of the seventh work package, is mostly based on the previous deliverable D7.1 (Evaluation Plan) and aims to show the first results of the ongoing evaluation which has started after the site became public and will end with the termination of

the project. It is a report of the progress achieved so far in each separate domain of the evaluation plan.

## **4. Evaluation of quality**

According to the Deliverable D7.1, evaluating the project results requires acting in two different directions: First, to evaluate the usefulness and convenience of the groupware used in order to develop the site and second, to evaluate the quality and value of the web site developed.

### **4.1 EVALUATION OF THE GROUPWARE**

In order to collect the views of all partners for the usefulness of the groupware in a representative way, we have developed and circulated to them via e-mail a simple questionnaire, a copy of which is presented below, followed by the answers of the partners after each question.

#### **QUESTIONNAIRE APPLIED TO THE GROUPWARE USERS:**

**Before the project started did you have any previous experience of working with computers and for how long?**

Three of the seven users who answered the questionnaire had a previous experience from 20 to 40 years. The rest had a small experience with the simple use of the “word” and “office” utilities and Internet use.

**Did you understand quickly and easily how the groupware works?**

Four out of seven users answered yes. One out of seven answered no. Another thought that it was a bit difficult to enter the logic of the progress and one is still an apprentice.

**Did you find the training provided adequate to use the groupware? Was the technical department helpful enough?**

Three of the users entered the project at a later stage and did not have the opportunity to receive any special training. However all of them recognised the presence and help of the KTH staff.

**Do you regularly use the groupware for writing new FAQs?**

Six yes, one no.

**How well does the system support the writing of new FAQs?**

Well. There are different options for draft and the original. Sometimes I had minor problems with the German template of the groupware-system and posting new FAQs.(german partner)

Very well.(italian)

Well except that I would like to have a What You See Is What You Get (WYSIWYG)-editor integrated into the software better than Dreamweaver is integrated.(swedish-technical partner)

Well. When you've grasped the idea, it's not a complicated thing.(swedish apprentice)

Very well I think. I've only been here for a week so I could probably answer the question better in a week or two. (swedish apprentice)

I think it is most helpful. You can insert the text of the answer as you like it and modify it as you want it. It is directly exported and thus there is no need to send it by mail and someone else to put it in the website. This way it is far less time consuming.(greek partner)

**What do you think is unnecessarily complex in the support for writing new FAQs?**

Two users thought the problem is the use of HTML, two thought there is no problem, one thought Integration with Dreamweaver could be improved and one could not answer due to lack of experience.

**Do you regularly use the groupware for answering the ask-the-expert questions and how well do you think the system supports this function?**

Five of the users do not use this function. Of the two who use it one thinks it works well and the other uses a texteditor to minimize spelling errors.

**What do you think is unnecessarily complex in the support for answering the ask-the-expert questions?**

The one user does not have any complaints. The other (swedish-technical partner) thinks that there exists a problem with converting ask-the-expert answers to FAQs,

and that this is the same problem as with writing FAQs that there is no well-integrated WYSIWYG-editor. Also he thinks there should be integration with e-mail, so that the system can automatically send e-mails to an English expert with texts for language checking.

**Have you used the groupware for reviewing FAQs written by other partners or modifying your own FAQs? How well do you think the system supports this function?**

Six of the users regularly use it and think that it works very well. The last does not use it regularly.

**Do you regularly use the “News” page? Do you think it is useful? Do you think it should be improved?**

Three of the users do not use it at all. Two seldom use it because they find it difficult to use. The two left use it but find that there are too many unimportant information inside.

**Have you used the support for sending notifications by e-mail? How well do you think this function works?**

Four do not use it. Two of those who use it find it works well. The other always gets an error message, so it is impossible to use it.

**Please give your opinion on the translation system inside the groupware.**

The usability of the translation is weak, since the system causes HTML errors, so I have to clear the text and rewrite it. The English->German translation is poor, very very poor. Even the internal translation of single words of the webpages was not at all acceptable.

This is not a problem of terminology, but syntax and general expression or translation of every-day language. Usually the machine translation uses the least common term for translation. I think there was no proper adaptation of the dictionaries and software provided by Systran since I would expect a better adaptation to the vocabulary and use of common expression for this project. Rating on a scale 1-5 =2. (german partner)

The translation system is not working. It often is more amusing than correct!

