



European digital content for the global networks

EDC-22071 – KOM2002/27938 – 22071Y2C2DMAL2

Groupware for Distributed Content Production

KOM2002

D3.2 Report of Second Stage

Security (distribution level)	Public
Contractual date of delivery	April 2004
Actual date of delivery	May 2004
Deliverable number	D 3.2
Deliverable name	Report of Second Stage
Type	R
Status & version	Up-to-date until May 2004
Number of pages	12
WP contributing to the deliverable	WP3
WP / Task responsible	Omega
Other contributors	DSV
Author(s)	Michele Carenini, Jacob Palme
EC Project Officer	Kimmo Rossi
Keywords	Contents, usage, Web4Health, statistics, access
Abstract (for dissemination)	This report focuses on the main developments from Stage 1 to Stage 2 of the Project in the Web4Health web sites.

*Type: PU-public, LI-limited, RP-restricted

**Nature: PR-Prototype, RE-Report, SP-Specification, TO-Tool, OT-Other

Table of contents

1	Executive Summary	3
2	National Web Sites	3
3	Active services	5
3.1	Natural Language Question Answering System	5
3.2	Static FAQ's	5
3.3	Ask-the-expert	6
3.4	Forum	7
3.5	Chats	8
4	Contents and Statistics	8
5	Statistics of Web Site Accesses	9

1 Executive Summary

This report will mainly describe how the national Web4healths system web sites evolved in the last six months, in terms of both their structural characteristics and contents.

The Web4Health web site was opened to the public in English in the middle of July, 2003. Since then to April 2004, the web site had more than 7 million hits. In April 2004, there have been about 84.000 visitors, 24.000 page downloads, almost 1.5 million hits. The answers success rate reached by the natural language question answering system (automatically calculated) remains, as in previous evaluation, about at 80%. As it will be detailed in further in this report most visitors came from Google and Yahoo, although figures for other portals are far from relevant. In six months, the overall number of FAQ's (i.e. inclusive of translations of the same FAQ in different languages). has grown to 3063 (April 2004).

2 National Web Sites

National (English, Swedish, German, Greek, Italian) web sites all share the same access page, split in the following in order to highlight the main features.

Free medical advice on mental health, psychology, personality disorders, relationships, stress, anxiety, depression, emotional abuse, substance abuse, sexual abuse, types of mental illness, etc.

Home **Search** **Categories** **Discussion** **Ask the Expert** **Unseen** **Login/out** **My account**

Psychology Free Online Medical Advice

Answers to more than 700 questions about psychology, mental health and relationships, written by a team of experts appointed by the Commission of the European communities.

OTHER LANGUAGES

Italy Sweden Germany Greece

RECENT DISCUSSIONS

06 May 06:03
emotional shutdown

In the upper part of the page it is possible to log in, as well as to access main services offered by the system and to select the available languages by clicking on the flags on the right. It is also possible to join any on-going discussion or to read recent discussions on particular topics.

SEARCH FOR ANSWERS

Write a simple question in one sentence (Note: Our answers are not oriented towards somatic - body - medicine):

ⓘ

Advanced search

11 May 06:58
Re: Aisha


11 May 07:04
Re: Sibling causing problems


13 May 07:59
Re: Sibling causing problems

This part of the page allows the user to submit a question that will be analysed by the natural language question answering system (for English, Swedish or German). The Greek site does

not provide any search facility. The Italian site does not provide a natural language processing based tool, but uses Alkaline (alkaline.vestris.com/):

TESTO DA RICERCARE IN TUTTO IL SITO:

Inserisci una o più parole 

Alkaline is a traditional free text search engine, which of course gives less good precision and recall and shows a smaller degree of (linguistic) flexibility.

CATEGORIES			LATEST QUESTIONS
Addiction	Healthy living	Psychotherapy	15 Apr 00:32
ADHD	Instability	Relations	GP: About my relationship
Anorexia	Life	Sex	26 Apr 13:02
Anxiety	Mood swings	Sleeping problems	Re:
Bulimia	Obesity	Stress symptoms	14 May 23:42
Child care	Phobia	Workplace	EFFEXOR XR SIDE EFFECTS
Depression	Psychiatric drugs	More.....	18 May 18:02
<p>Help us make Web4Health better, fill in a questionnaire with your views and suggestions (10 minutes)</p>			Re: EFFEXOR XR SIDE EFFECTS
			All answers

This section displays all available categories, i.e. topics to be browsed by users. Each category is organised as a subcategory tree, whose leaves are in the question/answer form. When reaching a Q/A, it is also possible to search for related arguments (a few – more precision, possibly less information – or many – less precision, possibly more information).

