



European digital content for the global networks

EDC-22071 – KOM2002/27938 – 22071Y2C2DMAL2

Groupware for Distributed Content Production

KOM2002

D7.2A Web4Health Statistics

Security (distribution level)	Public
Contractual date of delivery	November 2003
Actual date of delivery	1 November 2003
Deliverable number	D 7.2A
Deliverable name	Web4Health Statistics
Type	R
Status & version	Up-to-date until October 2003
Number of pages	11
WP contributing to the deliverable	WP7
WP / Task responsible	Ioannina
Other contributors	
Author(s)	Jacob Palme
EC Project Officer	Kimmo Rossi
Keywords	statistics, usage, Web4Health
Abstract (for dissemination)	This report Gives statistics on the content and usage of the Web4Health web site.

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

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

Table of contents

1	Executive Summary	2
2	Introduction	3
3	Number of FAQs Produced	3
4	Statistics of Web Site Accesses	4
5	Most Popular Pages	5
6	Entry Pages.....	6
7	Referrers and Search Strings.....	6
8	Usage of the Natural Language Question Answering System.....	6
9	Summary of How People Find Information in Web4Health	6
	Appendix A: More Detailed Statistics.....	8

1 Executive Summary

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 where accessible using the subject tree in English. The web site in other languages than English had not been opened yet at that time.

2 Introduction

As part of the evaluation of KOM2002, this report gives some statistics on the content and usage of the Web4Health web site. The report is valid up to the end of October 2003.

3 Number of FAQs Produced

Table 1 shows the total number of FAQs ready in each language. This includes both FAQs originally written in that language, and FAQs originally written in another language and translated to that language. In addition to this, the FAQ data also contains 39 (in English) subject tree pages and 26 (in English) pages for creating headings in the list of all FAQs at <http://faq.web4health.info/w4h-faq-en/faq.exe?>

Table 1: Total number of FAQs produced and translated by language

Language:	de	el	en	en ¹	it	sv
Number of written (and translated if originally in another language) FAQs	376	12	699	699	44	585
FAQs which do not yet have text	0	0	2	2	0	2
Number of searchable FAQs using the natural-language search system	333	10	679	679	7	417
Number of FAQs written but not yet searchable	43	2	20	20	37	168
Number of FAQs written and in the subject tree	61	0	504	611	0	29
Number of FAQs written and not in the subject tree	315	12	195	88	44	556
Number of FAQs in the subject tree but without text	64	0	7	10	0	3
Total entries in the FAQ data base	376	12	701	701	44	587

Table 2: Produced FAQs by main subject area

Area shortname	Area description	Number of written (and translated if originally in another language) FAQs
ed	Eating disorders	208
sex	Sexual disorders	107
bipolar	Mood disorders (depression, bipolar disorder,	104

¹ Includes FAQs in an internal data base for the subject tree, but some of which have not yet been entered into the user-available subject tree.

Area shortname	Area description etc.)	Number of written (and translated if originally in another language) FAQs
bio	Somatic disorders and symptoms	78
life	Healthy life issues, relational issues, etc.	56
adhd	ADHD, ADD, etc.	50
anx	Anxiety disorders	34
psy	Other psychological issues	32
child	Child rearing issues	24
project	Information about the project	18
border	Borderline disorder	16
therapy	Therapeutic methods	15
sleep	Sleep disorders	12
phobia	Phobia	5
experts	Presentations of project experts	3
health	General health issues	2

4 Statistics of Web Site Accesses

Table 3 shows the number of visitors and page hits on the Web4Health web sites.

Table 3: Total Web site access statistics

	Jun-03	Jul-03	Aug-03	Sep-04	Oct-03
web4health.info visitors	294	5162	17277	42663	40307
faq.web4health.info visitors	558	810	1250	6590	6154
web4health.info hits	106314	143575	335425	747096	626678
faq.web4health.info hits	4413	2918	6556	32427	21442
total hits	110727	146493	341981	779523	648120

Note 1: All of these are not human visitors, some of them are visits by search engines. We have no exact figures, but one indication of how many page hits were made by search engines is that the file “robots.txt”, which is mainly of interest to search engines, had 923 downloads in October 2003. From some other statistics, one can infer that the number of robots represent about 15-20 percent of all hits on our web site, and that the rest are real humans. Probably less than 5 % are our own accesses when testing and developing the system.