(Italian partner)

More fun than really useful, but some people find it useful to get a draft when making a translation. (swedish-technical partner)

Have not used (netdoctor)

It works very well. (swedish apprentice)

The translation system is not so very good. It often is more amusing than correct!

(swedish apprentice)

I think the translation works as good as a machine translation could work. It is not very correct (especially syntactically) but one can have a general idea of the content of the FAQ. At least this is my opinion of the translation considering the greek language. (greek partner)

**Do you believe machine translation has reached a point where it could be useful to similar projects?**

No. Usually I like machine translation very much and use it for translation of webpages or technical information. But the use of machine translation for answers or forum discussion should not be recommended. (german partner)

Not as Systran works now. But Systran works well for translating office letters, so maybe Systran could be made to work well for our area. Note that most important for our area is not that Systran handles medical terminology well, but rather that Systran handles ever-day-life and human behaviour words like “body” or “gay” better. (swedish technical partner)

No. The automatic translations are of no use. The translations have to be done manually. (swedish apprentice)

No, and I doubt it will ever be. Language is a living organism which is changing all the time and a machine will never be able to translate sentences in such a way that for example irony or comforting will come through in a satisfying way. (swedish apprentice)

Well, I think it depends on the needs of the project. It can certainly be helpful to a project where syntactical translation is not so important. (greek partner)

No. (italian partner and netdoctor)

**Would you like some additional personal settings to adjust the groupware to your needs?**

Four answered no. Two were not able to answer the question. The rest made the following suggestions:

- Direct connection to word processor / spelling correction
- Allow forums and discussion for each FAQ to get user feedback
- Include a rating system for common treatment experiences (e.g. rate the use of Cipralopram for anxiety disorders)
- Customizing the news page to reduce clutter.

**How complicated did you find the groupware to be, compared to working by meetings? Please state below 3 advantages and 3 disadvantages of this way of international collaboration.**

Three were not able to answer this question. The others made the following remarks:

(german partner)

Advantages :

1. Possibility to use the groupware in different countries or places
2. Connect to other experts and write and discuss FAQs, texts, decisions...
3. No need to travel

Disadvantages

1. Not all partners are using the system to the same extend
2. Personal communication is lacking.
3. Problem of using a non-native language for communication

(swedish technical partner)

We have never really used the groupware as a replacement for meetings, except for the chat sessions. The chat sessions worked well, advantage is that you can have meetings more often and also that it is easy to misunderstand each other in spoken language using a language which you are not perfect in.

(swedish apprentice)

Working by meetings? Well, it's comfortable to work by a computer. It's interesting to hear the opinions of people in other cultural settings. But you cannot talk face to face. You cannot expect the FAQs to consider your own specific cultural demands. And... sometimes you cannot trust your computer to function perfectly...

(greek partner)

It is certainly very easy and useful for everyone to work in their own home or office whenever they can or want. Furthermore the groupware gives the chance to people even from different countries to interact. However it can not replace meetings completely because problems occur and decisions have to be made which cannot always be done by e-mails. It is also easier for people to cooperate when they have met each other. Also, for many people communication via e-mail is not their favourite way of cooperation.

**Did you find that the chat sessions between the partners inside the groupware worked for the benefit of the project and gave solutions to problems?**

Four have not participated in such sessions, but one thinks it seems to be a good way of solving problems which appear suddenly without having to meet. Those who have used it think they work quite well.

**How often was the intervention of the technical department required while you were using the groupware?**

Four were not able to answer this question.

The german partner reports that though he had several contacts for the improvement of the groupware, the general function was very good.

The swedish technical partner reports such a function usually once every two weeks.

The greek partner reports such interventions needed mostly to correct wrong moves or problems occurring with the greek language.

**When problems occurred how quick was the response of the technical department?**

Five were not able to answer. Two reported a very quick response.

**What was your overall opinion of the groupware and its contribution to the project?**

All the users claim it was good for the project, though some difficulties in getting used to it did appear.

**What would you change in the groupware for it to become “very good”?**

Three of the users could not answer.