[Sources, references](#)

It is also possible to go and check both the Source of the information and References (if any). The last part of the access page contains disclaimer and copyright notices, and a copy of the main menu showing services.

Disclaimer: The documents contained in this web site are presented for information purposes only. The material is in no way intended to replace professional medical care or attention by a qualified practitioner. The material in this web site cannot and should not be used as a basis for diagnosis or choice of treatment. If you find something which should be corrected, please write to cmc@dsv.su.se.

[Home](#) | [Search](#) | [Categories](#) | [Discussion](#) | [Ask the Expert](#) | [Unseen](#) | [Login/out](#) | [My account](#)

[About Web4Health](#) | [Contact us](#) | [Medical experts](#) | [Communities](#) | [Join us](#) | [External reviews](#) | [Your privacy](#) | [Rules](#)

Copyright 2003, 2004 Web4Health

3 Active services

In this section a brief description of the active services functioning is provided. Information on content (inclusive of statistics) should be looked for in section 4.

3.1 Natural Language Question Answering System

Instead of browsing the FAQ list, the user¹ may insert a question using his/her language, which is analysed and matched to the most proper FAQ. The Natural Language Question Answering System (NLQAS) depends heavily on the classification process of FAQ's. NLQAS is based on manually specifying templates for each FAQ. The templates are matched against the questions asked by users; thus success of the question-answering depends a lot on the quality of these templates.

3.2 Static FAQ's

The term "FAQ" then is used here to describe *the record in the data base describing an answer*. The term "query" refers to the actual query written by a user.

An FAQ is defined by the following attributes:

1. **ID:** A unique identifier to separate this from other FAQs.
2. **Title:** A user-oriented textual description of the FAQ. Important for finding the document by search engines such as Google.
3. **Required keywords:** Keywords and phrases which must exist in the user query for a match to be done.
4. **Priority:** Keywords which will cause this FAQ to be shown before FAQs without this keyword.
5. **Optional keywords:** Keywords which may or may not exist in the user query. Words in "Priority" need not be repeated here.
6. **Forbidden keywords:** Keywords which may not occur in the user query.
7. **Number of non-envisaged keywords:** If the user query contains no more than this number of non-specified keywords, the response is shown to the user as a *probable match*. If the user query contains more than this number of non-expected keywords, the response is shown to the user as a *possible match*.
8. **Body:** The HTML or XHTML text of the answer to the FAQ.

Keywords (3., 5. and 6.) are actually groups of synonyms, in order to be as much inclusive as possible of the real user's phrasing. In order to do this, a structured stemming procedure is implemented. An example of an FAQ's definition of keywords could be the following:

bulimic* [binge; eat*] ; obes* overweight* fat* corpulent* stout* fleshy [weigh* # \$much]

¹ The natural language question answering system is available for Swedish, English and German only. No search engine is available for the Greek web site, while the Italian web site uses a traditional text retrieval tool, Alkaline.

Such a set will match questions like:

- *Are bulimic people obese?*
- *Will binge eaters weigh unusually much?*

The * symbol at the end of one string means any additional characters at the end, for example *obes** will match both “obese” and “obesity”.

Phrases are matched if specified within square brackets [].*[binge; eat*]* will match with “binge eating”, “binge eaters”, “binge-eaters”, etc.

Phrases may include # to indicate an arbitrary text between the first element and the second element - *[weigh* # \$much]* will match “weight as much”, “weights very much”, “weight not too much”.

If the same synonym occurs in many FAQs, it is possible to include it in a separate synonym list, and refer to this list with, for example, “\$much” to refer to a list of synonyms for “much”. This list is named the *substitutes* list.

A complete description of the classification system for FAQ in Web4Health is given in J. Palme and E. Sneiders, KOM2002 D1.2-part B, “Natural Language Question Answering System Classification Manual”, December 2002.

3.3 Ask-the-expert

This service allows the user to submit questions that will be answered by psychological-psychiatric experts. Questions and answers are made publicly available. Initially, users could choose whether to be anonymous or not, but after suggestions from the CEC, we have changed the system so that all questions are completely anonymous. If users specify their name in the text of their questions, we remove these names before publishing the question. If the users wish to receive their answers by e-mail, they must register their e-mail address before writing the question, but this address is not used for anything else than sending the answer to the user and is not even visible even to us. Users cannot be guaranteed to receive an answer to their questions. Short questions (max 10-20 lines) are more likely to be answered. Average time to receive an answer is one to two weeks.