Note 2: The slight decline in October 2003 was because our web site was “Cool site of the day” on 26 September 2003 on a popular American portal, and we then got ten times the normal number of visitors on that single day. *Excluding the additional number of visitors on*

that single day, the monthly number of visitors would have been as shown in Figure 1 and the number of page hits as shown in Figure 2.

Figure 1: Adjusted number of visitors

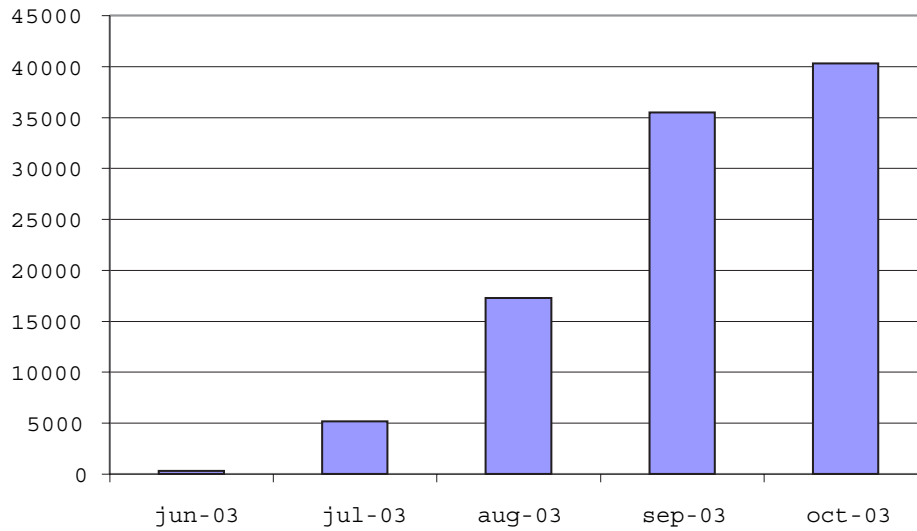
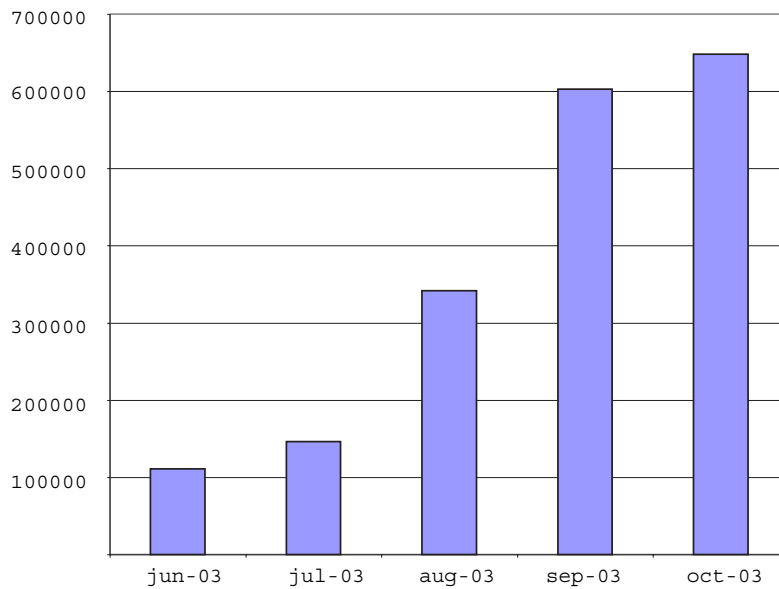


Figure 2: Adjusted page hit statistics



5 Most Popular Pages

As can be seen in Table 4 in appendix A, the two most popular pages was our main home site and our home site for forums and ask-the-expert question.

