

Grant Agreement Number: 257528

KHRESMOI

www.khresmoi.eu

Use case definition including concrete data requirements

Deliverable number	<i>D8.2</i>
Dissemination level	<i>Public</i>
Delivery date	<i>29 of February 2012</i>
Status	<i>Final version</i>
Author(s)	<i>Celia Boyer (HON), Manfred Gschwandtner (GAW), Allan Hanbury (TUW), Marlene Kritz (GAW), Natalia Pletneva (HON), Matthias Samwald (TUW), Alejandro Vargas (HON)</i>



This project is supported by the European Commission under the Information and Communication Technologies (ICT) Theme of the 7th Framework Programme for Research and Technological Development.

Abstract

This document illustrates use cases of how layman users and physicians would use Khresmoi for an online health information search. These user profiles were drawn based on the survey results described in D8.1.1 and D8.1.2, and typical scenarios were created for both of them. The use cases focus on searching for information about issues related to diabetes and pneumonia. This document describes the user profiles, scenarios, as well as the use of different tools to assist users in the search process. Based on these scenarios, 6 types of searches have been identified: two typical for members of the general public and four typical for physicians. According to these types, various requirements were identified in respect to data sources, interface, representation of results, and multilingual/translation features to further advance the work of technical work packages within the Khresmoi project.

Table of Contents

1	Executive summary	5
2	Introduction	6
3	Physician use cases	8
3.1	The path towards use case categorization.....	8
3.2	The choice of the physician use case topic: LADA diabetes (Latent Autoimmune Diabetes of Adults) and UIP (Usual interstitial pneumonia).....	10
3.3	Physicians user scenarios.....	11
3.3.1	Use case 1: Self-employed general practitioners (SEGPs): To find treatment information on a new drug to treat diabetes	11
3.3.2	Use case 2: Self-employed specialists (SEs): To find information on diagnosis of LADA diabetes	14
3.3.3	Use case 3: Physicians in training (PITs): To find the meaning of an abbreviation, obtain descriptive information, and do an online CME course	17
3.3.4	Use case 4: Research physicians (RPs): To prepare a presentation on the most recent advances in diagnosis and treatment.....	20
3.3.5	Use case 5: Hospital clinicians (HCs): To confirm the diagnosis proposed by a radiologist	23
3.4	Categorization requirements for physician use cases	25
3.4.1	Document level annotation.....	25
3.4.2	Paragraph level annotation	26
3.5	Preliminary analysis of a PubMed query log.....	27
4	The general public use cases.....	28
4.1	The path towards use case categorization.....	28
4.2	Interviews with representatives of the general public	29
4.3	General use tools.....	31
4.3.1	Initial choice of language/country	31

D8.2 Use case definition including concrete data requirements

4.3.2	Spelling correction and query completion	31
4.3.3	“Conversion” into medical terms where applicable	32
4.3.4	Definitions of medical terms	32
4.3.5	Preview of results	32
4.3.6	Consult search history	32
4.3.7	Medical thesaurus	32
4.3.8	Risk factor assessment tools	32
4.3.9	Body 3D visualization	33
4.3.10	Filtering by images, audio and video files	33
4.3.11	Filtering by relevancy or recency	33
4.3.12	Recommended by Khresmoi search results	33
4.4	Translation and multilingual tools	33
4.5	Registration	34
4.6	Categorization requirements	35
4.6.1	Document level annotation	35
4.6.2	Paragraph level annotation	37
4.7	Preliminary HON query log analysis	38
4.8	Scenarios	39
4.8.1	First scenario: To find the information about a diet during gestational diabetes	39
4.8.2	Second scenario: To find the information about a blood test	41
4.8.3	Third scenario: To find information about cinnamon in treating type 2 diabetes	42
4.8.4	Fourth scenario: Patient with complications of type 2 diabetes: to find information about the side effects of a drug	44
4.8.5	“Rainy day” scenarios	46
5	Conclusions	47
5.1	Key messages of physician use cases	47
5.2	Key messages of the general public use cases	48
5.3	Summary of use case requirements	49
6	Acknowledgements	53
7	References	53
8	Annex	54
	Annex 1. Physician user profiles (based on questionnaire results)	54
	Annex 2. General public user profiles (based on detailed survey analysis)	66
	Annex 3. Set of questions for semi-structured interviews of general public	69
	Annex 4. HON logs analysis	70
1	Introduction	72
2	HON-log-analysis Description	72
2.1	Development Platform	73
2.1.1	Environment	73

D8.2 Use case definition including concrete data requirements

2.1.2	Team sharing.....	73
2.1.3	Programming Language	73
2.1.4	Database management system	73
2.2	Data Collection	73
2.3	Data preparation	74
2.3.1	Parsing, Cleaning the data	74
2.3.2	Database System	75
2.4	Data Analysis	75
2.4.1	Term analysis	75
2.4.2	Query analysis.....	77
2.4.3	Analysis by country origin	78
2.4.4	Number of queries per day	79
2.4.5	Countries by term	79
2.4.6	Number of queries and details by user	80
3	Conclusions	82
4	References	83
	Annex 5 Useful data sources for physicians	84

List of abbreviations

CT	Computer tomography
GAW	Gesellschaft der Ärzte in Wien (Society of Physicians, Vienna)
HC	Hospital clinician
HON	Health on Net Foundation, Geneva
LADA	Latent Autoimmune Diabetes of Adults
NSIP	Non-specific interstitial pneumonia
RP	Research physician
SEGP (GP)	Self-employed general practitioner
SES	Self-employed specialist
TUW	Technischen Universität Wien (Vienna University of Technology)
UIP	Usual interstitial pneumonia

1 Executive summary

This deliverable aims to define use cases and provide a list of concrete data and interface requirements for two users groups: physicians and the general public. The use cases were created in three main steps: categorization of user profiles, identification of user scenarios, and the verification of this identification. Most of the user profiles and use cases presented in this deliverable are based on a survey described in the Deliverables D81.1 and D8.1.2. However, certain aspects of these user profiles and use cases proved to be very difficult to identify due to methodological restrictions of online-based surveys. To obtain additional, quantitative insight, we conducted additional studies: query log files analysis and interviews. The results obtained through these additional studies helped us test the use cases in real-life circumstances.

Categorization: First, the detailed statistical analysis of the survey results described in D8.1.1 and D8.1.2 helped to identify user profiles with unique patterns of search behaviour and information need. The analysis of the physician questionnaire demonstrated that search behaviour and information need was primarily determined by the level of specialisation, status of employment, and level of qualification. As a result, five distinct user groups (self-employed general practitioners, self-employed specialists, hospital clinicians, physicians in training, research physicians) were identified. An overall trend was that all physicians preferred to access a secured professional platform or at least to filter information to exclude laypeople content and sought colleague feedback when having to choose what information was relevant and trustworthy (physician quality rating, relevance rating). These user groups were found to differ in the following aspects: level of search, expected resources, location of search, role of multilinguality, level of access (mobile device vs. fixed device), priority of output ranking (last update, simplicity, quality rating), and digital social support. In respect to general public users, several dimensions were identified. It's important to know the age and gender in order to filter the results and customize the interface requirements. Information about personal health status is required in order to propose the most appropriate content in respect to the user's level of health literacy.

Identification: The second step involved implementation of user requirements within the context of real-life search scenarios for each of the proposed groups. A link between the general public and physicians was ensured by the choice of an overlap in one of the chosen medical topics; diabetes. Furthermore, the second topic of the physician use cases, usual interstitial pneumonia (UIP), highlighted the importance of image search for non-radiologists.

Verification: After this, the user scenarios were verified and validated using semi-structured interviews with physicians and the general public. Interviews helped to gain some qualitative insight on the topic whilst ensuring that the proposed scenarios had validity in real life. Preliminary results of the work done on HON and PubMed query log analysis have been also taken into account. Further work on logs analysis is foreseen.

Based on the nine use cases described in the document, we have identified six types of searches: two performed by the general public and four by physicians: basic and advanced search by the general public, wide vs. narrow, research and CME search typical for different groups of physicians. A table listing the specific requirements for each type of search was constructed to provide a coherent overview of the use case requirements in respect to data sources, interface, representation of results, and multilingual/translation features.

2 Introduction

Identification of profiles of future users of Khresmoi, their current online search behaviour, and their possible interactions with the Khresmoi search engine are important aspects in guiding the work of technical work packages within the project. The main goal was to identify typical information needs, data source requirements, results presentation, interface and translation tool requirements, as well as possible tools that could be helpful in facilitating the search process in a range of possible real-life scenarios. Results of two large-scale surveys, conducted in the summer of 2011, analysing information needs and online search behaviour of physicians (D8.1.2), and the general public (D8.1.1), have provided the major basis for identification of user profiles. An extensive analysis of the survey results, follow-up interviews, and query analysis enabled the construction of different user scenarios for each of the different user groups. Eventually, we created five different user profiles and corresponding “typical scenarios” for physicians as well as four different user profiles and “typical scenarios” for members of the general public.

In order to keep the main focus on search strategies and user requirements as opposed to the medical topics themselves, all except one of the physician user scenarios (hospital clinicians) stuck to the common medical topic of diabetes. The choice of diabetes as a topic is based on the following:

1. Diabetes is one of the biggest public health threats worldwide. According to WHO Europe, there are approximately 60 million people in the European region currently living with this disease (1). Prevalence of diabetes is increasing, and in some member states within the European region the disease is already affecting 10-12% of the population (1). Thus, both the epidemiological significance and the worldwide economic impact of the disease are making diabetes a central topic within the public health domain (2, 3).
2. A great variety of online health content is dedicated to diabetes, including general consumer websites, personal health applications, blogs, continuous research, and publications of results into the presence of contradicting treatment concepts.
3. As a long-term chronic disease, diabetes encompasses a vast majority of potential users’ interactions while searching for online health information and communicating with their peers.
4. For the general public, diabetes is a topic of general interest due to its prevalence which may lead to active health information seeking behaviour, and in case of diagnosis, to further investigative research tasks.
5. For physicians, diabetes is a major topic on all levels of specialization. While general practitioners are involved in both the diagnosis and treatment of diabetes, specialists are primarily concerned with difficult cases and the wide array of associated tertiary complications.

The idea was to emphasize how one medical topic is perceived and dealt with from the perspective of different types of user groups. An individual who is affected by diabetes may use the Internet to seek for information about the treatment and accompanying lifestyle interventions (e.g. diet). He/she may carry out additional research on the use of homeopathic substances to assist his/her health condition, or he/she may communicate with peers on everyday issues. In contrast, a general practitioner may use the internet to seek general or advanced treatment information, while the specialist may use the internet to obtain specialized information on atypical cases of diabetes. On the other hand, a research physician may seek recent primary literature to prepare a presentation on the latest developments in research. A physician in training could, at the same time, look for a CME course and require a medical dictionary

D8.2 Use case definition including concrete data requirements

to make sense of unknown abbreviations used in medical reports. Different needs and search behaviour within different types of user groups are highlighted in this manner.

One physician use case, hospital clinicians, was assigned a different topic, UIP (Usual interstitial pneumonia), in order to highlight the need for image search amongst non-radiologists. While radiologists are expected to primarily search on an image level, the need of image search amongst non-radiologists should not be underestimated. We have described a case of a pulmonologist looking for typical cases of UIP in order to verify the diagnosis made by a radiologist. Other specialists, such as dermatologists, neurologists as well as the general practitioners, may be interested in image search on a more sophisticated level.

The structure of use case scenarios was inspired by the work published by Cockburn A. (2001) on writing effective use cases (4). Each use case scenario includes information about the scenario itself, user specific characteristics (pre-condition, log-in status, willingness to pay, context, time available, and language skills), goal/intention of the search, ideal search process (link description, automatic categorization, manual categorization, tools, and suggested websites), successful search scenario and alternative scenarios (unsuccessful scenario and suggested settings for a log-in scenario). In addition, further information on examples of useful data resources for each use case scenario is provided.

Deliverable 8.2 is structured in the following way:

Part 4 describes the use cases of physicians. The first section describes the categorization of user profiles, the choice of scenario topics, and the search scenarios of different types of physicians with respect to data requirements. The second section describes the query log analysis and interviews to verify user scenarios on a qualitative level.

Part 5 describes the use cases of the general public. Firstly, the different kinds of tools to be used by the members of the general public are proposed, the results of interviews and observations of members of the general public on the topic of online health information search are reported. Secondly, the user scenarios of different types of users are detailed.

Part 6 contains conclusions and main insights of the use cases in the form of a summary table of requirements to guide further work of the technical work packages within the project.

Part 7 is an appendix including detailed tables of the survey results providing the basis for user categorisation and the expected search behaviours and needs, as well as a detailed description of query log analysis.

3 Physician use cases

3.1 The path towards use case categorization

Whilst screening the data of the physician questionnaire (see Deliverable 8.1.2) for significant trends and differences, we noticed that the most prominent predictors of search behavior and information need were a) employment status (e.g. employed vs. self-employed), b) level of experience (e.g. physician in training vs. research physician), and c) level of specialization (e.g. general practitioner vs. specialist). We found that, for example, physicians in training primarily sought information to fulfill an immediate need (i.e. requiring an answer within 10 minutes) while research physicians were more concerned with information on an educational level (e.g. to prepare a presentation). The level of specialization was found to determine the type of resources physicians expected to access. For example, it was found that GPs mainly consulted secondary resources while specialists were drawn towards more specialized resources (e.g. to go to targeted society websites or to seek primary literature). In addition, employment status appeared to determine to what extent physicians sought the help of a “digital colleague” in the process of choosing relevant, trustworthy information. Our findings suggested that self-employed physicians were more likely to use physician society communities (despite the lowest awareness), post questions in medical forums, and use trustworthiness of a link as a choice criterion.

With respect to these “trends”, we identified five groups of physicians:

- 1) Self-employed general practitioners (SEGPs)
- 2) Self-employed specialists (SESSs)
- 3) Hospital clinicians (HCs) – including qualified physicians who reported being employed in a setting outside a university (hospital, private or public health care setting)
- 4) Research physicians (RPs) – including qualified physicians who reported working in a university health care setting and/or who had obtained the level of professorship/habilitation
- 5) Physicians in training (PITs) – including physicians pursuing specialty training or final year medical students

Detailed profiles are described in Annex 1. The following profiles are described based on the survey results.

Self-employed general practitioners (SEGPs):

Crucial information overshadowed by laypeople information, help in information selection required, direction towards secondary sources needed. General practitioners are usually time-constrained (i.e. usually have a maximum of 10 minutes devoted to do a search) and seek broad-surfaced information on various topics. Their main problem with current general search engines is that they are overwhelmed with laypeople information and don’t have help/time/skill to filter out trustworthy quality information. According to the SEGPs opinion, current medical search engines either focus on primary resources (e.g. PubMed), lack local resources in the user’s mother tongue (e.g. Best Practice), or are focused on a subspecialty of medicine (e.g. Yottalook). General practitioners usually have simple queries and are primarily interested in secondary and local resources providing simple, practice-orientated information (e.g. articles, guidelines, etc.). A general medical level is sufficient, and an overview is essential. For more specific cases general practitioners usually transfer their patients to specialists. The search engine must help them to pursue a quick search which leads to

D8.2 Use case definition including concrete data requirements

applicable, professional answers. General practitioners prefer a simple layout with a maximum of 10 links per page. They would like to filter information based on target group (i.e. to view physician information only), type of source (to view secondary resources only), country (to view local resources only), and date range. A good link description (indicating source, type of content, and quality rating) helps the SEGP to determine whether the link is relevant to the query and of good quality. General practitioners are likely to profit from translation tools and regard a “search help”, such as automatic completion of query and suggested links (where other users have found the answer), very helpful. They will be the physicians who primarily benefit from viewing ratings of quality and trustworthiness as they usually don’t have time and expertise to rate the content. They will also be the physicians who are likely to seek a “digital colleague” via a medical forum to discuss patient cases and to obtain collegial support.