(german partner)

Avoiding HTML-code and use of a spell correction. Optimization of the forum and discussion layout and functions.

(Swedish technical partner)

WYSIWYG-editor tightly integrated, support for sending texts for approval by e-mail, reformatting of ugly HTML to look nicer and thus be easier to edit, support for converting ask-the-expert to FAQs.

(netdoctor)

Needs to appear more simple and have a more standard user interface. It is just not very logical to use without a lot of training.

(greek partner)

I think what bothered me most was the need for the use of html.

## **4.2 EVALUATION OF THE WEB CONTENT**

### **a. External evaluation:**

According to the deliverable D7.1 we are planning to give the Discern instrument will be given to an external reviewer with the request that he visits the home page of our site, makes a specific question on the search engine and reviews the information given. The Discern project was funded by the British library and the NHS Research and Development program and aimed to develop an easy to use instrument that can be used by consumers to rate the quality of health related information published in a written form or through the internet. This instrument has been used to rate the quality of information of various sites with health material in general and mental health material in particular. It was also used for the detailed evaluation of the Greek site during the evaluation of the existing content (Deliverable D1.1).

This kind of evaluation is not ready yet. We are hoping that it will be ready and presented in one of the following deliverables concerning evaluation.

## **b. Evaluation with the help of control users:**

Another method to test the groupware along with the website is to test the site on control users while they are being observed and their moves are being noted down. Afterwards the observer can question them on their moves, write down any possible problems or omissions of the groupware or the site and ask them to comment on their degree of satisfaction. This method is of course better because it allows an "interview" of the user in multiple domains, but is also more time consuming and thus difficult both for the user as for the observer.

This kind of method was tested from KTH with the help of control users. In Appendix Part I we have attached the results of this trial, presented and commented from the responsible for this task KTH technician.

### **c. Evaluation with the help of real users (ongoing evaluation):**

As is understandable an evaluation based only on the opinions of control users is not adequate to estimate whether the site has achieved its original goal of providing medical information to those who really need it. For this purpose we aim to collect data from the real users.

In Appendix Part II one can find attached the questionnaire which was created for the purposes of this evaluation and inserted into our website in order for the users to complete. This was done in early September but unfortunately we still do not have a good sample to work on and draw conclusions from their answers. Therefore we are leaving this part of the evaluation for a following deliverable to be commented upon.

### **4.3 EVALUATION OF THE SITE IN OTHER LANGUAGES**

Until the moment this deliverable is written the site has not yet gone public in any other language but English. Our plan is for the site to have gone public in all partner's languages (Swedish, German, Greek and Italian) until January 2004. We hope that after that time we will also have results concerning separately the quality and popularity of each linguistic version of the site. Some thoughts were also made to try and evaluate each different linguistic version with the help of control users (as proposed in 4.2a) but this depends on each partner's ability to find and try their site with such control users.

## 5. Evaluation of the usage

### 5.1 Methodology

Through an external tool we can have the usage statistics of the site, that is how many users visit the site, from how many different computers, how often each one has visited the site, and from which different countries they are. This data can be collected every two or three months and compared to the previously collected data and in this way we can come to some important conclusions about the usability and popularity of the web4health site.

Furthermore we can also use the ratings provided to us from the various internet search engines about the popularity of the site, the position given in their list and how it varies with different search words. This data can also be collected every few months and be compared and analyzed. This way we will be able to see if the site has achieved its goal of attracting patients and people in need of its services.

A statistics report was created for the purposes of the evaluation from KTH and was delivered on 01/11/03 (D7.2A Web4health Statistics). A short summary of this report is presented below:

The Web4Health web site was opened to the public in English in the middle of July, 2003. Since then until 31 October, 2003, the web site has had 63 000 visits and 1,92 million page hits. In both September and October, the number of visits were more than 40 000, compared with the target of 10 000 visitors/month specified in the KOM2002 contract.

The natural language question answering system answered more than 8 000 questions in September 2003. According to the evaluation made by the software itself, 80 % of the

questions were successfully answered. A human evaluation of the quality is not ready when this report is written (2 November 2003).