Of the specialized pages, the most popular were the subject tree pages on sexual disorders, depression, sleeping problems and healthy living issues.

Of the non-menu specialized pages, the most popular were pages on “surgery against obesity”, “the incidence of eating disorders”, “the effects of sexual abuse on children”, “symptoms of depression” and “the prevention of self-harm behaviour”.

In the ask-the-expert area, most visited were answers on sleeping problems and on grandparents role in the life of their grandchildren.

6 Entry Pages

The most popular entry pages (pages where a visitor first came to our site, usually from a search engine such as Google) are shown in Table 5 in Appendix A. These pages were pages on the surgical treatment of obesity, on the incidence of eating disorders, on the effects on children of sexual abuse, on the cycles of life, on prevention of self-harm, on the symptoms of depression and some of our main subject tree pages.

7 Referrers and Search Strings

As shown in Table 6 in Appendix A, most of our visitors came through Google, Yahoo and other search engines and web portals. We also got quite a number of visitors from a document in Swedish on eating disorders written by one of our partners before the project start and placed on her personal web site, probably because this page is indexed in Swedish web portals.

As shown in Table 7 in Appendix A, the most popular search strings used in search engines to reach our web page were “medical advice”, “causes of eating disorders”, “stress symptoms” and “causes of homosexuality”. This does not mean that these were the most popular search terms, but rather that these were the most popular search terms for which major search engines like Google recommend our web site, i.e. probably because there were few competing web sites with as good information as we have on these questions.

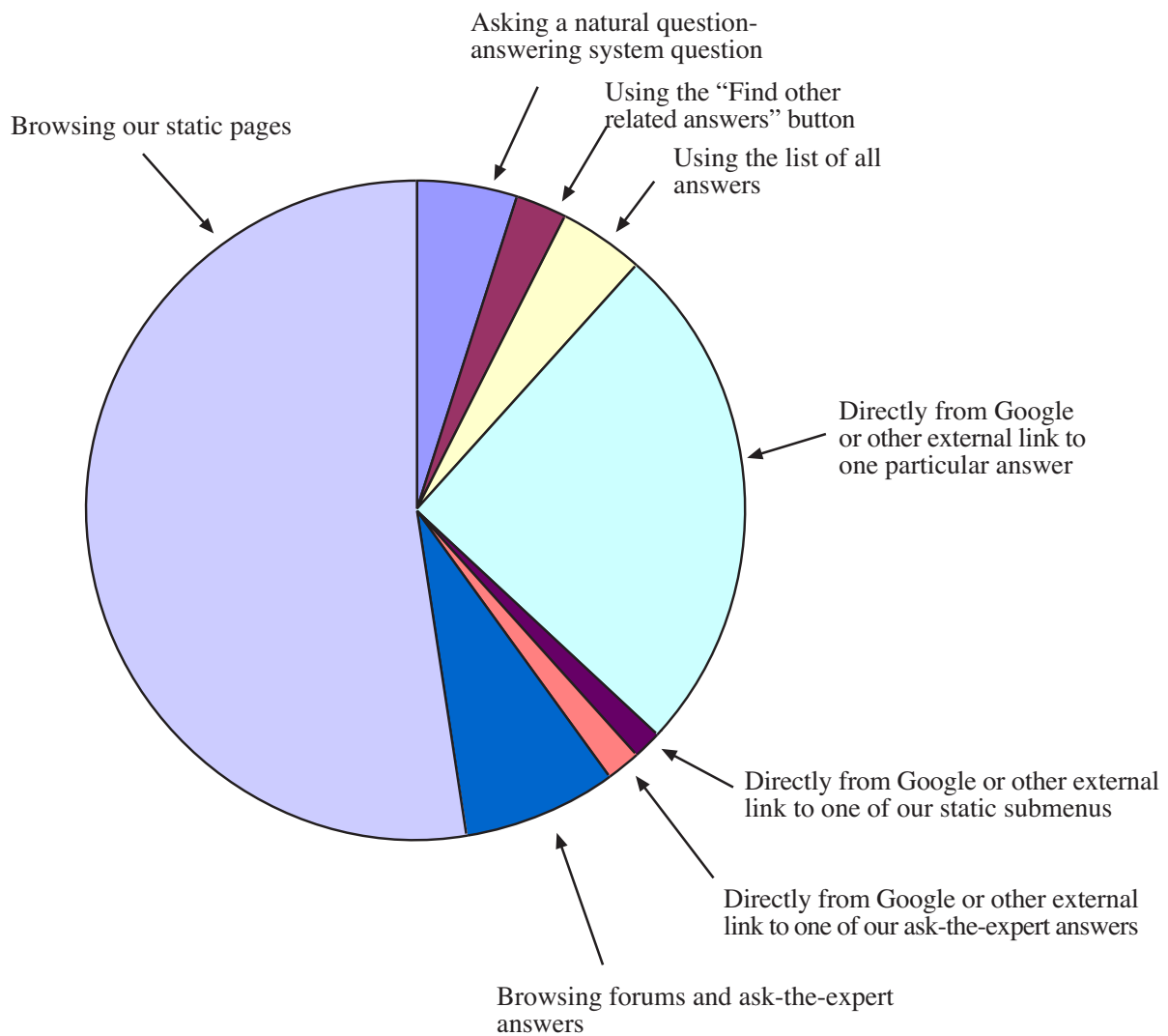
8 Usage of the Natural Language Question Answering System