Self-employed specialists (SES):

Search for specialized, practice-oriented secondary sources: Self-employed specialists are primarily interested in specialized information within their field of medicine. General search engines are sufficient for providing answers to definitional questions but too broad for answering specialized queries. (e.g. “What is the treatment for condition X in the case of Y?”) Primary resources either fail to answer practice-oriented queries in the time available, or are not accessible. Like SEGPs, at the point-of-care, SESs are primarily interested in local resources providing practice-oriented information (e.g. articles, guidelines etc.) in their mother tongue. However, the SESs have a more specialized approach to search, and their queries are more complex than those of the SEGPs. At the point-of care, the SESs primarily seek resources which provide in-depth information within their field of medicine (e.g. society websites, university websites). An overview of other fields of medicine is not necessary but an “in-depth level” within their field of medicine is essential. They seek primary sources, preferably summarized primary sources (e.g. Cochrane Reviews) as they are required to keep up-to-date in their field of medicine. The search engine should be able to help them execute a quick but comprehensive search. They prefer a simple layout with a maximum of 20 links per page. They would like to filter information based on target group (i.e. to view physician information only), type of content (to view information in their field of medicine only), type of source (to view both primary and secondary sources, depending on the case), on occasions country (to view local resources at the points-of-care), and date range (to ensure being up-to-date in their field). Like the SEGPs, they profit from a good link description to determine the quality of the source. However, they are more likely to rate sources (as experienced experts) themselves and to have good skills in evaluating reliability. They will profit from ratings of information sources outside their field of work or knowledge. Like the SEGPs, they profit from a “search help” in terms of language and navigating on the Web. Like the SEGPs, they are “alone” and also physicians who are likely to seek a “digital colleague” via a medical forum to discuss rare patient cases.

Physicians in training (PITs):

Requirement to compensate lack of experience and fill information gaps at point-of-care, obtain social support via forums, pursue CME to advance levels of medical knowledge: Physicians in training usually have high requirements for often very basic immediate/definitional but also specialized educational information. Basic resources, such as the Wikipedia, help them to solve most of their daily questions in the medical field. For more complex queries, an access to medical forums/discussions within a secured platform, as well as a targeted search for CME courses might be useful.

Research physicians (RPs):

D8.2 Use case definition including concrete data requirements

The research physicians who participated in the survey have been found to be mainly interested in primary literature (e.g. PubMed, DocGuide, Cochrane Reviews) that helps them to further their research and to prepare for presentations. Their English language skills are usually good (the best of all physicians in this survey), and they are interested in accessing both local and global information. Like the self-employed specialists and the hospital clinicians, the RPs are in “expert” positions and have skills to rate websites and thereby help other time-constrained physicians to judge the quality of sources.

3.2 The choice of the physician use case topic: LADA diabetes (Latent Autoimmune Diabetes of Adults) and UIP (Usual interstitial pneumonia)

In order to establish a valid real-life scenario, two important topics were chosen for the physician use cases: LADA diabetes and UIP. Diabetes was chosen to create a link with the laypeople use cases, while UIP was chosen to highlight the importance of image search of non-radiologists. The restriction to LADA diabetes for four of the physician use cases enabled us to construct scenarios viewed from different angles of user groups. LADA diabetes is a relatively unknown form of a “slowly progressive”, eventually insulin dependent form of diabetes of increasing prevalence and often misdiagnosed as type 2 diabetes due to its late onset and initially successful treatment with oral medication (5). It has been estimated that as much as 20% of people diagnosed with non-obesity-related type 2 diabetes may have LADA (6). Misdiagnosis and lack of information amongst physicians can lead to inappropriate and ineffective treatment procedures. Thus, information on LADA diabetes is essential for both the general practitioner and the specialist.

To establish a connection with the radiology use cases and to emphasize the importance of image search amongst non-radiologists, we have additionally described one use case (hospital clinicians) in the field of pulmonology. Usual interstitial pneumonia (UIP) is a form of interstitial lung disease with unknown cause, which develops due to progressive damage of the lung tissues (6). The condition is usually diagnosed via a CT scan by a radiologist. However, in many cases a pulmonologist is required to provide the final diagnosis as he/she has more profound expertise in the topic and patient contact. The presence of UIP can have significant impact on treatment procedures and prognosis. Thus it is essential that careful and correct diagnosis is made. Consequently, accessing image search in evaluating the diagnosis of a radiologist may be helpful for the pulmonologist.

Verification of categorization, use case topics and search behaviour:

The search processes described as part of the user scenarios were written and based on the physician questionnaire results. In order to verify the validity of proposed scenarios, topics, and search behavior, we conducted informal, semi-structured interviews with three physicians: 1) a self-employed specialist of pulmonology, 2) an employed research physician and a specialist of internal medicine, and 3) a self-employed general practitioner. Each interview lasted about 45 minutes. The aim of the interviews was to evaluate whether the chosen topics represented real problems in daily practice, and whether the use case scenarios had appropriate levels of validity. The interviews led to minor improvements on all levels of the user scenarios.

3.3 Physicians user scenarios

Each user group described in Section 2.1 was assigned its own use case scenario to emphasize different forms of search behavior and information needs amongst physicians. The purpose was to highlight the specific behavior of different user groups in the context of real-life scenarios.

3.3.1 Use case 1: Self-employed general practitioners (SEGPs): To find treatment information on a new drug to treat diabetes

A patient diagnosed with type 2 diabetes taking metformin has done a blood test determining his/her HbA1C status and is seeing a SEGP to talk about the blood test results and to discuss his/her health and maintenance/change of medication/treatment procedure. The SEGP skims through the medical report during the consultation and notices that the HbA1C levels are still rising despite the patient having been prescribed several types of oral medications. The SEGP wants to find a way to reduce the HbA1C levels and needs help in finding a new treatment/medication. He/she opens Khresmoi during the consultation while the patient is present.

Willingness to pay: The SEGP would be prepared to log in and pay a membership fee if this enabled him/her to access unbiased, quality-rated information and a secured platform where he/she can exchange information and/or communicate with other physicians.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: In his/her office while the patient is present, accessing the Internet from his/her laptop computer.

Time available: 5-10 minutes (i.e. immediate need).

Intention: To find easy-to-read, unbiased, trustworthy, practice-oriented information in guideline/recipe format on a new oral medication for type 2 diabetes.

Language: The mother tongue of the GP is German and he/she has average skills in medical English.

Search process:

The search bar: The SEGP types into the search bar (“Orale Therapie, Diabetes mellit.”) Khresmoi offers an automatic completion of the query. Khresmoi recognizes the query by translating the language internally to English (Oral Therapy, Diabetes mellitus).

Search output:

- **Number of links and layout :** Khresmoi delivers a search output containing a maximum of 10 links per page. The interface is simple and easy to understand. Advertisements are tolerated to keep the search engine free-of-charge.
- **What is shown first? Automatic ranking:** Khresmoi can detect the country of login. The search output is automatically ranked by **relevance** (i.e. number of search terms matching the query) and **language**. The most relevant and information in the language of query input (i.e. German in this case) is ranked first.
- **What is shown in the link? Automatic categorization/link description:** The link descriptions contain a) type of content, author, b) publisher, c) target audience, d) quality

D8.2 Use case definition including concrete data requirements

accreditation marks (if available), e) physician quality rating, f) publisher, g) time of the last update, h) indication whether the content is of restricted or free access, i) type of media (e.g. video vs. text).

- **Manual categorisation/Filter:** Khresmoi offers the possibility to filter the data based on the type of source (primary vs. secondary source vs. forums vs. news, CME courses), type of media (e.g. video vs. image vs. text vs. audio), type of content (e.g. diagnosis, treatment etc.), language (e.g. only pages in German), target audience (laypeople vs. physician) websites from Austria only (local information). The SEGP filters the information based on the type of source (secondary only), type of content (treatment), target audience (physician information only), and limits the results to websites from Austria to obtain information that contains local, secondary text resources for physicians.
- **Advanced search options/Ranking:** Khresmoi offers the possibility of an advanced search where the user can choose to rank the information based on date range (i.e. time of publication/last update), physician quality and relevance ratings (i.e. how many search terms match the query), level of accessibility (freely accessible websites to be displayed first). The SEGP ranks the information by date range, level of accessibility and physician quality ratings to find free, the most recent, and the most trustworthy information.
- **Tools:** Khresmoi offers the following tools which are relevant to the SEGP: automatic translation, dictionary (including explanation of abbreviations), integration of patient data within the search process as a diagnostic aid for complex cases. The SEGP skims through the second and third page, if necessary, and clicks on the link that appears to be most relevant and trustworthy.
- **Suggested relevant links:** Khresmoi suggests relevant/related topics and websites which users with similar queries have looked at.
 - The SEGP clicks on the suggested link of the FDA¹ which warns the SEGP that Sitagliptin should be used with caution and with appropriate monitoring of patients with a history of pancreatitis.
 - The SEGP clicks on the link about “Lifestyle change of diabetic patients – what a patient can do” and sees that the page has been rated to be of good quality by other physicians and has high “usefulness rating” by affected patients. The SEGP therefore decides to write down the link and the name of the medication. After gathering information about the medication, the SEGP checks his/her medical software database to check whether the medication is available and how much it costs.

¹

<http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHealthcareProfessionals/ucm183764.htm>

D8.2 Use case definition including concrete data requirements

Successful scenario: The SEGP is able to find information on diabetes treatment from a trustworthy (academic) website containing information about Sitagliptin, a new medication that has shown to be effective in reducing the HbA1c levels of diabetic patients. Khresmoi tracks the search history.

The SEGP makes sure that the patient does not have a history of pancreatitis, then prescribes the medication and provides the patient with the link to help him/her find better self-help strategies.

Alternative scenarios:

Unsuccessful scenario: The SEGP does not find the information

- The SEGP does another search using more specific search terms (e.g. new treatment, oral medication, diabetes mellitus).
- If again no relevant answer shows up, the SEGP stops the search due to the time constraint of 10 minutes. The SEGP chooses to send the patient to a specialist of internal medicine for further consultation.

Login scenario:

Khresmoi recognizes that the physician is a general practitioner (automatic settings for general practitioners are activated)

Automatic settings for general practitioners are (and can be manually adjusted by user):

- **Automatic ranking:** Local, practice-oriented information, secondary sources, guidelines/general health websites first, information in mother tongue first.
- **Filter:** Availability of important content filters such as “only drug information”, or “physicians in the area”.
- **Medical news:** Possibility to make comments and contribute to discussions.
- **Suggested relevant links are based on the type of content accessed in search history.**

Examples of useful Data sources for this use case²:

Drugs

- Austria Codex – Austrian Drug Database³
- Drug Safety of Sitagliptin⁴

Medical news

- Medical news⁵

² For an extended list of resources please see Annex 5

³ http://www.univadis.at/medical_and_more/de_AT_Login

⁴ <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHealthcareProfessionals/ucm183764.htm>

⁵ http://www.univadis.at/medical_and_more/de_AT_Login

D8.2 Use case definition including concrete data requirements

- Austrian magazine for general practitioners⁶

Physician Directory

- German Diabetic Society⁷
- List of hospitals/Institutions in Germany Specialized Diabetes⁸

3.3.2 Use case 2: Self-employed specialists (SEs): To find information on diagnosis of LADA diabetes

A diabetic patient has been transferred to the self-employed specialist (diabetologist) by the SEGP who did not know how to help the patient anymore, as his/her HbA1C levels were rising despite the previously prescribed medication. The SES needs to find out why the patient is no longer responding to the medication and/or what alternative treatment could possibly help him/her. The SES considers the possibility of LADA diabetes. He/she opens Khresmoi to find out which blood tests would indicate whether the patient suffers from the aforementioned condition. He/she uses Khresmoi and hopes to find information on the diagnosis and treatment of LADA diabetes or associated conditions of oral non-responders. Khresmoi should direct him to specialized secondary sources providing information on diagnostic procedures of LADA diabetes. Khresmoi should help him decide what tests to do in order to check whether the patient suffers from LADA diabetes.

Willingness to pay: The SES would be prepared to log in and pay a membership fee if that enabled him/her to access quality-rated information and a secured platform where he/she can exchange information and/or communicate with other physicians.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: In his/her office while the patient not is present, accessing the Internet from his/her PC.

Time available: 20 minutes (i.e. not immediate need) after the patient has given blood for testing and been sent home (and asked to come again next week).

Intention: To find out why the patient is not responding to the oral medication and which blood tests are suitable in checking whether the patient suffers from LADA diabetes. The physician is primarily interested in specialized secondary sources with guidelines for diagnosis and summarised primary sources to provide information on the current state-of-the-art diagnosis and therapy.

Language: The mother tongue is SES German and he/she has average skills in medical English. He/she prefers to access and type medical queries in German and can understand good medical English but finds automatic translation tools and an automatic completion of queries helpful.

Search process:

The search bar: The SES types the query in German into the search bar (Therapieversagen bei normalgewichtigen Typ 2 Diabetes Insulinpflichtig?). Khresmoi offers an automatic completion of the query and translates the query to English. English translation: Failure of treatment for normal weight type 2 diabetics insulin therapy?

⁶ <http://www.aerztemagazin.at/dynasite.cfm?dssid=4169&dszsid=246>

⁷ <http://www.deutsche-diabetes-gesellschaft.de/>

⁸ <http://www.deutsche-diabetes-gesellschaft.de/>

D8.2 Use case definition including concrete data requirements

Search output:

- **Number of links and layout?** Khresmoi delivers a search output containing a maximum of 10 links per page. The interface is simple and easy to understand. Advertisements are tolerated to keep the search engine free-of-charge.
- **What is shown first? Automatic ranking:** Khresmoi can detect country of login. The search output is automatically ranked by **relevance** and **language**. The language of query input (i.e. German in this case) is ranked first.
- **What is shown in the link? Automatic categorization/link description:** The link descriptions contain a) type of content, author, b) publisher, c) target audience, d) quality accreditation marks (if available), e) physician quality rating, f) publisher, g) time of the last update, h) indication whether the content is of restricted or free access, i) type of media (e.g. video vs. text). Publisher and time of the last update helps the SES to evaluate the quality and/or the usefulness of the link.
- **Manual categorisation/Filter:** Khresmoi offers the possibility to filter the data based on the type of source (primary vs. secondary source vs. forums vs. news, CME courses), type of media (e.g. video vs. image vs. text vs. audio), type of content (e.g. diagnosis, treatment etc.), language (e.g. only pages in German), target audience (laypeople vs. physician) websites from Austria only (local information). The SES filters the information based on the type of content (diagnosis) to obtain “quick-to-read, recent information (preferably in German) aiding the diagnosis of LADA diabetes” (e.g. Wikipedia, up-to-date).
- **Advanced search options/Ranking:** Khresmoi offers the possibility of an advanced search where the user can choose to rank the information based on date range (i.e. time of publication/last update), physician quality and relevance ratings (i.e. how many search terms match the query), level of accessibility (freely accessible websites to be displayed first). The SES ranks the information using date range and quality ratings to find the most recent, trustworthy information on the topic in German (which in this case is easier to comprehend).
- **Tools:** Khresmoi offers the following tools which are relevant to the SES: automatic translation, dictionary (including explanation of abbreviations), integration of patient data within the search process as a diagnostic aid for complex cases. The SES uses the automatic translation of results to obtain the translation of sentences and/or words which are too complex to understand, and/or the medical dictionary for terms and/or abbreviations that are unfamiliar to him/her. Khresmoi tracks the search history.
- The SES skims through the second and third pages, if necessary, and clicks on the link that appears to be the most relevant and trustworthy.

Successful scenario: The SES finds diagnosis information which is recent, rated to be of good quality, from a trustworthy (academic) publisher, and containing descriptive information (primary and secondary sources) about LADA diabetes and ways to diagnose it (i.e. indicating what tests are needed). After gaining the information about the possibility of LADA diabetes, the SES sends out the patient’s blood samples for laboratory analysis in order to confirm/reject the diagnosis.

Alternative scenarios

Unsuccessful scenario: The SES does not find the information

- a) The SES does another search using more specific search terms.

D8.2 Use case definition including concrete data requirements

- b) If again no relevant answer shows up, the SES logs in and posts the question in a medical forum.
- c) If there is no answer, the SES consults other sources of information.