Write Question to Ask The Expert Here

Author or secret name (pseudonym):

Reply wanted by e-mail to: This e-mail address will never be shown to anyone else.

Subject:

Language:

Enter the text... Plain Text Guess HTML

Plain text
 Reply
 Happy
 Sad
 Angry
 Question

[See other people's questions and answers](#)

The users may choose a pseudonym (a default one is provided by the system). They then may decide to specify a subject and a language. Further, the question may be submitted either in plain text or HTML (the users may see a preview of their texts before submitting the question), and completed with a symbol specifying if the message itself is a reply or a question. Emoticons may be added.

All the list of past questions and answers may be browsed. Ask-the-expert is available for the English, Swedish, German and Italian site. The relative link in the Greek site leads to an ask-the-expert service which the Greek partner is providing on their own web server. Since they already had this service running, the project saw no need to copy it on the Web4Health server.

3.4 Forum

The Forum service will be described in detail in D5.2.

3.5 Chats

The software has a chat function, but it has not been used except for internal chats within the partners.

4 Contents and Statistics

The following table contains the total number of FAQ's produced and translated for Greek, German, English, Italian and Swedish.

Language	Greek	German	English	Italian	Swedish
<i>Number of written (and translated if originally in another language) FAQ's</i>	317	612	899	479	756
<i>FAQ's which do not have text yet</i>	1	1	2	0	2
<i>Number of searchable FAQ's using the Natural Language Question Answering System</i>	28	577	852	121	703
<i>Number of FAQ's written but not yet searchable</i>	288	34	45	358	51
<i>Number of FAQ's written and in the subject tree</i>	271	120	831	0	691
<i>Number of FAQ's written and not in the subject tree</i>	45	491	66	479	63
<i>Number of FAQ's in the subject tree but not available in this language (users will be shown the English text instead)</i>	209	105	6	0	0

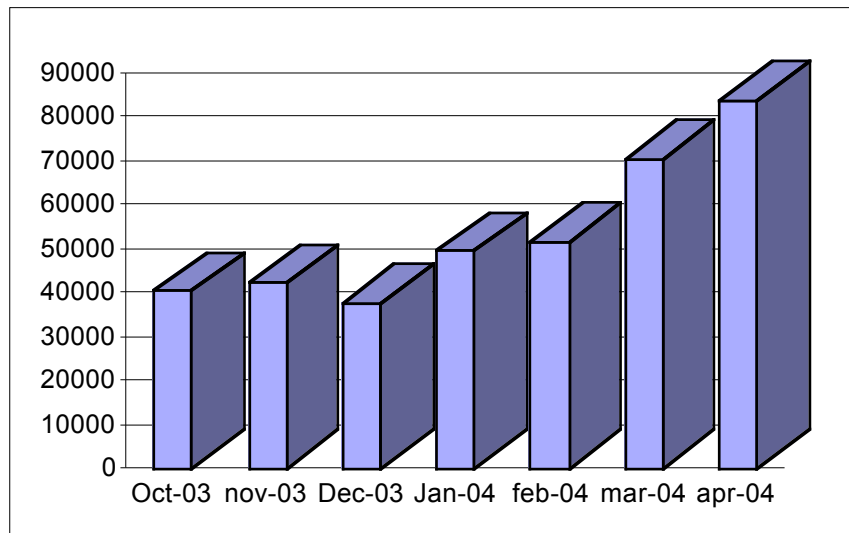
This table instead contains relative number of FAQ's about different topics for each language.

Language	Greek	German	English	Italian	Swedish
<i>Addictive disorders</i>	30	32	94	15	43
<i>ADHD, ADD, etc.</i>		81	49	9	44
<i>Anxiety disorders</i>	37	45	36	10	26
<i>Somatic disorders and symptoms</i>	99	18	58	13	23
<i>Mood disorders (depression, bipolar disorder, etc.)</i>	38	119	106	35	100
<i>Borderline disorder</i>		11	15	3	15
<i>Children and child are</i>		14	25	18	24
<i>Eating disorders</i>	16	105	211	122	209
<i>Presentations of project experts</i>		1	4	1	2
<i>General health issues</i>		1	2	8	2
<i>Healthy life, relational problems, etc.</i>	2	29	59	91	59
<i>Links to other web sites</i>	1	1	1	1	1
<i>Phobia</i>	17	16	6	7	6
<i>Information about the web site, user privacy, etc.</i>	2	12	26	8	17
<i>Other psychological issues</i>	43	37	38	12	32
<i>Sexual disorders and problems</i>	15	40	109	104	102
<i>Sleep disorders</i>	1	19	18	6	16
<i>Somatic aspects of psychiatric disorders</i>	1	20	25	3	17
<i>Therapeutic methods</i>	14	10	15	13	16