Table 8 and Table 9 in appendix A show that the natural-language question-answering system received 5630 questions in September 2003. The system itself judged that it could find a good answer for 80 % of these questions. Only 2 % of all users told us if the answer was what they wanted to find, and about half of these were satisfied with the answer. This does, of course, not say anything about the satisfaction of the 98 % who did not say whether the answer found was acceptable or not.

9 Summary of How People Find Information in Web4Health

Combining data from various logging sources, the following *very approximate* estimate can be made on how people in September 2003 found answers to their question in our web site:

Figure 3: How people found answers on our web site



A conclusion of this is that we should spend more effort on the most important ways in which people find information, browsing our static pages and using Google to get directly to individual pages in our site.

Appendix A: More Detailed Statistics

Table 4: Most popular pages, hits

	sep-03	oct-03	
	18615	7930	
	14195	15623	
	2023	1528	
	1965	2017	
	1643	953	
	2426	2697	
	1356	411	
	1334	601	
	1145	883	
Sleeping too much, wrong time and depressed	913	14	
	807	296	
	747	574	
	698	1077	
	733	497	
	693	1041	
	674	703	
	648	734	
ask-the-expert-here	http://web4health.info/en/answers/ed-causes-culture-example.htm	420	770
	344	678	
	50	638	
	563	599	
	468	564	
	470	523	
	380	514	
	584	505	

Table 5: Entry pages visitors

	sep-03	oct-03
	18615	7930
	813	
	713	
	634	1077
	580	1041
	559	883
	541	734

	sep-03	oct-03
http://web4health.info/en/answers/sleep-menu.htm	1356	
http://web4health.info/en/answers/sex-menu.htm		1528
http://web4health.info/en/answers/bipolar-menu.htm	432	
http://web4health.info/en/answers/somatic-menu.htm		883
http://web4health.info/en/answers/border-selfharm-prev.htm		523
http://web4health.info/en/answers/bipolar-phys-depr.htm		703