Most of our visitors came from Google and Yahoo, and the most popular pages were our menu pages on sexual disorders, depression, sleeping problems and healthy living and relational problems issues. Also some specific FAQs, such as on the incidence of eating disorders, the effects of sexual abuse on children and causes of feeling tired all the time were most popular.

On 31 October 2003, a total of 699 FAQs had been produced. Of these, 679 were searchable using the natural language question answering system in English, and 611 were accessible using the subject tree in English. The web site in other languages than English had not been opened yet at that time.

## 6. Conclusions and Timeplan

- a. At this point of the project we can say we have finished with the evaluation concerning the groupware and we are running an ongoing evaluation mostly about the website and its content.
- b. The results of the groupware evaluation showed that the groupware as created and used might have caused some practical problems, mostly to new users, but at the end it proved valuable to the progress of the project.
- c. The usage statistics show that the number of hits we get is far bigger than the one expected and this is due (at least to the extent we can trust the machine evaluation) to successful answers in common problems (most popular pages were those about sexual disorders, depression, sleeping problems etc).
- d. We are waiting in the next deliverables of one at least evaluation performed by an external reviewer and of a big enough sample of answers to the questionnaire inserted into the site and addressed to the users. These results will be most helpful to see how the site is judged by those it is aimed to serve.

## **BIBLIOGRAPHY**

1. Anonymous. Criteria for Assessing the Quality of Health Information on the Internet. Mitretek organization. <http://www.mitretek.org/hiti/showcase/documents/criteria.html> (accessed 23/12/2002).
2. Berland GK, Elliott MN, Morales LS. Health Information on the Internet. Accessibility, Quality, and Readability in English and Spanish. *JAMA* 2001;285:2612-2621.
3. Charnock D. The DISCERN Handbook. Quality criteria for consumer health information on treatment choices Radcliffe Medical Press, 1998.
4. Delamothe T. Quality of websites: kitemarking the west wind. Rating the quality of medical websites may be impossible. *BMJ* 2000;321:843-4.
5. Griffiths KM, Christensen H. The quality and accessibility of Australian depression sites on the World Wide Web. *MJA* 2002; 176: S97-S104.
6. Jadad AR, Gagliardi A. Rating Health Information on the Internet: Navigating to Knowledge or to Babel? *JAMA* 1998;279:611-614.
7. Kathleen M Griffiths KM, Christensen H. Quality of web based information on treatment of depression: cross sectional survey. *BMJ* 2000;321:1511-5.
8. Kim P, Eng TR, Deering MJ, Maxfield A. Published criteria for evaluating health related web sites: review. *BMJ* 1999;318:647-9.
9. Risk A, Dzenowagis J. Review of Internet Health Information Quality Initiatives. *Journal of Medical Internet Research* 2001;3(4):e28.
10. Winker MA, Flanagan A, Chi-Lum B, White J et al. Guidelines for Medical and Health Information Sites on the Internet. Principles Governing AMA Web Sites. *JAMA* 2000;283:1600-1606.
11. Wordtracker ([www.wordtracker.com](http://www.wordtracker.com))
12. HON (Health On the Net) [www.hon.ch](http://www.hon.ch)

**Medcircle** [www.medcircle.org](http://www.medcircle.org)

## APPENDIX

### PART I:

We are attaching the results of the evaluation testing which was performed from KTH to control users in a previous part of the project. The results are presented and commented from the responsible KTH technician.

I have tested the Web4health web site with three people. In all three cases I have the following instruction: "You have been told that there is interesting information about eating disorders on this web site. So you connect your computer to this page. I will try to not tell you what to do, just observe what happens, in order for me to learn how easy or difficult or web site is."

I then let the person use Web4health for about ten minutes, then asked some questions of why they did what they did, and in one of the cases pointed them in the right direction, and noted down what happened.

Here is a summary of what happened for the three persons (all women):

Person A:

She started by writing "huger" in the "Write your question in natural language" textbox, and was confused when she got no result. Not until I pointed out that she had misspelled "hunger" as "huger" did she note this. I also asked why she wrote only one word, she said she was accustomed to starting with one word queries in search engines.

She then tried with "hunger", looked at the list of answers, found answer no 4 most interesting and looked at that. She then scrolled down from the answer no 4 to the succeeding answers.