5 Statistics of Web Site Accesses

In April 2004, the Project web site had 84.059 visitors, 24.714 page downloads, 1.431.267 hits. Since the opening of the site, it had more than 7 million hits. The following table and graph show the monthly number of visitors since last October 2003 (included):

<i>August 2003</i>	17 277
<i>September 2003</i>	42 663
<i>October 2003</i>	40 307
<i>November 2003</i>	42 285
<i>December 2003</i>	37 383
<i>January 2004</i>	49 360
<i>February 2004</i>	51 760
<i>March 2004</i>	70 382
<i>April 2004</i>	84 059



In April 2004, national sites have been accessed as in the following table:

<i>English</i>	62 449	78,40%
<i>German</i>	3 617	4,50%
<i>Greek</i>	46	0,10%
<i>Italian</i>	3 617	4,50%
<i>Swedish</i>	9 940	12,50%

whose total is slightly different from the figure of total visits given above. This most likely derives from the fact that some pages which do not belong to any specific language area, and which should not be found by users (as, e.g., <http://web4health.info/test/>), are retrieved anyway.








The number of visits to a site in a certain language depends on a number of factors:

- Total number of pages, where English and Swedish have most pages, about 700 English and 600 Swedish.
- Competition, which is higher in larger language areas like English and German.
- Search engine optimisation, which has been done a lot for the English pages, and a little for the Swedish pages.
- Percentage of the population which uses the Internet, which is probably lower in Greece and Italy.
- Whether the most popular search engines and portals in a language area has found our site and gives it high rankings.

In April 2004, the portals which most often gave visitors to the service were:

www.google.com/search	24341
search.yahoo.com/search	7009
www.google.de/search	6473
www.google.se/search	6043
www.palme.nu/free/ideal.html	3424
www.google.it/search	2897
www.google.ca/search	2612
search.msn.com/pass/results.aspx	1676
search.msn.com/results.asp	1665
aolsearch.aol.com/aol/search	1636
www.google.co.uk/search	1556
search.msn.com/results.aspx	1495
www.eniro.se/query	1419
www.google.com.au/search	1185
www.ask.com/udir	923
search.msn.com/spresults.aspx	744
www.google.ch/search	603

Some of the search engines have multiple addresses (especially for national sites); the following table gathers them:

	Google	45710	69,60%
	Yahoo	7009	10,70%
	www.palme.nu/free/ideal.html ²	3424	5,20%
	MSN	5580	8,50%
	Eniro ³	1419	2,20%
	Ask	923	1,40%
	AOL	1636	2,50%

Finally, it may be interesting to list most popular search strings/phrases which, still through portals, got people to access ourservice:

accesses	percentage	phrase	position in Google
386	1.13%	<i>stress symptoms</i>	5
351	1.03%	<i>medical advice</i>	2
337	0.99%	<i>always tired</i>	2
315	0.92%	<i>human life cycle</i>	1
301	0.88%	<i>causes of eating disorders</i>	4
291	0.85%	<i>mood swings</i>	5
272	0.80%	<i>anorexia statistics</i>	8
270	0.79%	<i>eating disorders</i>	not among first results
257	0.75%	<i>bulimia statistics</i>	1
220	0.64%	<i>causes of homosexuality</i>	2
220	0.64%	<i>sex education in public schools</i>	3
212	0.62%	<i>cutting self mutilation</i>	2
204	0.60%	<i>eye twitches</i>	4
193	0.57%	<i>things to do when bored</i>	9
182	0.53%	<i>how to make yourself vomit</i>	not among first results
172	0.50%	<i>muscle twitches</i>	3
162	0.47%	<i>sleeping problems</i>	5
144	0.42%	<i>psychotic depression</i>	8
133	0.39%	<i>what is my ideal weight</i>	3
128	0.38%	<i>sexual education in schools</i>	1

² An article in Swedish by Gunborg Palme about eating disorders.

³ Eniro is a Swedish search engine using Google technology.