Login scenario:

Khresmoi recognizes that the physician is an SES (automatic settings for SESs are activated)

Automatic settings for self-employed specialists are (can be manually adjusted by user):

1. **Automatic ranking:** Local, practice-oriented information first, field-specific information (e.g. specialized and society websites) first, information in mother tongue first.
2. **Filter:** Availability of important content filters such as “only information on diabetes/for specialists of internal medicine”
3. **Medical news:** Possibility to make comments and contribute to discussions
4. **Suggested relevant links** are based on search history

Examples of useful Data sources for this Use case⁹:

Specialised Guidelines for Treatment and Diagnosis of Diabetes

- OEDG Guidelines German¹⁰
- OEDG Guidelines English¹¹
- International Diabetes Federation - Guidelines¹²

Secondary and tertiary resources for general information

- Springer medizin¹³
- Medizin Medien¹⁴
- Universimed¹⁵
- Univadis¹⁶
- Docguide¹⁷
- WebMD¹⁸

⁹ For an extended list of resources please see Annex 5

¹⁰ http://www.oedg.org/oedg_leitlinien.html

¹¹ http://www.deutsche-diabetes-gesellschaft.de/redaktion/mitteilungen/leitlinien/Uebersicht_engl_LL.php

¹² <http://www.idf.org/guidelines/guidelines-development>

¹³ <http://www.springermedizin.de/>

¹⁴ <http://www.medicin-medien.at/>

¹⁵ <http://universimed.com/>

¹⁶ <http://www.univadis.at/>

¹⁷ <http://www.docguide.com/general-practice/popular/30days>

- Medscape¹⁹
- Wikipedia²⁰

3.3.3 Use case 3: Physicians in training (PITs): To find the meaning of an abbreviation, obtain descriptive information, and do an online CME course

A physician in training has just started his/her training at a department of internal medicine. During an afternoon colloquium, a case of a patient diagnosed with LADA diabetes is discussed. The PIT comes across this relatively unknown condition for the first time. On the way home, the physician uses Khresmoi to quickly inform himself/herself of the meaning of the abbreviation and disease description, and wants to attend a CME (video) course on the topic in order to be prepared for the next day at work (i.e. for the morning colloquium).

For this particular use case, the possibility to filter information based on accredited, available CME courses is essential and might make him prefer Khresmoi over current solutions.

Willingness to pay: The physician in training is unlikely to pay and is looking for free-of-charge content on the topic but is not disturbed by advertisements.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: During travel, accessing the Internet from a tablet PC or a smart phone.

Time available: 40 minutes (i.e. educational need) while travelling.

Intention: To find definitional and descriptive information on the topic of LADA diabetes (the definition of the abbreviation) in the form of text and/or video.

Language: Mother tongue of the PIT is German and he/she has average skills in medical English. He/she prefers to access and type medical queries in German and can understand good medical English but is often unfamiliar with certain terms. He/she finds automatic translation tools and an automatic completion of queries and a medical dictionary helpful.

Search process:

Mobile device adaption: The PIT opens Khresmoi. Khresmoi detects that the PIT is accessing from a mobile device and automatically adjusts to the mobile device friendly format.

The search bar: The PIT types into the search bar “LADA diabetes“. Khresmoi offers an automatic completion of the query.

Search output:

¹⁸ <http://www.webmd.com/>

¹⁹ <http://www.medscape.com/>

²⁰ http://en.wikipedia.org/wiki/Main_Page

D8.2 Use case definition including concrete data requirements

- **Number of links and layout?** Khresmoi delivers a search output containing 20 links per page. The interface is simple and easy to understand. Advertisements are tolerated to keep the search engine free-of-charge.
- **What is shown first? Automatic ranking:** Khresmoi can detect country of login. The search output is automatically ranked by relevance and language. The language of query input (i.e. German in this case) is ranked first.
- **What is shown in the link? Automatic categorization/link description:** The link descriptions contain a) type of content, author, b) publisher, c) target audience, d) quality accreditation marks (if available), e) physician quality rating, f) publisher, g) time of the last update, h) indication whether restricted accessibility or free content, i) type of media (e.g. video vs. text). Publisher and author help the PIT to evaluate the quality and/or usefulness of the link.
- **Tools:** Khresmoi offers the following tools which are relevant to the PIT: Automatic translation, dictionary (including explanation of abbreviations). The PIT uses a medical dictionary to obtain meanings of unknown words while watching the video.
- **Suggested topics:** Khresmoi suggests relevant/related topics and a CME video course on the topic of LADA diabetes. The PIT clicks on the link and then activates the filter “CME videos only” and local information, ranks the output by quality and time of the last update. The PIT then types “LADA diabetes treatment” and filters it by “local websites only” to find a recent CME on the topic which is locally accredited. The PIT finds a free-of-charge CME video course (free content is indicated in the link description) and he/she is able to rank (show free content first) the video and watch it.
- **Smartphone usability – Application filter:** The PIT is interested in obtaining medical applications about diabetes for his/her smartphone and uses a filter provided by Khresmoi to show “applications only”, types in diabetes and then ranks the output by “popularity” to obtain the most recent applications.

Successful scenario: The PIT finds a website which defines the meaning of LADA diabetes and offers some descriptive text, for example, a secondary descriptive source such as the Wikipedia.

Alternative Scenarios

Unsuccessful scenario: The PIT does not find the information

- a) The PIT does another search using more specific search terms, e.g. “definition of LADA diabetes”.
- b) If again no relevant answer shows up, the PIT stops the search, looks up the definition in a textbook, or asks a colleague.

Login scenario: Khresmoi recognizes that the user is a physician in training and accessing from a mobile device (automatic settings for PITs and mobile users are activated)

Automatic settings for the PIT users are:

Automatic ranking (can be manually adjusted by user):

D8.2 Use case definition including concrete data requirements

- Type of information: local information, general health websites, online CME.

Suggested links: Educational, local career websites, lead to definition, basic information resources (highest immediate need), smartphone-friendly format, video.

Language: Local information in mother tongue first.

Automatic settings for the mobile users are:

- Application filter: The user can search applications and rank them according to time of the last update and popularity.
- Adjusted clear screen presentation for mobile devices

Useful Data sources for this use case²¹:

Overall information and state of the art

- Springer medizin²²
- Docguide²³
- Wikipedia²⁴

App Resources

- Tapforms (App Database)²⁵
- Itunes Medical Apps²⁶

Online CME

- Gesellschaft der Ärzte Wien²⁷
- Medizin Medien²⁸
- Universimed²⁹
- Mein DFP³⁰
- Fortbildungskalender³¹

²¹ For an extended list of resources please see Annex 5

²² <http://www.springermedizin.de/der-latent-autoimmune-diabetes-im-erwachsenenalter-lada-eine-sonderform-des-diabetes/161384.html>

²³ <http://www.docguide.com/general-practice/popular/30days>

²⁴ http://en.wikipedia.org/wiki/Latent_autoimmune_diabetes

²⁵ <http://www.tapforms.com/>

²⁶ <http://itunes.apple.com/app/dg-apps/id445912798?ign-mpt=uo%3D5>

²⁷ http://www.billrothhaus.at/index.php?option=com_vf_elearning&task=dfp

²⁸ <http://www.medizin-medien.at/dynasite.cfm?dsmid=69263>

²⁹ <http://www.med-diplom.at/>

³⁰ <http://www.meindfp.at/>

³¹ http://www.der-fortbildungskalender.de/index.php?bundesland=&typ=online&datum=&datum_%20max=&Buchungscode=&action=search

- OEDG³²

3.3.4 Use case 4: Research physicians (RPs): To prepare a presentation on the most recent advances in diagnosis and treatment

The research physician is preparing a presentation on LADA diabetes and uses Khresmoi to find out about the most recent developments in terms of diagnosis and treatment. He/she is looking for recent papers and reviews to help him/her create a good presentation.

Willingness to pay: The RP is unlikely to pay since he/she already has access to most journals through the university.

Login status: The user is not logged in.

Precondition: Access to the Internet

Context: At home accessing from a tablet computer.

Time available: 40 minutes (educational need) while at home from a table PC.

Intention: To find scientific information in an organised format (ranked by date of publication and relevance) in order to prepare a presentation on the advances in the diagnosis and treatment of LADA diabetes.

Language: The RP has very good skills in medical English, he types medical queries in English. He/she can understand medical English but finds automatic translation tools and an automatic completion of query helpful.

Search process:

The search bar: The RP types into the search bar (“Advances in LADA diabetes”) Khresmoi offers an automatic completion of query: “Advances in LADA diabetes”

Search output:

- **Number of links and layout:** Khresmoi delivers a search output containing a maximum of 10 links per page. Advertisements are tolerated to keep the search engine free-of-charge.
- **What is shown first? Automatic ranking:** Khresmoi can detect the country of log in. The search output is automatically ranked by **relevance and language**. The language of query input (i.e. English in this case) is ranked first. Most relevant is defined as to what extent the search terms match the results or based on user relevance ranking.
- **What is shown in the link? Automatic categorization/link description:** The link descriptions contain a) type of content, author, b) publisher, c) target audience, d) quality accreditation marks (if available), e) physician quality rating, f) publisher, g) time of the last update, h) indication whether restricted accessibility or free content, i) type of media (e.g. video vs. text). Author, publisher, and time of the last update help the user to evaluate the quality and/or usefulness of the link.

³² <http://www.oedg.org/onlinevortraege.html>

D8.2 Use case definition including concrete data requirements

- **Tools:** Tools offered by Khresmoi: perceiving quality ratings (being able to rate through logging in), possibility to store self-made compendia, links where most users with a similar query have found the answer, integration of patient data within the search process, spelling correction, automatic translation, dictionary, possibility to view search history. If the most valuable/recent information is in English (RP switches the language filter on and off) the RP uses automatic translation to obtain translation of unknown words/texts while watching the video³³.
- **Advanced search options/Ranking:** Khresmoi offers the possibility of an advanced search where the user can choose to rank the information based on date range (=time of publication/last update), physician quality ratings and relevance (i.e. how many search terms match the query), level of accessibility (freely accessible websites to be displayed first). The RP ranks the information using date range and quality ratings to find the most recent, trustworthy information on the topic (like the application “doc guide”).
- **Manual categorisation/Filter:** Khresmoi offers the possibility to filter the data by the type of source (primary vs. secondary source vs. forums vs. news vs. books), type of content (e.g. diagnosis, treatment), type of media (e.g. video vs. image vs. text vs. audio vs. apps vs. ppt), language (e.g. only pages in German), target audience (laypeople vs. physician, apps, websites from Austria only (local information)).
- **Find information: look at previous presentations/work in the area, store relevant material in compendium.** The RP then types in “Advances in LADA diabetes treatment” and uses the filter to obtain journals and reviews only (e.g. PubMed and Cochrane). The RP switches the with news filter and image filter on and off in order to help him obtain material for the presentation. The RP filters the information based on the type of source (e.g. only books to get some ideas of definitions and introductory text on the topic, only primary resources to obtain access to papers on the topic) content/format to obtain slide presentations on the topic to see other work done in the area. Again, the RP is able to rank based on how recent and popular information is. Khresmoi suggests a video presentation (information resource: e.g. society of physicians) about the advances in diagnosis and treatment of LADA diabetes. The RP clicks on the video and uses relevant information for his presentation. The RP uses a compendium/cloud to store all information found on the topic and found useful for the presentation.
- **Suggested relevant links:** Khresmoi suggests relevant/related topics and websites which users with similar queries have looked at. The RP sees suggested relevant topics “Guidelines to diagnosis of LADA diabetes” and clicks on the link to find out more on current guidelines on how LADA diabetes is diagnosed (secondary source). The RP uses this for the introduction of his/her presentation and again stores it in the self-made compendia.

³³ Example: <http://www.linguee.de>

D8.2 Use case definition including concrete data requirements

Successful scenario: The RP is able to find definitional information on the topic (e.g. source: the Wikipedia, general websites), some papers containing news in diabetes treatment (source: PubMed, Cochrane etc.) which is recent, rated of good quality, from a trustworthy (academic) publisher and containing information about Sitagliptin, a new medication which has been shown to be effective to reduce the HbA1C levels of diabetic patients. He is able to find (if available) previous slide presentations on the topic (e.g. possible source: society of physicians). The RP stores the interesting data in the compendium. Khresmoi tracks the search history in case the RP wants to go back to a link.

Alternative scenarios

Login Scenario: Khresmoi recognizes that the physician is a research physician (automatic settings for RPs are activated)

Automatic settings for research physician are:

Automatic ranking (can be manually adjusted by user):

Type of information: Primary sources with some element of definitional secondary resources, videos, presentations, applications, images, etc.

Language: The physician is interested in the world-wide access (restriction to local information is not necessary for primary literature)

Tools:

- User can rate pages based on the quality and view/contribute to medical social forums.
- Integration of patient data within the search process as a diagnostic aid for complex cases.
- Self-made compendium/cloud (similar to Dropbox³⁴).
- Possibility to search for slide presentations

Examples of useful Data sources for this use case³⁵:

Primary sources and scientific secondary sources

- PubMed³⁶
- Medscape³⁷
- Cochrane³⁸

Past presentations in the field³⁹

³⁴ <http://www.dropbox.com>

³⁵ For an extended list of resources please see Annex 5

³⁶ <http://www.ncbi.nlm.nih.gov/pubmed?term=lada%20diabetes>

³⁷ <http://search.medscape.com/news-search?newSearchHeader=1&queryText=lada>

³⁸ <http://www.cochrane.org/search/site/lada>

³⁹ www.pitt.edu/~super4/36011-37001/36831.ppt

3.3.5 Use case 5: Hospital clinicians (HCs): To confirm the diagnosis proposed by a radiologist

After suffering symptoms such as shortness of breath, a patient is referred to a radiologist by a GP. The patient visits the radiologist at the hospital, who does a computer tomography (CT) scan of his/her lungs and refers the patient to a pulmonologist for further diagnosis and treatment. With the CT scan and the suspected diagnosis of UIP/IPF, the patient visits the pulmonologist. The pulmonologist looks at the CT scan which looks unusual for UIP and uses Khresmoi to verify the diagnosis made by the radiologist and to establish whether a biopsy for further confirmation is necessary. He/she uses Khresmoi to find typical cases of CT scans of UIP and NSIP (differential diagnosis) and to decide which type of treatment is best for this patient.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: In the office (hospital) while the patient is not present, accessing the Internet from a tablet computer.

Time available: 10 minutes (immediate need) while the patient is present.

Intention: To find images of typical UIP and NSIP (case reports) to gain confidence in the diagnosis, to find information on what's the best treatment for this patient.

Language: The HC has good skills in medical English, he/she prefers to access and type medical queries in German. He/she can understand simple medical English but finds automatic translation tools and automatic completion of queries helpful.

Search process:

The search bar: Khresmoi identifies the location of the HC as Vienna and therefore the language setting is German. HC types into the search bar "Usual interstitial pneumonia NSIP" Khresmoi offers automatic completion of query and can understand the abbreviation used to identify the search as "usual interstitial pneumonia non-specific interstitial pneumonia".

Search output:

- **Number of links:** Khresmoi delivers a search output containing a maximum of **20 links per page**
- **What is shown in the link? Automatic categorization/link description:** The link descriptions contain a) type of content, author, b) publisher, c) target audience, d) quality accreditation marks (if available), e) physician quality rating, f) publisher, g) time of the last update, h) indication whether restricted accessibility or free content, i) type of media (e.g. video vs. text).
- **Manual categorisation/Filter:** Khresmoi offers the possibility to filter the data by the type of source (primary vs. secondary source vs. forums vs. news vs. books), type of content (e.g. video vs. image vs. text vs. audio vs. apps vs. slide presentation), language (e.g. only pages in German), target audience (laypeople vs. physician, apps websites from Austria only (local information)). The HC activates the Image filter and target audience to only obtain images of UIP for physicians.
- **Advanced search options/Ranking:** Khresmoi offers the possibility of an advanced search where the user can choose to rank the information based on date range (i.e. time of

D8.2 Use case definition including concrete data requirements

publication/last update), physician quality and relevance ratings (i.e. how many search terms match the query), level of accessibility (freely accessible websites to be displayed first). The HC ranks the information using date range and physician quality ratings to find the most recent, trustworthy images on the topic.