Table 6: Mostly frequent referrers

	sep-03	oct-03
http://www.google.com/search	10282	13964
http://search.yahoo.com/search	5355	6686
http://www.palme.nu/free/ideal.html	2170	2611
http://www.google.ca/search	1169	1758
http://aolsearch.aol.com/aol/search	1014	1190
http://www.google.com.au/search	831	865
http://www.google.co.uk/search	745	805
http://www.google.com	743	
http://www.web4health.info	439	
http://www.palme.nu/web4health/	312	
http://www.zeit.de/	266	
http://216.239.41.104/search	262	
http://216.239.37.104/translate_c	231	
http://216.239.39.104/translate_c	213	
http://search.netscape.com/nscp_results.adp	212	
http://www.komando.com/kool_show.asp	197	
http://216.239.39.104/search	194	
http://web4health.info:9999/w4h/search.html	194	
http://www.palme.nu/web4health/swedish.html	190	
http://www.google.co.in/search	183	
http://www.dsv.su.se/jpalme/web4health/answers.html	167	
http://drs.yahoo.com/S	163	
http://dir.yahoo.com/Health/Mental_Health/	160	
http://216.239.51.104/search	157	
http://www.seelenfarben.de/wunderkiste.htm	149	
http://uk.search.yahoo.com/search/ukie	137	
http://www.google.com/custom	128	
http://web4health.info	124	
http://www.google.co.nz/search	121	
http://216.239.53.104/search	120	

Table 7: Search strings in external search engines

	sep-03	oct-03
medical advice		585
causes of eating disorders	125	306
stress symptoms	264	286
causes of homosexuality	215	268
human life cycle	188	240
sexual disorders	85	197
bulimia statistics	124	190
always tired	151	181
effects of child abuse		167
signs of anorexia		142
free medical advice		135
xanor	171	133
hypnagogic hallucinations	152	130
autism symptoms	120	122
anorexia statistics	102	115
sexual education in schools		113
media and body image		111
grief stages	89	106
kleptomania		101
mood swings	174	95
sleeping problems	302	
symptoms of panic attacks	146	
stomach stapling	130	
gastric banding	106	
causes of anorexia	103	
psychotic depression	90	
Cipramil side effects	81	

Table 8: Natural language question answering system usage absolute numbers

		sep-03
faqById	Lookups from list of all FAQs	4585
simpleQuestions	Intelligent questions	5630
	Pushing the "Find more related FAQs" button which causes a search in the natural-language question-answering system with the current FAQ title as query	2035
anyWordQuestions	Questions using the any word option	1210
allWordsQuestions	Questions using the all words option	44
success	Successfully answered question in the opinion of the FAQ software	6862
failure	Questions which could not be answered at all (might be questions on somatic medicine or irrelevant questions like "What time is it")	1355
onlyRelated	Questions where only related answers were found (might be good answers, or might not be good answers)	939

		sep-03
unfamiliarWords	Questions containing words not known to the FAQ system (might be good answers, or might not be good answers)	591
replyToQ	Total number of user ratings	187
replyToQnone	User rating = No opinion	12
replyToQno	User rating = Did not tell what I wanted to know	88
replyToQsomewhat	User rating = Partly what I wanted to know	50
replyToQwell	User rating = Well answered what I wanted to know	24
replyToQverywell	User rating = Very well answered what I wanted to know	13
replyToQcomment	User wrote a text string explaining the user rating	39

Table 9: Natural language question answering system usage in percentages

	Sep-03
Percentage of questions which the FAQ system believed were successfully answered	79,8 %
Percentage of questions where the FAQ system did not find any answer at all	15,8 %
Percentage of questions where the FAQ system found only answers it believed to be related	10,9 %
Percentage of questions which contained words not known to the FAQ system	6,9 %
Percentage of questions where user provided a rating of at least one answer	2,2%
User rating = Did not tell what I wanted to know	47,1 %
User rating = Partly what I wanted to know	26,7 %
User rating = Well answered what I wanted to know	12,8 %
User rating = Very well answered what I wanted to know	7,0 %

Table 10 Entry pages 1-12 November 2003, very approximate figures

Category	No. of entry hits	Percent
Home page	3397	7,8 %
Individual specific static FAQ pages	27223	63,0 %
Submenu pages	1690	3,9 %
Forum system home pages	8747	20,3 %
Forum system individual answers	1624	3,8 %
Internalpages only open to project partners	202	0,5 %
Project description pages	300	0,7 %
Total	43183	100,1 %