I asked why she started with the text box, she said she thought that was what I expected. I then asked what she might have done in the start page other than writing in the text box. She then looked at the menus on the home page, found the menu "Links to related sites" and clicked on an item in that menu.

Person B:

This person had some previous experience with using Web4health. She started with a question, and rapidly started finding non-working links in some of our pages.

Person B had too much experience with using our system to be a good testing of the system as it appears to new users.

Person C:

(Person C said at the start that she did not have very much computer experience, only used e-mail, not very much WWW.)

Person C first clicked in the "Write your question in natural language" then looked a little at the screen, then clicked on "Anorexia" in the subject tree below.

(I asked her after the session why she did not write anything in the text box, she said that since the subject tree seemed to contain what she wanted, she did not feel any need to type a question herself.)

She read the page which "Anorexia" gets to, clicked on a link, read a page, clicked on a link to a related page, etc. until she came to a page which did not have any links to related pages.

She then clicked on home, went into the subject tree again and looked at one or two pages.

She then clicked on "join us" in the home page. She then got to <http://cmc.dsv.su.se/eu/kom/1;register>, she decided she did not feel willing to register as a user. But this page contained a command "Go to: All forums" (but actually "all forums" was in Swedish, i.e. "Go to: Alla forum". She clicked on this, then saw a \_ome on that page saying:"Language: Swedish ( 0 ) , English ( O )" She clicked on "Swedish" expecting to switch to a Swedish user dialogue, but the user dialogue stayed in English.

We stopped the session at that point.

#### Conclusions:

From the experiment with person A, I conclude

- (a) We should have handling of spelling errors in questions.
- (b) Also I conclude that it may not be good to have the main subject tree with blue background, person A (but not person B and C) sort of disregarded the subject tree in blue background and instead went to the menu with white background with links to related sites. And we certainly do not want our visitors to start by going to related sites!!

From the experiment with person C I conclude:

- (c) It is very good if all pages have links to each other, so that people can move around from page to page just by clicking on the links in the pages.

(d) People do not like to register, the page

<http://cmc.dsv.su.se/eu/kom/1:register> was good because it had a menu to the left to bypass registration.

Also "Ask the expert" and "All forums" should lead to pages where there is an option to bypass registration (which they do not have now). People should be allowed to read in forums, and to write ask-the-expert question, without having to log in or register. Whether people should be allowed to write in open forums without registering first is something we have to discuss.

(e) The user was confused by a mixture of English and Swedish text, like for example "Go to: Alla forum".

(d) The user was confused when clicking on "Language Swedish" did not switch to a fully Swedish page.

(She did not understand that this command was meant only to show the text of the current message in Swedish - not easy to understand!)

(e) The "Help" command did not work in any meaningful way. It should probably go to contextual help like the "(i)" icons instead of to general help.

## **PART II:**

### **QUESTIONNAIRE TO BE APPLIED TO REAL USERS:**

**What did you search information about?**

**Was the content of the site useful to you? Please rate on a scale from 1 to 5:**

- 1 (not at all useful)*
- 2*
- 3*
- 4*
- 5 (very useful)*

**Please rate the information you obtained on a scale from 1 (very badly written-not understandable) to 5 (very well written-totally understandable):**

- 1*
- 2*
- 3*
- 4*
- 5*

**On a scale from 1 to 5 please rate to which point the site made clear what sources of information were used to compile the content:**

- 1 (no sources of evidence for the information are mentioned)*
- 2*
- 3*
- 4*
- 5 (the sources of evidence are very clear)*

**On a scale from 1 to 5 please rate whether it was clear when the information was produced:**

- 1 (no, no dates existed on any text or answer or source of information)*
- 2*
- 3*
- 4*
- 5 (yes, there were dates on every text of information I was given from the site)*

**Please rate to what point you thought the information was balanced and unbiased?**

- 1 (the information is completely unbalanced or biased)*
- 2*
- 3*
- 4*
- 5 (the information is completely balanced and unbiased)*

**Please rate to what point you thought the information covered your questions in depth:**

- 1 (no, I was left with many questions unanswered)*
- 2*
- 3*
- 4*
- 5 (yes, it had all the necessary details)*