- **Suggested relevant topics:** Khresmoi suggests a link to a recent Cochrane study (the link is clearly marked as restricted access) about the treatment of UIP. The HC is interested, clicks on the link and is directed to the article (required to log in for free access as a hospital member) and obtains a quick overview on a new medication for the treatment of this condition.
- **Tools:** Khresmoi offers the following tools which are relevant to HCs: Automatic translation, Dictionary (including explanation of abbreviations). The HC uses an automatic translation of results to obtain the translation of texts/words which are too complex to understand and/or the medical dictionary for terms/abbreviations that are unfamiliar.

Successful scenario: The HC is able to find the requested image, can confirm the suspicions of the radiologist (UIP) together with the clinical information he receives from the patient.

Alternative Scenarios

Unsuccessful Scenario: The HC does not find the information. He changes his search terms to “Usual interstitial pneumonia NSIP”

Login scenario: Khresmoi recognizes that the physician is a hospital clinician (automatic settings for hospital clinicians are activated)

- Manual categorisations/ranking are automatically set in the login scenario.
- User is able to rate content in his field of specialization.
- User is able to access medical forums anonymously.

Examples of useful data sources for this use case⁴⁰:

Scientific secondary resources

- Cochrane⁴¹
- Universimed⁴²
- UptoDate⁴³
- Best practice⁴⁴

Image database:

- Talisman Database⁴⁵

⁴⁰ For an extended list of resources please see Annex 5

⁴¹ <http://www.cochrane.org/>

⁴² <http://pulmologie.universimed.com/>

⁴³ <http://www.uptodate.com/index>

⁴⁴ <http://bestpractice.bmj.com/best-practice/welcome.html>

3.4 Categorization requirements for physician use cases

In order to guide annotation at both document and paragraph level, the following categories are suggested:

3.4.1 Document level annotation

Audience:

- Laypeople content (content targeted at patients rather than physicians)
- Physician content

Origin:

- Local content (i.e. country specific)
- International content

Level of Accessibility:

- Free content
- Payment required
- Subscription required
- No subscription required

Type of content:

- Primary source (e.g. first author paper)
- Secondary source (e.g. second author articles, magazines, reviews)
- Tertiary source (e.g. books, the Wikipedia)
- Forums
- Medical news
- Blogs
- Online CME courses
- Mobile device applications

Author/Institution:

- Medical portals
- Medical societies
- University websites
- Hospital websites

Format of content:

⁴⁵ http://www.sim.hcuge.ch/medgift/01_Talisman_EN.htm

D8.2 Use case definition including concrete data requirements

- Video
- Image
- Text
- Audio
- Presentations

Topic:

- Anaesthesiology and Critical Care Medicine
- Dermatology
- Emergency Medicine
- General Practice
- Child- and Adolescent Medicine
- Internal Medicine
- Pulmonary Disease
- Neurology/Psychiatry
- Obstetrics and Gynaecology
- Ophthalmology
- Orthopaedics and Orthopaedic Surgery
- Otolaryngology (Nose, Ear, Throat)
- Physical Medicine and Rehabilitation
- Surgery
- Urology
- Orthodontist and Dental Medicine
- Radiology

3.4.2 Paragraph level annotation

Disease and diagnosis:

- Definition
- Abbreviations
General descriptive information
- Causes, risk factors
- Symptoms
- Diagnosis
- Guidelines

D8.2 Use case definition including concrete data requirements

Treatment:

- Treatment-specific information
- Associated conditions
- Self-help groups for patients

Drugs:

- General drug information (the chemical constitution of the drug, when to use, how to use, definition)
- Drug interactions
- Drug contraindications
- Drug side effects
- New medication/medication in development
- Drug prices/local availability

Medical updating:

- Clinical trial information
- Medical education
- Research and scientific publications (document level for physicians?)

Health care:

- Directory of hospitals
- Directory of physicians
- Health insurances
- Social security
- Job/career opportunities

3.5 Preliminary analysis of a PubMed query log

PubMed is a very popular web search interface for biomedical literature and is made freely available on the web by the United States National Library of Medicine. In addition to articles from basic research, it also contains large amounts of medical literature. Studies conducted by the Khresmoi project and other research groups show that it is considered an important resource for medical practitioners, primarily by specialists and research physicians pursuing in-depth research for answering certain questions arising in their daily routine (see Khresmoi Deliverable D8.1.2).

A query log capturing queries submitted to PubMed during a 24-hour period is available. A basic analysis of this log file is also available in the scientific literature (7). The analysis of this query log for the purposes of Khresmoi is hampered by the fact that not only medical practitioners with contact to patients are using PubMed, but also researchers who do not use the search engine for clinical decision making. Therefore, we conducted a preliminary analysis of this log file to determine the percentage of

D8.2 Use case definition including concrete data requirements

queries in PubMed that are of possible relevance in clinical settings, and how these queries are structured.

A sample of 500 queries from the log file were selected and annotated as either ‘Khresmoi relevant’ or ‘Not Khresmoi relevant’. ‘Khresmoi relevant’ queries were those where one could imagine a scenario in which the query is used to answer an immediate or non-immediate medical question, focused on a specific person (e.g. a patient) or a specific population of people. Non-relevant queries were those that were done in the context of scientific research or veterinary medicine. The annotation was done by Matthias Samwald (TUV), who has domain knowledge in biomedical research and clinical practice.

Out of 500 selected PubMed queries, 103 (i.e. roughly 20%) were annotated as ‘Khresmoi relevant’.

Some examples of queries that were annotated as relevant: “salt restriction in hypertension”, “timecourse inflammation”, “interactions methamphetamine and MDMA”, “patient teaching for crohns disease”, “tobacco cessation”, “toradol and postpartum haemorrhage”, “intravenous injection”, “weeds and ingestions”, “alexithymia”, “nutrition disability”, “inflammation and cancer”, “triangular alopecia”, “blood glucose testing AND finger care”, “fractional anisotropy”, “fiber loss”, “prognosis first attack ulcerative colitis”.

These can be contrasted by queries that were annotated as not relevant: “The Arabidopsis CDPK-SnRK Superfamily of”, “apoptosis”, “breast cancer p53”, “Eugenol GC/MS”, “morris water maze”, “DCoH”.

This was a very basic approach for getting an overview of the queries in PubMed and their possible relevance for Khresmoi, and it needs to be decided if more in-depth analysis is warranted (e.g., with a larger group of annotators to cover more queries, more detailed annotation and to measure inter-annotator agreement). The current annotations can be downloaded from http://wiki.khresmoi.eu/index.php5/File:500_Pubmed_queries_annotated_with_khresmoi_relevance_-_Samwald.xls.

4 The general public use cases

4.1 The path towards use case categorization

To create user profiles, survey results were analysed with respect to stratification of respondents into 6 age/gender groups:

- Males, below 40 y.o.
- Females, below 40 y.o.
- Males, from 40 to 60 y.o.
- Females, from 40 to 60 y.o.
- Males, above 60 y.o.
- Females, above 60 y.o.

Some differences have been identified for the interface requirements, for example:

- Women of all ages give more preference to forums and blogs, radio and TV than men do.
- The youngest group would rather search in online patient groups.

D8.2 Use case definition including concrete data requirements

- The oldest group would appreciate risk factors assessment tools comparing with younger groups of users.
- The importance of a doctor consultation is growing with the age, hence in the interface the possibility to share the results with a physician should be highlighted for those aged above 60.
- On the contrary, the youngest users would rather verify the information on the website they trust than ask a physician, hence the possibility to rate results and to make a list of favourite websites are especially important to them.

Detailed user profiles are available in the Annex 2.

Another possible way to classify the users of Khresmoi could be the one depending on health status:

- A generally healthy person, browsing information for self-education.
- A person just diagnosed with an acute or chronic condition, searching information about an acute condition.
- A person just diagnosed with a chronic condition, searching a wide variety of information about a chronic condition from a definition of disease and treatment to a social worker assistance.
- A person, being a chronic patient for a certain period of time, “expert patient”.

Khresmoi can classify each user and his/her query based on this classification. For example, registered users may specify this information in the registration form. If a patient has had type 2 diabetes for a year, the results will be displayed targeting this group of users. In case the same patient is looking for another health topic, this search will be classified as the first or second category.

4.2 Interviews with representatives of the general public

Several semi-structured interviews with members of the general public were conducted in order to enhance the user profiles (for the questions, see Annex 3). The basis for conducting the interviews were the survey results presented in D8.1.1. However, some questions were difficult to address within this survey. For example, one finding of the survey results was that the users are interested in the automatic translation. However, we could not understand the specific features translation tools should propose, neither how users would like to use it. Another reason to conduct the interviews was a possibility to contrast the survey results where users reported self-perceived online health information search with the actual one. The interviews aimed to identify search processes about health issues which affected participants on a personal level. Interviews were conducted either in a face-to-face environment allowing direct observation of actual search behaviour or via Skype. One of the biggest difficulties was to find participants who would be willing to be interviewed. Several patients' organizations were approached, but direct contact with the patients (members of these patients' organizations) was rejected due to ethical considerations. Consequently, few participants were recruited at random.

Our initial hypothesis, based on a detailed analysis of the survey results, was that the search behaviour of internet users differs with respect to gender and age. As a part of the recruitment process, we therefore tried to represent both gender, all age groups as well as different knowledge levels of English. Overall, 6 interviews were conducted:

- a woman below 40 doing exploratory health search about a drug with high level of English

D8.2 Use case definition including concrete data requirements

- a woman below 40 searching for acute disease information, native English speaker
- a man below 40 doing exploratory search about a disease with high level of English
- a woman from 40 to 60 who is a professional searcher of online health information for patients with average level of English
- a woman from 40 to 60 with long-term chronic disease, “expert patient” with high level of English
- a man over 60 with long-term chronic disease, “expert patient” with low level of English

All interviewees have at least a higher education, none of them has a professional medical background.

A sample of questions for semi-structured interviews is available in Appendix 9.3.

Main findings:

- 1) All participants of this study went to national versions of Google Search.
- 2) They usually typed a name of disease into the search bar, preferably in their mother tongue, except for the expert patient with fluent English.
- 3) Most of the participants checked only the first page of results (often the top sponsored results were not identified as such by the participants of this study), some clicked on the first link (often including the sponsored ones), others scrolled down and clicked on up to 5 results. To decide which result “to click on” the participants checked the URL. Two of the participants reported a distinct preference to directly visit the Wikipedia.
- 4) The information about the disease or drug should be structured in a classical manner, the educated users are used to perceive medical information in this way as in a (medical) encyclopaedia. Illustrations and videos to explain the procedures were appreciated.
- 5) To evaluate information found on the Internet, different sources were compared. However, none of the participants checked the source, author, publisher, date etc. On the other hand, the Wikipedia was appreciated for its clear indication of original sources.
- 6) The interviewees reported that the time they spent for a search varied from 5 minutes to 1.5 hours.
- 7) In cases where the answer was not found, the two common strategies used were to either reformulate a query or to ask someone (a physician or a trusted friend with medical knowledge).
- 8) Participants with low and average knowledge of English in a familiar health topic preferred to obtain search results in their mother tongue. However, when information was not found, automatic translation of both the query and the results showed itself as an attractive possibility. Participants with average knowledge of English reported a preference of being able to view both the original and translated versions side by side to be able to read the original document when translation is not clear/precise. A medical thesaurus with explanation of medical terms in mother tongue was demanded by all users.
- 9) Being in a different country, in case of an emergency situation, participants would rather seek “human” professional help than use a search engine.

To conclude, based on the interviews, users searching for online health content fall under one of the following three categories:

D8.2 Use case definition including concrete data requirements

1) A general interest search characterized by browsing information as a result of news coverage on a certain medical topic, awareness campaigns, or when someone close has got a certain disease. Definitions and basic, general information on the disease are usually sufficient.

2) An active search around the time of a diagnosis of an acute or chronic condition (may be a bit before when there is already a suspicion and usually after the diagnosis). In contrast to the “general interest search”, users pursuing an active search require more detailed, unambiguous information from trusted sources of a wider spectrum, seek social support (patient organizations, other patients experience). The users are under stress.

3) An advanced search of very particular information in a long-term management of a (chronic) disease performed by so-called “expert patients”. This search is characterized by retrieval of very detailed information that is not widely available nor easily accessible.

First two categories would prefer to have a content in their mother tongue, often definitions and basic “gold standard” information detailed to different extent are sufficient to answer their questions.

Expert patients carrying out more complicated searches with good English skills were more likely to be familiar with medical terms in English typical for their condition and can therefore search and perceive the results in English. In case of average knowledge of English, a side-by-side comparison of the original and translated versions is suggested.

In all cases, medical thesaurus as an interpretation tool to understand the meaning of medical terms in a plain language is demanded. For example, a Spanish-native participant mentioned using RAE which is a common, not health/medical specific dictionary of Spanish/Latin-American Spanish.

4.3 General use tools

4.3.1 Initial choice of language/country

The user for the first time accesses Khresmoi. He/she types a query in a search bar. Between the bar and “search” button he/she can change a country he/she is coming from (by default the country is suggested based on the IP address) and the language he/she is the most comfortable with (by default the official language of the country determined by the IP address).

4.3.2 Spelling correction and query completion

A female user wants to type “diabetes during pregnancy diet”. The system corrects spelling (diabetes → diabetes) and suggests several options:

- Diabetes mellitus
- Diabetes mellitus type 1
- Diabetes mellitus type 2
- Diabetes symptoms
- Diabetes treatment (etc)

She selects “diabetes mellitus” and keeps writing “dur” when the system suggests “during pregnancy”. She selects this query and adds “risks”.

4.3.3 “Conversion” into medical terms where applicable

Once the query is typed, a “conversion” into medical terms is proposed, i.e.

"Do you know that the medical term "diabetes during pregnancy" is "gestational diabetes"? See more results on "gestational diabetes".

4.3.4 Definitions of medical terms

When Khresmoi recognizes a medical term in a query typed by a user, it provides brief definitions of this term. In case the query contains more than one medical term, all definitions are provided.

4.3.5 Preview of results

When the search results appear on the page, the user can preview them before accessing by hovering over the links. In this case, the whole paragraph is highlighted and when the user clicks on the link, the paragraph on the page is also highlighted.

4.3.6 Consult search history

On the right side menu, there is a box where the last 5 queries are displayed. User can check all of them by clicking on “See more” when a drop-down list appears containing all queries. By clicking on any of them a query is copied to the search box so the user can modify the query and a new search is executed. In this case, there are new results (top-rated) comparing with the previous search, and they are highlighted. Another possibility is to have a side-by-side comparison of the previous search and the current one.

4.3.7 Medical thesaurus

A medical thesaurus in various languages enables to understand medical terms unfamiliar to a user in a layman language. Terms, for which the interpretation is provided, are highlighted within each link description. When scrolling over, a pop-up window with definition of the word is provided. Ideally, when previewing or accessing the results, this tool is still available.

http://depts.washington.edu/respcare/public/info/Plain_Language_Thesaurus_for_Health_Communications.pdf

http://diabetes.niddk.nih.gov/spanish/pubs/dictionary/Diabetes_Dictionary.pdf

<http://diabetes.niddk.nih.gov/dm/pubs/dictionary/>

4.3.8 Risk factor assessment tools

These kinds of tools are based on questionnaires to fill. They could be extracted from different websites and collected to a repository in the tools section of the interface.

Some examples:

<http://www.yourdiseaserisk.wustl.edu/english/> (US, English)

D8.2 Use case definition including concrete data requirements

http://www.heart.org/HEARTORG/Conditions/HeartAttack/HeartAttackToolsResources/Heart-Attack-Risk-Assessment_UCM_303944_Article.jsp (US, English)

http://www.nhlbi.nih.gov/health/public/heart/obesity/lose_wt/risk.htm (US, English)

<http://www.liguepulmonaire.ch/nc/fr/maladies/syndrome-dapnees-du-sommeil/test-de-risque-dapnees-du-sommeil.html> (Switzerland, French/Italian/German)

The repository of risk assessment tools should explicitly explain that these tools were automatically selected from trustworthy sources, however, they are served to assist and to inform users, but in no case to replace the doctor-patient relationship.

4.3.9 Body 3D visualization

This tool helps users to visualize and understand human body concepts. This tool is highlighted when a query contains names of human body parts, for example, liver or kidneys. In another case, a user can access this service by selecting it from the list of all tools offered by Khresmoi. When accessing an image of a body part, the possible health conditions for this body part are listed aside.

4.3.10 Filtering by images, audio and video files

Users have the possibility to filter the search output based on the file format (image, audio, video).

4.3.11 Filtering by relevancy or recency

By default, all results are filtered by relevancy. However, it is possible to manually filter them by recency.

4.3.12 Recommended by Khresmoi search results

Similar to Google Ads appearing on the top of search results, Khresmoi can highlight and recommend a certain number of governmental, hospital, or university websites.

4.4 Translation and multilingual tools

Translation and multilingual tools depend on the mother tongue of the user and his/her proficiency in English as the majority of the sources are available in English. Based on the interviews, we can categorize future users of Khresmoi as:

1) native English speaker

→ query typed in English

→ no translation needed unless a person is staying in a non-English-speaking country and looking for local information, i.e. Spanish/French/German/Czech to English is required

→ thesaurus for medical terms in English

D8.2 Use case definition including concrete data requirements

2) good to excellent knowledge of English in a researched health topic (example of user – “expert patient”)

- query typed in English
- no full text translation needed, English texts are mostly easy to understand
- thesaurus for medical terms in the user’s mother tongue or in English

3) average knowledge of English in a researched health topic

- query typed in the user’s mother tongue
- preference is given to content in the mother tongue;
- in case information in the mother tongue is not found, automatic translation is needed with access to both original and translated versions;
- thesaurus is provided in a mother tongue for the translated results and in English for the original text

4) no or very basic knowledge of English in a researched health topic

- query typed in a mother tongue
- preference is given to a content in the mother tongue
- in case there is no information found, a full automatic translation to the mother tongue is needed
- thesaurus is provided in the mother tongue

Based on this, a medical thesaurus in all languages included in Khresmoi appears to be the core element for interpretation of the information.

4.5 Registration

Khresmoi users have a possibility to register.

Registration process:

First time a user access Khresmoi, a pop-up window should appear asking to login or register. In case the user does not want to register, there should be two boxes to tick:

- 1) remind me later
- 2) do not ask me again

Then the user starts searching.

In case the user wants to register he/she clicks on ”Register now”, then he/she is redirected to a registration page.

To log in the user goes to the upper right corner of the page and introduce his/her login and password in the relevant fields.

Every registered user has his/her profile page where he/she can modify the data and access the repository.

Registration form, apart from basic elements (name, email address, password, etc.), should include:

D8.2 Use case definition including concrete data requirements

1. The boxes to enter the gender (male/female) and age (below 40, from 40 to 60 and above 60). Based on the gender, the search results are customized. Based on the age and gender, the interface is customized.
2. Possibility to share the results with a physician/friend or peer: an optional part of registration form has an icon with “Physician” image and “Friend” image. Under both images there are fields to enter multiple email addresses. When a user is looking for information, near each search result a tool tip appears proposing to automatically share this search results with his/her physician or friend. By clicking this tool tip, a pop-up window appears in a form of email which contains email address(es) of recipient(s), topic of the email (a query of the user) and text by default in the body of the message (Hello, I would like to share with you the information I found in Khresmoi: <http://...> Regards, X). The user can modify this email or directly click on ”Send” if no modifications are required. After sending the email, a pop-up confirmation window appears. The recipient of the mail may then directly answer the user to his/her personal mailbox (stated during the registration process).
3. A possibility to save search results into a repository for all queries, audio and video files. When the search results appear, next to each of them there is a tool tip by clicking on which this result can be saved in the repository. The user should select one of two categories when he/she is saving the result: “Favourite” or “To check later”. When a user accesses his/her profile he/she sees the links for the search results and files saved in the repository into both categories. The URLs selected as “Favourite” form a list of favourite sites, so when doing other searches, the websites which are listed as favourites are highlighted/prioritized.

The benefits for the registered users are:

1. To have the content and interface adapted according to gender and age group.
2. Possibility to share the results with a physician/friend.
3. Receive updates by email/RSS.
4. Possibility to have a repository for all queries, audio and video files in one place.
5. Possibility to rate the search results, save a list of favourite websites (top-rated) which are highlighted when appearing in search results.
6. Saved results of assessment tests (for example BMI, or blood sugar level) with the possibility to see a graph showing the changes of the results over time.

It should be explicitly explained which data is stored as well as where and who has access to this data for both registered and non-registered users in accordance with the EU legislation.

4.6 Categorization requirements

Once a query is typed, it is automatically categorized into certain categories on two levels: documents and paragraphs. The results are categorized into the same two categories.

4.6.1 Document level annotation

The document level annotation provides a structured overview of possible health documents which are of interest to the general public. Overall, it represents 42 health-related classes of information.

1. Alcohol
2. At home
3. At work

D8.2 Use case definition including concrete data requirements

4. Babies and children
5. Biological risks
6. Bio terrorism
7. Cancer
8. Cardiovascular diseases
9. Carers
10. Chemical risks
11. Consumer safety
12. Drugs
13. E-health
14. Elderly
15. Environmental health
16. Food safety
17. HIV/AIDS
18. Influenza
19. Insurance
20. Long term care
21. Medicines and treatment
22. Men
23. Mental health
24. Nutrition
25. Other infectious diseases
26. Other non-communicable diseases
27. Patient safety
28. People with disabilities
29. Physical risks
30. Prevention and promotion
31. Rare diseases
32. Research
33. Road safety
34. Sex
35. Social environment
36. Sports and Leisure
37. Statistics
38. Tobacco
39. Travel
40. Vaccination
41. Women
42. Young people

Examples to illustrate some of these classes:

- Alcohol: effects of alcohol on health, combination of alcohol with certain drugs etc.
- Drugs: search of information about a drug, comparison of drugs, where to buy certain drugs (for example, Suboxone which can be obtained via certified providers and pharmacies).
- Environmental health: search for asbestos-related information and mesothelioma⁴⁶, the possible risks, prevention, health effects etc.

⁴⁶ Mesothelioma is a form of malignant cancer often caused by exposure to asbestos.

D8.2 Use case definition including concrete data requirements

- Influenza: search of information regarding periodical outbreaks of influenza, related information about vaccinations.
- Mental health: searches may refer to depression and its treatment, suicide prevention etc.
- Road safety: search for information about trauma caused by road accidents, about local facilities in case of road accidents (emergency phone numbers, closest hospitals etc.).

4.6.2 Paragraph level annotation

In a paragraph level annotation, we simply focus and expand the most "important" (for document level) classes, such as drugs, diseases etc. The classification for paragraph level annotation is based on the extensive and continuous evaluation overview of health websites for the purposes of the HONcode Certification by the Health On the Net Foundation. Some subcategories can overlap, thus it is important to initially use categories, not subcategories. Present classification is not exhaustive and covers only the needs of the general public.

Disease

- Definition
- Reasons, risk factors
- Mechanism of disease development
- Symptoms
- Diagnosis
- Treatment
- Prognosis
- Associated conditions
- Patient support
- Day-to-day life
- Clinical trials
- Research and scientific publications

Drugs and medicinal substances

- Definition
- When to use?
- How to use?
- Side effects
- Precautions
- Interactions
- Overdose
- Clinical trials

D8.2 Use case definition including concrete data requirements

- Research and scientific publications

Healthcare

- Directory of hospitals (healthcare organizations)
- Directory of physicians
- Health insurance
- Social security

Test and procedures

- Blood tests
- Urine tests
- Biopsies
- Radiology
- Physiotherapy
- Surgical operations

First aid

- Animal bites
- Burns
- Bleeding
- Bruise
- Food poisoning
- Allergy

Healthy lifestyle

- Balanced diet
- Diet for specific health conditions
- Exercise and weight loss
- Quitting harmful habits
- Environmental health

4.7 Preliminary HON query log analysis

A specific open-source software was developed to perform HON query log analysis (detailed in Annex 4). By now information about queries themselves and sessions was analysed. We were able to extract information about terms and number of queries made by users. For example, it was found that most queries comprised a single word, followed by two word queries, and least frequently three word queries. If we compare this with the survey results (D8.1.1), we can see that 60% of respondents reported typing 2-3 words, and only 3% reported typing a single word. The most common terms

D8.2 Use case definition including concrete data requirements

related to environmental fluctuations, as the top single and cooccurring terms are “sleep”, “apnea” and “symptoms”. These searches were mostly coming from China. We made a search, and it appeared that in the end of November 2011 in China there was a car accident caused by a driver who had fallen asleep while driving⁴⁷. We assume it had a wide coverage on news that caused an increased interest in this topic.

There is no guarantee that only the general public has been using the search engines from which the query log was collected. At the current stage of the project, it is not yet possible to extract and analyse the information regarding which results users have clicked. However, work is being continued to retrieve further information. Since most queries contain no more than one or two terms, it is not possible to identify the relationships between the terms. Most of the inquiries appear to be in a definitional level. Detailed report on the software development and preliminary logs analysis can be found in the Annex 4.

4.8 Scenarios

In the following scenarios, we intend to focus on very specific, yet typical situations related to diabetes and describe different types of health searches done by members of the general public.

4.8.1 First scenario: To find the information about a diet during gestational diabetes

Alice is a pregnant woman (26 weeks along) of 30 years old, who was diagnosed with gestational diabetes during a routine visit to her physician. She was prescribed a diet and regular blood sugar tests to control the level of glucose in her blood. Once back at home, Alice still has various questions and she particularly wants to get more information about how to follow the right diet to avoid taking medications or insulin.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: At home, accessing the Internet from her PC/laptop.

Time available: Until all information is found and understood.

Expectations:

- Content: diabetic diet in text, video format, food calculators
- Technical: Information should be trustworthy and relevant, targeted to the general public, females.

Language: Easy to understand for people new to diabetes, English is the user’s mother tongue.

Willingness to pay: No.

Postcondition: The requested information is found and understood.

Steps of the search process:

1. Search bar:

⁴⁷ <http://news.cn.yahoo.com/ypen/20111201/733216.html>

D8.2 Use case definition including concrete data requirements

Alice goes to Khresmoi as suggested by her physician and wants to type “diabetes during pregnancy diet”. While typing, the system corrects spelling and suggest several options.

2. “Search” button:

Right below the search bar Khresmoi proposes the following:

“Do you know that 'diabetes during pregnancy' is called 'gestational diabetes'? Get more results on 'gestational diabetes'”

Below this phrase there is a definition of “gestational diabetes” in English.

3. Categorization of query:

As her query can be categorized into two categories: “disease” and “diet and healthy lifestyle”, these two sections of the left side menu are highlighted. In this way, if Alice wants to get more relevant information she can browse in the subcategories of “disease” and “diet and healthy lifestyle.”

4. Categorization of results:

Khresmoi delivers a search output containing a maximum of 10 links per page. The results are automatically ranked by relevancy. The links are categorized in several ways:

- By target audience: by default the results are displayed for the general public
- By language and country: English, UK

5. Results description

A result description contains a title of the page, the URL and an extract of the text with key words from the query. Each result also contains ratings of the source by the general public and physicians and a tip to evaluate this result.

Tools

- On the right side menu, Alice can read the latest news on this topic, access social media content (forums and blogs) and check interactive tools (food calculators).
- Medical thesaurus is available, so if while reading Alice does not understand a term, she can scroll over the underlined term and its definition in layman terms appears.

Conclusions:

Alice has found websites, which are relevant, rated of good quality by both categories of users, and contains information about the “gestational diabetes diet”. Khresmoi tracks the search history. Alice also rates the search result where she found the information, and her rating is summed up with other layman users. She has found information and has understood how to eat to keep blood sugar level under control.

Summary:

- Information requirements: diabetes diet
- Data sources requirements: general consumer health content; text, video, calculators
 - <http://www.diabetescare.net/>
 - <http://www.nlm.nih.gov/medlineplus/ency/article/007430.htm>
 - <http://www.everydayhealth.com/diabetes-awareness/gestational-diabetes-diet.aspx>

D8.2 Use case definition including concrete data requirements

- Interface requirements: highlighted patient groups, forums and blogs, share results with a physician
- Results representation requirements: up to 10 results
- Multilingual/translation requirements: NA

4.8.2 Second scenario: To find the information about a blood test

Jorge is a 45-year-old Spanish man, overweight. During his last visit to his physician, he was recommended to have a blood test. He can see in the prescription that the recommended test is A1C. He is stressed due to this test, wants to know why it needs to be done, and goes online to find this information.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: Jorge is in a waiting room and accesses the Internet from his mobile phone.

Time available: 5 to 10 minutes.

Expectations:

1. Content: description of the test, why it is taken and the interpretation of results, the link between the test and obesity
2. Technical: Information should be trustworthy and relevant, targeted to the general public.

Language: Easy to understand for people new to diabetes, Spanish is the user's mother tongue.

Willingness to pay: No.

Postcondition: The requested information is found and understood. Jorge has understood that the test is used to see the average level of glucose in the blood, and can be used to screen for diabetes. He is aware of which level of A1C is considered healthy and which level can signal the presence of diabetes.

Steps of the search process:

1. Search bar:

Jorge goes to Khresmoi because he has seen its name advertised in a waiting room and wants to type in «a1c obesidad⁴⁸». While typing, the system corrects spelling and suggests several options.

2. «Buscar⁴⁹» button:

Right below the search bar Khresmoi proposes definitions of «La prueba A1C» and «Obesidad» in Spanish.

3. Categorization of query

As his query can be categorized into two categories: «disease» and «tests and procedures», these two sections of left side menu are highlighted. In this way if Jorge wants to get more relevant information, he can browse in the subcategories of «disease» and «tests and procedures».

4. Categorization of results

⁴⁸ «a1c obesity» in Spanish

⁴⁹ «Search» in Spanish

D8.2 Use case definition including concrete data requirements

Khresmoi delivers a search output containing a maximum of 10 links per page. Results are automatically ranked by their relevancy. The links are categorized in several ways:

- By target audience: by default the results are displayed for the general public.
- By language and country: Spanish, websites from Spain.

5. Results description

A result description contains the title of the page, the URL and an extract of the text with key words of the query. Each result also contains the ratings of the source by the general public and physicians and a tip to evaluate this result.

Jorge previews/accesses several results, and if necessary, checks the second page.

Tools:

- On the right side menu he can check the interactive tools (A1C calculator).
- Medical thesaurus in Spanish is available, so if while reading he does not understand a term, he can scroll over the underlined term so that its definition in layman terms appears.

Conclusions:

Jorge has found websites which are relevant, rated of good quality by both categories of users and contains information about the blood test A1C, how it is taken, why and which levels can signal the presence of diabetes. Khresmoi tracks the search history. Jorge also rates the search result where he found the information, and his rating is summed up with other layman users. He has found the information and has understood that his physician has asked him to do this test because obesity is one of the risk factors of diabetes development, and the test is used to screen for diabetes.

Summary:

- Information requirements: A1C test, screening for diabetes, obesity as a risk factor of diabetes
- Data sources requirements: general consumer health content; text, video about how the test is taken.
 - <http://www.hormone.org/Spanish/upload/type-2-diabetes-and-A1C-spanish-042010.pdf>
 - http://ndep.nih.gov/media/KnowYourNumbers_Spanish.pdf
 - <http://www.estudiabetes.org/video/906848:Video:32824>
- Interface requirements: patient support groups are highlighted
- Results representation requirements: up to 10 results
- Multilingual/translation requirements: As this type of content is available in Spanish, there is no need for translation. In case of a different query for which the content in Spanish is not available, results are automatically translated from English to Spanish.

4.8.3 Third scenario: To find information about cinnamon in treating type 2 diabetes

Julie is 50 years old and has just been diagnosed with type 2 diabetes. Her physician prescribes metformin, recommends her to control her blood sugar levels, to do lifestyle changes and suggests to consult a nutritionist. Julie is overwhelmed, she tries to cope with all this information and follow the recommendations. About a month later she meets some friends and tells them about her new health

D8.2 Use case definition including concrete data requirements

status. One of the friends mentions the studies to investigate whether cinnamon can be used to cure diabetes. Julie wants to find more information about it and considers starting to take cinnamon.