**Did you think the site was satisfactory to the links provided for additional information?**

- 1 (no, it provided no links at all/all links were unuseful)*
- 2*
- 3*
- 4*
- 5 (yes, it provided many useful links)*

**Were the aims of the site clear to you? Please rate on a scale from 0 to 5:**

- 1 (no, the site does not include any indication of its aims)*
- 2*
- 3*
- 4*
- 5 (yes, the site has clear aims)*

**Do you think the site achieves its aims? Please rate on a scale from 1 to 5:**

- 1 (none of the information you were expecting from the aims has been provided)*
- 2
- 3
- 4
- 5 (all the information expected from the description of the aims has been provided)*

**Who do you think has the ownership of this site?**

**Please rate how easy/difficult to navigate the site was (navigation=finding your way in the web site) on a scale from 1 (very difficult) to 5 (very easy):**

- 1
- 2
- 3
- 4
- 5

**Please rate the mechanism of feedback (interactivity) of this site:**

- 1 (not at all satisfactory-very difficult to ask questions and obtain extra information)*
- 2*
- 3*
- 4*
- 5 (very satisfactory-very easy to ask questions and obtain extra information)*

**Based on the answers to all of the above questions, rate the overall quality of the site as a source of information about mental health:**

- 1 (very bad)*
- 2*
- 3*
- 4*
- 5 (very good)*

**What is more important for you when seeking a medical information on the internet? (you can click more than one options)**

- Quick information*
- Anonymity*
- Ability to find information from various sources*
- Currency of the information*
- Objectivity*

- Understandability*
- Detailed information*
- Ability to ask an expert*

**How old are you?**

- Less than 16*
- 16-20*
- 21-25*
- 26-30*
- 31-40*
- More than 41*

**Which country do you come from?**

**Are you currently employed?**

- Yes, full time.*
- Yes, part time.*
- No, I am unemployed.*
- No, I am on pension.*
- No, I am disabled to work.*
- I am economically inactive (student, housewife).*

### How often do you visit web4health?

- Only once*
- It is the second time*
- Few times*
- Quite often*
- Almost daily*

### PART III

Here is a preliminary, partial report of ongoing evaluation work done by FP and KTH.

#### Test of the natural-language question-answering system

The natural-language question-answering system was tested by randomly selecting 100 questions which actual users had given to the system. For each question, a psychology student evaluated the question in five classes, as shown in Table 1

**Table 1: Manual evaluation of the quality of answers given by the natural-language question-answering system to 100 randomly selected actual user questions**

	Of all questions	Of relevant questions
Good answer exists and was found	43 %	53 %
Answer found, but not so good	4 %	5 %
Answer exists, but was not found	10 %	14 %
No answer exists, irrelevant questions <sup>1</sup>	20 %	
No answer exists, but should exist	23 %	28 %

From Table 1 one can conclude that recall was  $43/53 = 81\%$  and that 53 % of all relevant questions got a good answer from the system. One can note that user satisfaction was actually higher than 53 % (see XXX), maybe users were satisfied also when they got answers which were not so exactly relevant to their questions.

---

<sup>1</sup> Questions about purely somatic illnesses were evaluated as irrelevant, since this is not within the scope of our web site. Also obvious tests or jokes or questions like “What time is it” were regarded as irrelevant.

## User satisfaction questionnaire

Table 2 shows preliminary results of the user satisfaction questionnaire described on page 31ff.

**Table 2: Preliminary result of user satisfaction questionnaire**

Not at all useful	8 %	A total of 88 % were thus satisfied with the information provided.
Not very useful	3 %	
Somewhat useful	29 %	
Useful	24 %	
Very useful	35 %	

Note that only 78 out of 30000 visitors filled in this questionnaire!

## Controlled user testing

Controlled user testing was done on (when this report is written) 11 patients at a psychotherapeutic practice in Stockholm. Each patient was first asked to try out the system, and then answer question on their experience. Preliminary results are shown in Table 3.

**Table 3: User satisfaction in a controlled trial with a small number of patients**

Worthless	1	Thus, 81 % were satisfied with the system
Not useful	1	
OK	1	
Good	7	
Very good	1	
Total	11	