Login status: The user is not logged in.

Precondition: Access to the Internet.

Context: Julie is at home and accesses the Internet from her PC.

Time available: Up to 20 minutes.

Expectations:

1. Content: secondary sources summarizing the existing evidence of cinnamon use for diabetes, original articles of the studies.
2. Technical: Information should be trustworthy and relevant, targeted to the general public.

Language: Easy to understand for people new to diabetes, French is the user's mother tongue, however she also feels comfortable with English.

Willingness to pay: No.

Postcondition: The requested information is found and understood. Julie has understood that cinnamon intake may have effects on blood sugar level; however, there is no sufficient scientific evidence to prove the effects. She does not add cinnamon to her diet.

Steps of the search process:

1. Search bar:

Julie goes to Khresmoi and wants to type "recherche cannelle diabète⁵⁰". While typing, the system corrects spelling and suggest several options.

2. "Recherche⁵¹" button:

Right below the search bar, Khresmoi proposes definitions of "diabète" and "cannelle" in French.

3. Categorization of query

As her query can be categorized into two categories: "disease" and "drugs and medical substances", these two sections of left side menu are highlighted. In this way if Julie wants to get more relevant information, she can browse in the subcategories of "disease" and "drugs and medical substances".

4. Categorization of results

Khresmoi delivers a search output containing a maximum of 10 links per page. The results are automatically ranked by relevancy. The links are categorized in several ways:

- By target audience: by default the results are displayed for the general public
- By language and country: initially French, websites from France; further she accesses original articles in English and reads them in English but needs a medical thesaurus to interpret unfamiliar terms.

5. Results description

⁵⁰ «research cinnamon diabetes » in French

⁵¹ «Search» in French

D8.2 Use case definition including concrete data requirements

A result description contains the title of the page, the URL and an extract of the text with key words of the query. Each result also contains ratings of the source by the general public and physicians and a tip to evaluate this result.

Tools:

- On the right side menu, she can check news about this topic.
- Both original and translated versions of the text are provided side by side.
- Medical thesaurus in both French and English are available, so if while reading she doesn't understand a term, she can scroll over the underlined term and its definition in layman terms appears.

Conclusions:

Julie has found websites which are relevant, rated of a good quality by both categories of users and containing information about the studies investigating the effects of cinnamon on diabetes. Khresmoi tracks the search history. Julie also rates the search result where she found the information, and her rating is summed up with other layman users. She adds some of the websites as favourites. She has found the general online health content in French explaining that so far there is no solid evidence that cinnamon intake improves blood glucose control. She has also accessed and read the original publication in English and with the help of a medical thesaurus, has understood the major findings. She decides not to take cinnamon. Next time she meets her friends Julie explains it to them.

Summary:

- Information requirements: information about studies of cinnamon for diabetes
- Data sources requirements: news and articles about studies, original scientific articles PubMed; text.
 - <http://www.lanutrition.fr/les-news/diabete-la-cannelle-ne-ferait-pas-de-miracles.html>
 - <http://www.ncbi.nlm.nih.gov/pubmed/22093965>
 - <http://www.ncbi.nlm.nih.gov/pubmed/22007625>
- Interface requirements: patient support groups are highlighted
- Results representation requirements: up to 10 results
- Multilingual/translation requirements: health news are available in French, no translation needed; original studies in English require medical thesaurus in both French and English.

4.8.4 Fourth scenario: Patient with complications of type 2 diabetes: to find information about the side effects of a drug

Andrej is a 63-year-old Czech man who has just moved to Geneva, Switzerland from the US and does not speak French. Recently he has started having blurred vision. He has had diabetes for 13 years and is aware that diabetes affects vision. He wants to find more information about this problem and an ophthalmologist to consult.

Login status: User is not logged in.

Precondition: Access to the Internet.

Context: Andrej is at home and accesses the Internet from his PC.

Time available: Up to 20 minutes.

D8.2 Use case definition including concrete data requirements

Expectations:

1. Content: description of the eye pathology typical for people with diabetes, information about ophthalmologists.
2. Technical: Information should be trustworthy and relevant, targeted to the general public.

Language: Andrej is an expert patient and fluent in English, Czech is the user's mother tongue.

Willingness to pay: No.

Postcondition: The requested information is found and understood. Andrej has learnt that most probably his condition is called diabetic retinopathy, but he needs to consult an ophthalmologist to be sure and to know how to manage it. He has also found the information about an ophthalmologist in his area.

Steps of the search process:

1. Search bar:

Andrej goes to Khresmoi and wants to type "diabetes eye problem". While typing, the system corrects spelling and suggest several options.

2. "Search" button:

Right below the search bar, Khresmoi proposes definitions of "diabetes" and "eye" in English.

3. Categorization of query

As his query can be categorized into one category: "disease", this section of left side menu is highlighted. In this way if Andrej wants to get more relevant information he can browse in the subcategories of "disease".

4. Categorization of results

Khresmoi delivers a search output containing a maximum of 10 links per page. Results are automatically ranked by relevancy. The links are categorized in several ways:

- By target audience: by default the results are displayed for the general public, however it is possible to change manually, so the results are targeted to physicians.
- By language and country: French from Switzerland, but Andrej does not speak French, hence he needs to find the information in English from Switzerland.

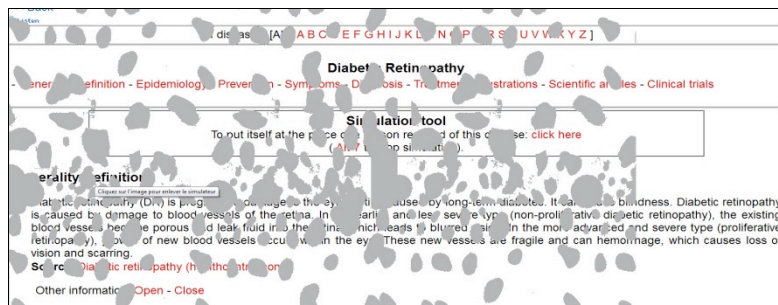
5. Results description

A result description contains the title of the page, the URL and an extract of the text with key words of the query. Each result also contains the ratings of the source by general public and physicians and a tip to evaluate this result.

Tools:

- Accessibility: as Andrej has vision problems, increasing letter size and changing the background colours facilitate his experience with Khresmoi
- Simulation tools: among other tools, simulation tools are proposed:

D8.2 Use case definition including concrete data requirements



http://www.provisu.ch/PROVISU/Structure/C11.768.257_en.html#Skip

- A medical thesaurus in Czech is available, by default it is provided in English, but Andrej can also consult the Czech one, so if while reading he does not understand a term, he can scroll over the underlined term and its definition in layman terms appears.

Conclusions:

Andrej has found websites which are relevant, rated of good quality by both categories of users and contains information about diabetic retinopathy. Khresmoi tracks the search history. Andrej also rates the search result where he found the information, and his rating is summed up with other layman users. He has found the general online health content in English explaining the reason, development and consequences of diabetic retinopathy. Andrej also finds an information about an ophthalmologists in Geneva. He calls for an appointment to discuss possible treatment.

Summary:

- Information requirements: information about diabetic retinopathy and ophthalmologists dealing with this condition
 - http://www.provisu.ch/PROVISU/Structure/C11.768.257_en.html
 - http://www.provisu.ch/cgi-prov/cache.pl?orig=http://www.docteur.ch/ophtalmologues/ophtalmologues_k_ge.html
- Data sources requirements: general health content, directory of ophthalmologists
- Interface requirements: simulation tools are highlighted
- Results representation requirements: up to 10 results
- Multilingual/translation requirements: information is available in English, no translation needed; some information may require a medical thesaurus in Czech.

4.8.5 “Rainy day” scenarios

1) No relevant result appear

- check relevant topics,
- reformulate query making more general,
- check the tutorial on successful strategy of online health information search,
- search for the information in English with the help if necessary of automatic translation of query and results

2) Too many results

- reformulate query making more specific

There should be a button in the end of search results “have not found the answer to your question? Do...” which can redirect to a tutorial on how to search health information on the Internet.

5 Conclusions

In this deliverable, two large scale surveys helped us to obtain insight into both the search behaviour and the information needs of physicians and the general public. An extensive analysis of the survey results, follow-up interviews and query analysis enabled identification of different user scenarios highlighting search specific needs and data requirements of distinct user groups. In the final section, the key messages for both categories of users are listed.

5.1 Key messages of physician use cases

The categorization of different user profiles and creation of corresponding possible scenarios attempted to provide a first understanding about different levels of search requirements amongst physicians. The expectation of a user towards the search engine is mainly driven by the amount of time available and the level of knowledge required to answer a query. The following conclusions forms the basis of proposed physician user requirements listed in Table 2.

Primary, secondary vs. CME search? There is a distinct difference between a primary and secondary source and a CME search. A search targeting primary sources usually requires both time and expertise to read and to evaluate such sources. In contrast, secondary resources enable a quick overview of applicable but often superficial information. CME resources enable physicians to pursue medical updating. Within the user profiles we found that both the level of specialization and the time restrictions determined the type of search a user did.

Local resources vs. the whole world? A clear distinction should be drawn in whether a physician wants to look for local information or “world-wide” information. Scientific medical updating usually targets recent world-wide information to get informed about the “state-of-the-art” in the specified field. On the other hand, typical point-of-care queries (e.g. what drug can I prescribe) require a connection to local resources (e.g. drug resources) in order to find out whether the information is applicable within the existing local medical system.

The role of multilinguality? While most physicians who participated in the survey (D8.1.2) reported adequate skills of medical English, there is a distinct preference to filter content in the user’s mother tongue. Especially time-constrained physicians requiring quick information at point-of-care look for information in their mother tongue, as it is quicker to understand. On the other hand, physicians seeking “specialised “ and scientific information may be more open to content in English. However, query translation to English is useful on all levels of search, as there is a clear difference between understanding a language and speaking (i.e. knowing the relevant words) a language.

The help to filter good quality information: A good link description, relevance and quality ratings. Several findings of the physician questionnaire described in deliverable 8.1.2 provide solid support for the notion that physicians require help in the quality evaluation of information obtained in the internet. Quality ranking was regarded as helpful, a good link description indicating the author, publisher, relevance of a source seen as essential. Furthermore, the possibility of the integration of

D8.2 Use case definition including concrete data requirements

patient data within the search process as a diagnostic aid for complex cases was viewed as an ideal solution amongst all groups of physicians. Furthermore, suggested relevant links and where other users have found the answer was another useful tool for the “help to search” amongst physicians. Thus, one upcoming challenge of Khresmoi may be to provide sufficient tools which help physicians to choose reliable, quality content whilst maintaining interface simplicity.

Towards a secured physician platform: A help of a “digital colleague”. It was found that physicians are overwhelmed by a search output in which physician content and lay people content is mixed. The possibility to filter out laypeople content or in the case of log-in access a secured physician platform is therefore essential. The finding of the questionnaire that most physicians seek a tool in which they can view quality rating of other physicians brings up the requirement of a “digital colleague”. It appears that physicians seek “social support” in their medical decision making and to help decide what content is trustworthy. Thus, the availability of a secured platform allowing communication amongst physicians, the possibility of rating and perceiving the quality of content and discussing patient cases make up an essential part of user requirements.

Promoting efficient medical updating: Availability of CME search filter, ranking by time of the last update, efficient content annotation. Another concept which may advance current systems is to provide a search engine promoting medical updating amongst physicians. Two functionalities may realise such a vision. First, the possibility to rank information by the time of the last update (most recent first) and the possibility of pursuing a local CME search is essential. At the moment there is lack of search engines extracting local CME content in an organised format. However, since all physicians must pursue medical updating efficient access to CME content is an essential requirement.

Mobile device friendly: In the physician questionnaire it was found that there is a growing trend towards accessing medical information via mobile devices. Thus, the possibility to filter apps and the adjustment of a format visible on mobile devices is essential.

5.2 Key messages of the general public use cases

In this document, four layman user profiles and their possible search scenarios were presented. Apart from it, some general-use tools, categorization requirements and registration process are described.

We have seen that there are several ways to categorize the layman users: based on their age, gender, health status, English language proficiency, and familiarity with a certain health topic. Ideally, all these factors should be taken into account while interacting with the users. Knowing their age and gender allows to customize the search results, as well as the interface. Knowing their personal health status allows to predict the above average level of health literacy in the topics directly related to personal health problems. On the other hand, we can expect a low level of eHealth literacy when the users are searching information about non-familiar conditions. Most of the users would prefer a local content in a mother tongue, especially those who are new to certain diseases and health topics. They will also benefit from an automatic translation of query and search results in case local information in their mother tongue is not available. “Expert patients” with good level of English may search and navigate in online health content dedicated to their disease originally written in English without automatic translation. On the other hand, they may require translation while searching for unfamiliar conditions. All users appreciate a medical thesaurus to explain the medical terms in a layman language. This tool should be provided in a mother tongue.

All the mentioned dimensions when combined provide an enormous range of combinations and variations, especially taking into account the variety of health topics. The biggest challenge is how to

D8.2 Use case definition including concrete data requirements

try to identify as much as possible of these dimensions automatically so the results are adapted to user knowledge and understanding of the particular health domain.

Another important issue to mention is the quality of information. Although in the survey the participants stated that the quality of the information is important to them, it appeared that none of the interviewed persons understand the dimensions of quality, as having the author, references, dates etc. From this, we can draw the conclusion that only trustworthy and curated content should be exposed to layman users when by default any search result user clicks is of a high quality.

Further work to enrich the understanding of users behaviour includes more detailed analysis of query log.

5.3 Summary of use case requirements

Based on the use cases described in the document, we have identified six types of searches: two performed by the general public and four by physicians:

- **Basic search**, characterized by definition and easily accessible health content search (as symptoms of a disease, drug description etc.) by members of the general public.
- **Advanced search**, characterized by more profound, detailed search by members of general public, typically by “expert patients”.
- **Wide search** is performed when a physician (typically a GP or physician in training) seeks out definitional, general or descriptive information from secondary resources.
- **Narrow search** is performed when a physician (typically a self-employed specialist or hospital clinician) seeks information in his/her field of specialisation, usually from specialised or society resources, to answer a usually concrete, complex question.
- **Research search** is performed when a physician (typically a research physician or hospital clinician) seeks information, usually from primary resources, describing the newest findings or “state of the art” on a topic.
- **CME search** is performed when a physician seeks local, online CME content to pursue required levels of medical updating.

Detailed explanation is provided in Table 1.

		General Public Tasks		Physician Tasks			
	<i>Users \ Search Tasks</i>	Basic Search	Advanced Search	Wide Search	Narrow Search	Research Search	CME Search
Public	Generally Healthy Person						
	Just Diagnosed Acute						
	Just Diagnosed						

D8.2 Use case definition including concrete data requirements

	Chronic						
	Chronic for longer period						
Physicians	Self-Employed GP						
	Self-Employed Specialist						
	Physician in Training						
	Research Physician						
	Hospital Clinician						

Table 1. Definition of types of searches by users and tasks

Further on, we identify specific requirements for each of these types of searches. All details are provided in the table 2.

<i>Functions \ Search Tasks</i>	General Public Tasks		Physician Tasks			
	Basic Search	Advanced Search	Wide Search	Narrow Search	Research Search	CME Search
Resources						
Primary resources		Potentially		Partly		
Secondary resources					For background	
Guidelines						
Certified websites						
Specialised websites (medical societies, hospitals)						
Drug information						
Further product information beyond drugs (glucometers, syringes, lenses)						
News						
Forums						
CME courses						
Wikipedia						
Local career websites						
Physicians directory						
Local healthcare information (patients/physician associations, health insurance, social security)						
Events (conferences)						
Classification/Filtering						

D8.2 Use case definition including concrete data requirements

(Default value shown)						
Primary/secondary resources	Don't show	Secondary	Secondary	Secondary	Primary	Secondary
Physician/general public resources	General Public	General Public	Physician	Physician	Physician	Physician
Level of detail (overview->very specific)	Overview	Specific	Overview	Specific	Specific	Overview
Type of resource (see list above)						
Type of content (diagnosis, treatment, ...) - also paragraph level - see list			Treatment mainly	Diagnosis mainly		
Country of origin of source	Local	Local	Local	Local	International	Local
Date Range				Recent	Recent	Recent
Trustability	Trustable	Trustable	Trustable	Trustable	Trustable	Trustable
Readability	Readable	Readable	Readable			Readable
Language	Mother tongue	Mother tongue	Mother tongue	Mother tongue	English	Mother tongue
Physician speciality				Own field	Own field	
Target general public group (gender, age, ...)						
Media type (video, audio, image, ppt, ...)						
Ranking by						
Relevance						
Language	Mother tongue	Mother tongue	Mother tongue	Mother tongue	English	Mother tongue
Location (independent of language)	Local	Local	Local	Local	International	Local
Physician social ratings						
General public social ratings						
Level of accessibility						
Date of publication						
Interface Layout						
Simple						
Comprehensive						
Link Description						
Type of Content						
Author						
Publisher						
Target Audience						
Quality accreditation marks						
Physician quality description						
General public social ratings						
Time of the last update						
Indication whether free access						
Type of media						
Page name						
URL						
Snippet						
Translation Tools						
Query translation (entered language -> English)						

D8.2 Use case definition including concrete data requirements

Document translation (English->chosen language)						
Translation and explanation of terms/abbrev (medical dictionary)						
Medical thesaurus (medical->layman translation)						
Abbreviations/Medical dictionary?						
Original/Translated Side-by-side						
Query Completion/Support						
Suggested query completion/expansion						
Integration of patient data						
Search history						
Spelling correction						
Suggestion of Scientific/Medical Synonym						
Term definition shown below search bar						
Query categorisation (content type)						
Tools						
Self-made compendium						
Image search						
List of favourites						
Share result with physician/peer						
Accessibility options						
Result preview						
3D body visualisation						
Collaborative/Social Aspects						
Physician quality rating	Sees	Sees	Mostly uses	Uses and rates	Uses and rates	Uses and rates
General public quality rating	Uses and rates	Uses and rates	Sees			
Forums						
Suggested Links/Topics						
Pages/topics related to the one viewed (drug contraindications, lifestyle changes)						
Recent publications						
Websites that users with similar queries have visited						
Educational/Career Websites						
Latest News						
Forums and Blogs						
Interactive Tools (Calculators)						
Patient Groups						
Simulation Tools						

Table 2 The summary table of requirements for all general public and physician use cases by task

6 Acknowledgements

We would like to thank all the physicians and representatives of the general public who agreed to be interviewed in order to enhance and validate the user profiles and user scenarios. We would like to thank also Pavel Pecina (CUNI) for reviewing of this document.

7 References

1. Diabetes. *WHO/Europe*. [On line] WHO. [Citation : 10 Jan 2012.]
<http://www.euro.who.int/en/what-we-do/health-topics/noncommunicable-diseases/diabetes>
2. *Global prevalence of diabetes: estimates for the year 2000 and projections for 2030*. **Wild S, Roglic G, Green A, Sicree R, King H**. 5, s.l. : Diabetes Care, 2004, Vol. 27.
3. **WHO IDF**. *Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia* : Geneva : WHO Document Production Services, 2006. ISBN 92 4 159493 4.
4. **A., Cockburn**. *Writing Effective Use Cases*. s.l. : Addison-Wesley Professional, 2000. ISBN-10: 0201702258.
5. *Latent Autoimmune Diabetes in Adults*. **M, LANDIN-OLSSON**. s.l. : Annals of the New York Academy of Sciences, 2006, Vol. 958.
6. *American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias*. **American Thoracic Society/European Respiratory Society**. 2, s.l. : Am J Respir Crit Care Med, 2002, Vol. 165. PMID:11790668.
7. *A Day in the Life of PubMed: Analysis of a Typical Day's Query Log*. **Herskovic JR, Tanaka LY, Hersh W, Bernstam EV**. 2, s.l. : J Am Med Inform Assoc., 2007, Vol. 14. PMCID: PMC2213463.

8 Annex

Annex 1. Physician user profiles (based on questionnaire results)

The following table is based on the physician survey results (D8.1.2). Overall trends (marked as Overall), Common trends (All groups) as well as additional “group specific” and significant (“More likely as”) trends search behaviour, need, and preferred tools are highlighted. If relevant % of total valid samples are indicated.

SEGP	SES	HP	RP	PIT	Young	Old
Information Need						
General information needs ?	Overall top Information needs are:					
	Drug information					
	Clinical trial information					
	Medical Education					
Overall	Treatment specific information					
	Disease descriptions					
All groups	All groups look for :					
	Drug information					
	Treatment specific information					

D8.2 Use case definition including concrete data requirements

Group specific	Disease descriptions	Clinical trial information Medical Education	Clinical trial information Medical Education Disease descriptions (Diagnostic aid Highest need for diagnostic aid !	Clinical trial information Medical Education Highest need for clinical trial information	Clinical trial info Med. Educat. Disease descr. Abbreviatio Diagnostic Aid Highest need for abbreviations Highest overall informatio need!	Drug information Clinical trial information Medical Education Disease descr.	Drug Information Clinical trial info 3. Medical Education
----------------	----------------------	---	--	--	--	---	---

	SEGP	SES	HP	RP	PIT	Young	Old
What drug info?	Overall Top 3 Drug Information Needs:						
Overall	1. Drug side effects 2. Drug Interactions 3. Drug dosage						
All groups	<u>All groups place an above average requirement on the following drug info needs:</u>						
	Side effects						
Group specific	Drug interactions Drug contraindications More likely to look for: Drug interactions and Drug prices	Drug interactions Drug contraindications	Drug interactions Drug dosage More likely to look for: New medication/ Medication in development Drug safety	Drug contraindications Drug interactions More likely to look for:, Drug description, Drug safety, Drug dosage, Drug prices	Drug dosage Drug contraindications More likely to look for: Drug description	Drug Dosage Drug contraindications	Drug interactions Drug contraindications More likely than younger to look for: New Medication in development +Drug contraindications

D8.2 Use case definition including concrete data requirements

on

What type of need?

Overall the most popular needs were:

immediate need and the educational need

Overall

Group specific	Immediate need (48,4 %)	Mainly immediate (41,3%) + educational (37%) need Highest non-immediate need (20,7%)	Immediate (42,3%) and Educational (41,4%)	Educational (46,7%) Highest level of educational need	Immediate need (61,3%) Highest level of immediate need	Mainly immediate (56,4%)	Mainly Educational (48,1%)
----------------	-------------------------	--	---	---	--	--------------------------	----------------------------

SEGP	SES	HP	RP	PIT	Young	Old
Resources						

What data sources ?

Overall Top resources are:

Overall

1) GSE

All groups

2) Med. Research Databases

3) Wikipedia

4) Society Websites

All groups use General search engines

D8.2 Use case definition including concrete data requirements

Group specific:	General health websites	Med. Res. DB.	Med. Res. DB	Med. Res. DB.	Med. Res. DB	Med. Res. DB	Med. Res. DB
	Wikipedia	Society Websites	Society Websites	Society Web. Hospital/University Websites	Wikipedia	Wikipedia	Society Websites
			Wikipedia	Wikipedia			Wikipedia
	Only secondary resources		Targeted/Area specialized websites	Targeted/Area specialized websites		Targeted/Specialised Websites	
		Primary and specialised, communicative secondary resources	Hospital/University Websites		Highest use of GSE	Society Websites	
			Primary resources, secondary resources, specialised and academic resources	Primary resources		Hospital/University Websites	
				Specialised secondary resources	Primary and secondary resources		

Use of Physician society communities	Most likely to use physician society communities (33,9- or 51% of those who are aware of them use them!)	10,8% use physician society communities	18,8% use physician society communities	15,4 % use physician society communities (Most awareness of their existence- 72,5 % are aware)	Least likely to be aware and use physician society communities (49,2% do not know of their existence, 9,8% use them)	16,7 % use them	18,1 use them
Group specific							

	SEGP	SES	HP	RP	PIT	Young	Old
Format of resources ?	Among all groups most popular format of Information: 1. Print and Image 2. Video						
Overall							
Group specific			More likely than other groups to place importance on video (2,88)		More likely than other groups to place importance on video (2,88)		
What type of queries?			Overall :				
Overall			Most likely to search by keywords, diagnosis				
Group specific	More likely than other		More likely than other groups to	More likely than other groups to	More likely than other groups to		

D8.2 Use case definition including concrete data requirements

groups to search by symptom (24,6%)	use question format 10,8%	search by diagnosis (77%)	search by symptom (22,6%)
	More likely than other groups to search by diagnosis (76,4%)		

SEGP	SES	HP	RP	PIT	Young	Old
------	-----	----	----	-----	-------	-----

Results presentation requirements:

**Most
important for
ideal search
engine ?**

Among all groups popular characteristics:

Relevance of results

trustworthiness of results

Group specific	Quality of link description	Quality of link description	Speed	Speed	Readability of results	Speed	Quality of link description
How many links/page ?	5-10 (32,3%) links/page or 10-20 (32,3%) links/page	5-10 links per page (40,2%)	10-20 links per page (41,4%)	10-20 links per page (37,8%)	5-10 (38,7%) links/page or 10-20 (38,7%) links/page	10-20 links per page (34,6%)	5-10 links per page (37,7%)
Group specific	More likely than the other groups to prefer below 5 links/page (25,8%)			More likely than the other groups to prefer 10- 20links/page (27,8%)			

**Link
description**

**Automatic
categorization
? All groups**

All groups (cut-off 40%) Type of content and type of author

Group specific	Additionally used type of target audience and quality accreditation marks (more likely than other groups)	Most likely (and more likely than other groups) used "type organization" More likely than other groups to use "type of author"	Additionally uses type of "target audience"	Additionally uses type of target audience	Additionally uses target audience
			Most likely group to categorize by "Type of content"	Most likely to use "Type of content" than older physicians	

Among all groups important quality criteria: Publisher

Quality markers ? What criteria physicians want to view to judge the	Time of the last update	References to original source	Author	Author	Author	Author	Time of the last update
---	----------------------------	----------------------------------	--------	--------	--------	--------	----------------------------

D8.2 Use case definition including concrete data requirements

quality of a link

	SEGP	SES	HP	RP	PIT	Young	Old
Search behaviour ? Possibility to skim through pages ?	Overall						
The first click ?	Checks second and third page of results						
Overall	First click on the link that looks the most relevant						
Group specific	Checks second and third page of results	Checks second and third page of results	Checks second and third page of results	Checks second and third page of results	Checks second and third page of results	Checks second and third page of results	Checks second and third page of results
	Faces situations where they cannot find the answer to a medical question in the Internet	More likely than other groups to click on link that is most trustworthy (34,1%)		More likely than other groups to click on the first link that appears	More likely than other groups to click on the link that is most relevant (77%)	More likely than older physicians pages that are relevant	
	More likely than other groups to click on link that is most trustworthy (39,3%)						

	SEGP	SES	HP	RP	PIT	Young	Old
Interface Requirements:							
Advanced search options- Ranking ?	Overall: In all groups						
	Date range						
	Language						
	were the most important advanced option						
Group specific	1. Date range 2. Language		Date range	Language	1. Date range	1. Date range 2. Language	
		More likely than other groups (except GP) not to use advanced search options and most likely to use Date		More likely than other groups to use "Country" and "Document format" as an advanced search			

D8.2 Use case definition including concrete data requirements

	range			option			
Advanced search- What filter?	Type of Source	1. Type of Source	1. Journals	Journals	Journals	1. Source	1. Journals
		2. Journals	2. Type of Source			2. Journals	
Group specific	More likely than other groups to use "Type of source"		More likely than other groups (except research physicians) to restrict by "News"	Most likely group to restrict results by "Journals"	Most likely group to restrict output by Books, Videos, Blogs/Forums and Images		
	More likely than other groups (except hospital physician) do "not use restriction criteria"			More likely than other groups (except hospital physicians) to restrict by "News"	Most likely group to "not use restriction"		
	More likely than other groups to ask for Automatic translation of results						
	SEGP	SES	HP	RP	PIT	Young	Old
What tools are required ?	Popular tools mentioned as important by ALL groups:*						
Overall	1. Being able to quality rate information/websites and perceiving ratings of other physicians (51,7%) 2. Advanced search (44,3%) 3.Possibility to chose between content filters (43%,) 4. A list of popular websites (where most users have found the answer to a query) (45,6%,) 5. Possibility to chose between content filters (43%,) 6. Suggested relevant topics (38,2%) 7. Search of images (32,9%) 8. Possibility to view search history (31,2%)						
Group specific	Integration of patient data within the search process as a diagnostic aid for complex cases (42,9%)	Integration of patient data within the search process as a diagnostic aid for complex cases (36%)	Integration of patient data within the search process as a diagnostic aid for complex cases (42,9%)	Possibility to store self-made compendia (34,9%)	Use from mobile platforms (44,8%)	Use from mobile platforms (44,6%)	FAQ- How to use a search engine (31,1%)
	Medical calculators (42,1%)	FAQ- How to use a search engine (28,4%)	Use from mobile platforms (40,2%)	Integration of patient data within the search process as a diagnostic aid for complex cases (32,2%)	Spelling correctionMedical calculators (35,1%)	Younger physicians more likely to seek:	Older more likely than younger to seek:
	FAQ- How to use a search engine (28,1%)	Word cloud (33,3%)		Use from mobile	Possibility to store self-made compendia (32,1%)	Use from mobile platforms**(p>0.01) (44,6%)	FAQ- How to use a search engine** (31,1% p<0.01)
	Automatic	Medical calculators					

D8.2 Use case definition including concrete data requirements

[illegible]

D8.2 Use case definition including concrete data requirements

Possibility to
view search
history (34%)

Automatic
completion of
query (23,8)

SEGP	SES	HP	RP	PIT	Young	Old
------	-----	----	----	-----	-------	-----

Barriers

**Why do they fail to
find an answer ?**

All groups most common reason for failing to find answer to a query is:

Search is too time consuming

Overall

Group specific	More likely than other groups to perceive that search is too time consuming (25,9%)	More likely than other groups to fail finding an answer due to formulation difficulties (26,2%)	More likely than other groups to fail finding an answer due to formulation difficulties (63,6%)	More likely than other groups to perceive that search is too time consuming (74,1%)	Older Physicians (26,3%) report higher levels of formulation difficulties than younger physicians and other groups
----------------	---	---	---	---	--

**What happens
when an answer is
not found ?**

For all groups most popular action when not finding an answer:

Do another search getting more specific

All groups

Group specific	2) Stop searching on the Internet 3) Chat to colleague Most likely group to chat to a colleague (13,1 %) or post the question into a medical forum (4,9%)	2) Stop searching on the Internet 3) Less specific Most likely group to stop searching in the Internet (17%)	2) Pursue another less specific search 3) Stop searching on the Internet Most likely group to pursue a more specific search (66,7%)	2) Pursue another less specific search 3) Chat to colleague	2)) Pursue another less specific search 3) Chat to colleague Most likely group to do another less specific search (25%)	2) DO a search getting LESS specific (21,8) 3) Chat/Skype with a colleague (7,4%) 3) I stop searching on the Internet (13,2%)
----------------	---	--	---	--	--	---

Further barriers?

All physicians highlighted the following barriers:

D8.2 Use case definition including concrete data requirements

All groups		Restricted accessibility to good quality resources, Time consuming to find relevant information Search results are too general, Questionable trustworthiness					
Group specific	Additional: Absence of good quality/filter	Additional: Search results too general Dominant barrier= Restricted accessibility and barrier: Time consuming More likely than other groups: Time consuming to find relevant information (56, 2%)	Additional: Search results too general Time consuming Dominant barrier: Restricted accessibility	Additional: Time consuming Questionable trustworthiness Dominant: Restricted accessibility More likely: To find advertisements distracting	Dominant restricted accessibility Questionable trustworthiness Search results too general Time consuming More likely than other groups: Restricted accessibility to good quality resources (81,7%- GP least likely)) Search results too general	Young more likely than older: to find search results too general, find information restricted access for mobile device users	Older more likely to find it time consuming to find relevant information

Access

How connect All groups in order of importance access via 1. Modem/Cable DSL 2. WIFI 3. Smartphone

All groups

Group specific	More likely to use Modem/DSL than other groups (85,5%)		More likely that other groups to use USB (16,2%)	More likely than other groups to use WIFI (55,6%)	More likely than other groups to use Smartphone (38,7%)		
Main Device used during patient consultation	Laptop	Laptop	Laptop	Laptop	Smartphone	Laptop	Laptop,
Groups specific	Most likely to access the Internet in this situation		Most likely to use a Tablet in this situation (4,2%) compared to the other groups		Most likely to use a Smartphone in this situation (16%). Least likely to use the PC in this situation and most likely to report never accessing the Internet in this situation (38%)	Younger more likely than older to use a Smartphone and to access the Internet in this situation	Older more likely than younger to use a Tablet

D8.2 Use case definition including concrete data requirements

Main device used at work?

All groups: PC

All groups

Group specific Tablet (2,2%) Tablet (4,6%) Tablet (1,2%) Smartphone (5,4%)

Main device used during conference

Overall:

Top 2 devices among all groups is the Laptop and Smartphone

Group specific Least likely to access the Internet in this situation (27,6% never access) Most likely to use a tablet in this situation (9,2%) Most likely to use a PC (3,8%) or Laptop in this situation (58,2%) Most likely to use a Smartphone (40,7%) in this situation Younger physicians more likely access via a Smartphone (36%) than older physicians (23%) Older Physicians more likely access via Tablet (9,4%) than younger physicians (4, 5%)

Main device used in meeting colloquium ?

All groups: 1. Laptop 2. Smartphone

Least likely group to access the Internet in this situation (42,4% never access) Most likely group to use a tablet in this situation (5,1%) Most likely group to use a Laptop (46,3%) in this situation Most likely group to use a PC (21,27%) and a Smartphone (25%) in this situation More likely than older to access via a Smartphone (23%) than older physicians (11,6%) More likely Than Younger (1, 2%) to access via tablet (6,8%)

Main device used at home?

Overall: Most popular devices were PC and Laptop (varying extent of popularity- see below for ranking)

1. Laptop	1. PC	1. PC	1. Laptop	1. Laptop	1. Laptop	1. PC
2. PC	2. Laptop	2.Laptop	2. PC	2. PC	2. PC	2.Laptop

Other information

Willingness to pay ? All groups: Most prefer to have ads in order to keep a search service free

Overall

Group specific	More likely than other groups to prefer membership subscription	More likely than other groups to prefer to donate and avoid ads	More likely than other groups to prefer to donate and avoid ads	More likely than other groups to prefer having ads in order to keep	More willingness than older physicians to keep ads to keep a search	More Willingness than younger physicians,
----------------	---	---	---	---	---	---

D8.2 Use case definition including concrete data requirements

	to avoid ads (26,7%)	placement (17,8%)	placement (17,8%) or pay a fee for service (11,1%)	search service free (Both 73%)	service free (71,2 vs, 55,5%)	to subscribe to a membership subscription to avoid ads (22,3% vs. 14.2%)
Consequences of online research	All groups: Conduct research using other resources					
Overall						
Group specific	Print out information for the patient or recommend a website to the patient ($p>0.05$)			Ask a colleague for his/her opinion ($p>0.05$)		

Annex 2. General public user profiles (based on detailed survey analysis)

	Less than 40 y.o. males	Less than 40 y.o. females	From 40 to 59, males	From 40 to 59, females	From 60 y.o., males	From 60 y.o., females
1. Typical information needs:	Treatment, detailed and general disease description					
	Patients groups are more important	Patients groups are more important		Detailed and general disease description		Detailed and general disease description
2. Data sources:	1) GSE52, 2) Wiki, 3) Links, 4) Hospital-University-Pharma, 5) Suggested by a Healthcare provider, 6) Forum and blogs 7) Specialized search tools					
	GSE	GSE		Forum, blogs		Forum, blogs
4. Results presentation requirements:	1) Relevance 2) Readability of results 3) Quality of description 4) Trustworthiness 5) Simplicity					
			Quality (3) is especially important	Quality (3) is especially important		
		2), 3), 4) and 5) are more important		2), 3), 4) and 5) are more important		2), 3), 4) and 5) are more important
3. Interface Requirements:						

D8.2 Use case definition including concrete data requirements

1) Before search:	<ul style="list-style-type: none">• Advanced search (country, language, dates)• Spelling correction• Suggested filling of query• Search of images, audio, video podcasts					
<i>No differences revealed</i>						
2) After search:	<ul style="list-style-type: none">• Suggested relevant topics,• Word cloud,• Automatic translation of the results,• Sharing the results with Doctor/friend					
<i>No differences revealed</i>						
3) Tools:	<ul style="list-style-type: none">• Med dictionary/thesaurus,• Risk factors assessment tools• Body 3D visualization• Tutorial on strategy of search• Updates by email/RSS• Use from mobiles• Hotline/web chat support• Listed and save as mp3• Letter size and font customization					
					Risk factors assessment tool is regarded as more important	Risk factors assessment tool is regarded as more important
6. Multilingual features: <i>No differences revealed</i>						
7. Other notes:	Internet is the top preference	Internet is the top preference	Doc consultation moderately important comp with 60 plus	Doc consultation moderately important comp with 60 plus	Doctors consultation, friends, family member (use for the interface)	Doctors consultation, friends, family member (use for the interface)

D8.2 Use case definition including concrete data requirements

	Radio, TV			Radio, TV		Radio, TV
Internet because of support for diseases	Internet because of support for diseases	Internet because of support for diseases	Internet because of support for diseases	Internet because of support for diseases	Internet because of lack of money	Internet because of lack of money
Checking a web site he/she trusts to verify the answer	Checking a web site he/she trusts to verify the answer				Asking doctor after search to verify the answer is more important	Asking doctor after search to verify the answer is more important
Search results do not help to find the answer	Search results do not help to find the answer	Search results do not help to find the answer	Search results do not help to find the answer	Search results do not help to find the answer	Too many search results	Too many search results
	Tend to ask more often the physician when no answer was found online			Tend to ask more often the physician when no answer was found online		Tend to ask more often the physician when no answer was found online

Annex 3. Set of questions for semi-structured interviews of general public

Which web site you use for searching of the information?

What do you type?

Do you type in your mother tongue?

Search results

- Which links do you click?
- Do you check the second page?
- Do you know what the upper highlighted links are about?
- Are you happy with the results which appear in your mother tongue?

If you are happy

- Have you found an answer to your question?
- Where have you found an answer?
- How much time have you spend to find an answer?
- If you want more specific information, you go back to search engine, or stay on ther web site?

If you are not happy:

- What will be your strategy?
 - to reformulate the query,
 - to go to another web site,
 - to type it in another language,
 - to contact someone?
- Would you like the system to propose you these options in stead of you doing it?

Translation :

- Are you interested in checking the results different from your initial query language? For example, your query is in French, would you like to see the results in English?
- Do you want the system automatically translate your query into English, French, Spanish? Or you prefer to manage it yourself?
- Do you prefer to read the results in your mother tongue only, or translated from other languages to your mother tongue?
- Do you prefer to read the results in the original language (for example, English) as you feel comfortable with understanding English?
- Imagine, you look for the same information while you are based in a different country.
- How would you prefer the results to be displayed?

Annex 4. HON logs analysis

Grant Agreement Number: 257528

KHRESMOI
www.khresmoi.eu

HON Logs Analysis

Deliverable number	<i>Attached to D8.2</i>
Dissemination level	<i>Public</i>
Delivery date	<i>8 February 2012</i>
Status	<i>Version Final</i>
Author(s)	<i>Alexandre Masselot, Daniela Sinjari</i>



This project is supported by the European Commission under the Information and Communication Technologies (ICT) Theme of the 7th Framework Programme for Research and Technological Development.

Abstract

The main motivations of HON Logs Analysis project described in this paper, are to study user behaviours by examining and exploiting the underlying structure and nature of query log coming from different search engines. Query log, which are files of searching interaction between users and the retrieval system, often contain indications of relevance judgements from the perspective of selected documents in the context of a query.

This paper reports the work being done on gathering, preparing, analysing the query log independently from their source origin and file format. Furthermore, the first results of HON log analysis are presented.

1 Introduction

HON-logs-analysis is an open-source project which is hosted on Google Code server. Its goal is to improve information retrieval by exploiting historical queries contained in query log coming from different retrieval system such as Khresmoi, HON search engine etc. The work is focused on gathering, preparing, analysing the query log independently from their source origin and file format. These steps are presented in details throughout this paper.

The main motivations of this project are to study user behaviours in the context of the work being done by wp8 which aims at the current stage to define user profiles and elaborate user cases.

Query log, which are files of searching interaction between users and the retrieval system, often contain indications of relevance judgements from the perspective of selected documents in the context of a query. We can gain a clearer understanding of the interactions among searcher, content and system by conducting an analysis with the aim of achieving some objectives, such as better understanding of some user information searching behaviour, search engine design improving with advanced searching assistance etc. HON-log-analysis project aims to extract the most significant information from query log for further analysing data using different statistical metrics.

Currently, HON logs are collected and analysed. In the future this software will be also used to understand Khresmoi user behaviours.

2 HON-log-analysis Description

HON-logs-analysis project treats different query log file format coming from different Web searching servers (HON, Khresmoi etc). Its services logic is well distinguished so that it can support query log parsing independently from the coming sources and query log file format. Any kind of coming sources and query log file formats can be treated by this software.

Basically, the search log file contains data such as the client computer's Internet Protocol (IP) address, search engine access time, user query, and referred site, among other fields. Once the server collects and stores the data in a file, it's important to analyze this data in order to extract relevant information.

2.1 Development Platform

2.1.1 Environment

HON-log-analysis is a web project hosted on Google code server. SpringSource Tool Suite has been used as development environment, supplied with Grails and Groovy based technologies. Grails framework simplifies testing by supporting unit and integration testing. In addition, it is based on the Model-View-Controller (MVC) architectural pattern by isolating domain logic from user interface permitting independent development, testing and maintenance.

2.1.2 Team sharing

Subversion support for SpringSource Tool Suite is used to allow team members working parallelly on the HON-log-analysis project hosted on Google code server, by getting the latest version from the server and updating it frequently with new changes (meanwhile maintaining current and historical versions of project files).

2.1.3 Programming Language

The language used for implementation is Groovy, a dynamic language for java platform.

2.1.4 Database management system

HON-log-analysis project based on Grails framework is so versatile in the way that it can support many kinds of database management system (DBMS) for database storing. By following the paradigm of 'convention over configuration', it's a matter of choice to define what kind of DBMS one prefer to use to store query log data in database. HON-log-analysis design permits to store different type of query log format into database.

Grails Object Relational Mapping (GORM) API provides the means to save object properties from a grails application into database tables dynamically. It permits to create relationships between tables, enforce data validation rules, and change the object-relational database in our Grails application.

PostgreSQL (often simply Postgres) is used in HON-log-analysis project as database management system for storing cleaned and prepared data from query log into a database. It's an open source object-relational database system which fits pretty well with Grails applications.

2.2 Data Collection

This is the process of collecting the interaction data for a given period in query log. Web servers can record various types of data from the interactions between searchers and search engines in log file depending on the file format that the server software supports. In different levels of analysis, there will be used HON query log collected for a period from 24 November till 19 December of 2011.

D8.2 Use case definition including concrete data requirements

Fields in HON query log

Remote Ip	Date	Query Text
76.168.71.58	11/Dec/2011:00:00:15 +0100	Adie+Syndrome
76.168.71.58	11/Dec/2011:00:00:43 +0100	Caroli+Disease
88.74.122.100	11/Dec/2011:00:02:42 +0100	Infections+bact%C3%A9riennes+et+mycoses

Tab. 1 The query log format collected by HON search engine

Each row in Table 1 represents a searching interaction. The format of each searching interaction is as follows:

- Remote IP: the IP address of the client's computer.
- Date: the date and time of the interaction as recorded by the search engine server.
- Query Text: the user query, as recorded by the search engine server software in its supported query log file format.

2.3 Data preparation

Once the data are collected, the focus is on parsing and cleaning the user interactions data , in order to extract the most significant information and to import it into a database which will serve as the basis for further analysis.

2.3.1 Parsing, Cleaning the data

Several services are implemented in HON-log-analysis project for parsing different query log format which come from different sources (dbs, hon, tel). Their logic design permits to parse query log files independently from their format.

User interaction data extracted from query log may contain non-human interactions as recorded by search engine server software which is desirable to be removed. The query text may also need to be cleaned to remove stopwords, bad data like numbers, punctuation, special characters.

Stopwords are considered as extremely common words used in search engines which mostly don't have weight in information retrieval. A list of English and French stopwords was taken into consideration such as prepositions, auxiliary verbs, pronouns etc (i.e. fr: aux, avec, ce, il, je etc. i.e. en: after, again, can't, cannot etc.)

Regular expressions have been used in HON-log-analysis project for extracting fields of Remote IP, Date and Query Terms (common to all Web query log), from query log records. A regular expression provides a concise and flexible means for matching patterns of characters. Records in query log can contain noisy data. A way of rapidly ignoring most noisy data is to take into account only the fields of interest in sequence. The list of query terms is checked if it contains stopwords and therefore is cleaned.

2.3.2 Database System

HON-log-analysis design can handle different kinds of database management systems, it's a matter of choice to define what kind of DBMS one prefer to use to store query log data in database. For the aim of this project, PostGres database was just an optional choice among the others. In addition, HON-log-analysis services design permits to store different type of query log format into database, no matter where they're coming from. Three tables have been created dynamically, each belonging to the source where query log come from. Each table corresponds to a specific query log distinguished by its fields format, many of which can provide valuable information (see Fig. 1). This database containing several query log, will serve for further data analysis.

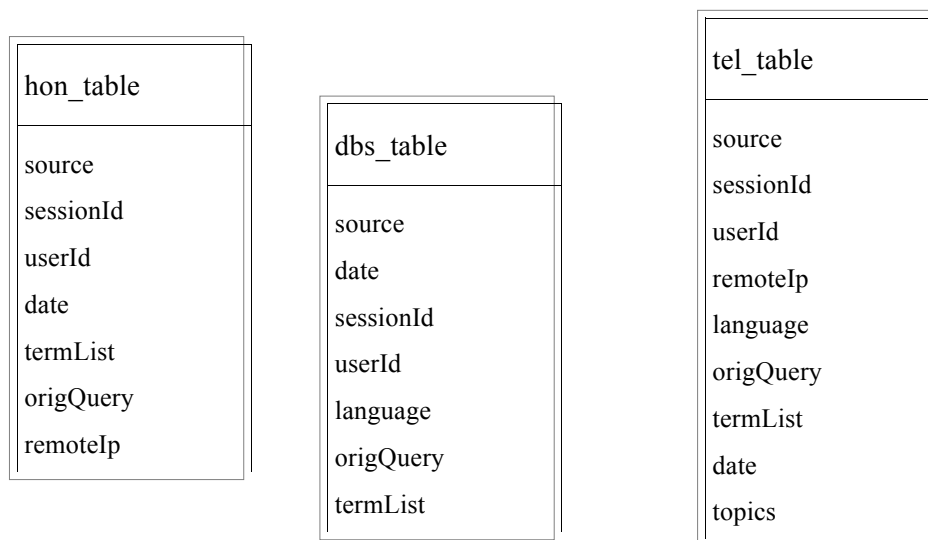


Fig. 1. Tables format for storing query log, each belonging to the source where they come from

2.4 Data Analysis

This step focuses on the process of analyzing the prepared data. We're going to step through different levels of data analysis by using HON query log collected for a period from 24th of November till 19th of December 2011. Some metrics of analysis for examining query log such as term frequency, term co-occurrence, query length, nb. of queries by language and country origin, nb.of queries per day etc. are discussed below.

2.4.1 Term analysis

At this level of analysis, we are focused on measuring term occurrence, which is the frequency of a specific term occurring in the query log. In the Fig. 2 below we can observe the high usage terms which are those terms that occur most frequently in the data set.

Total nb terms in query log: 5165

D8.2 Use case definition including concrete data requirements

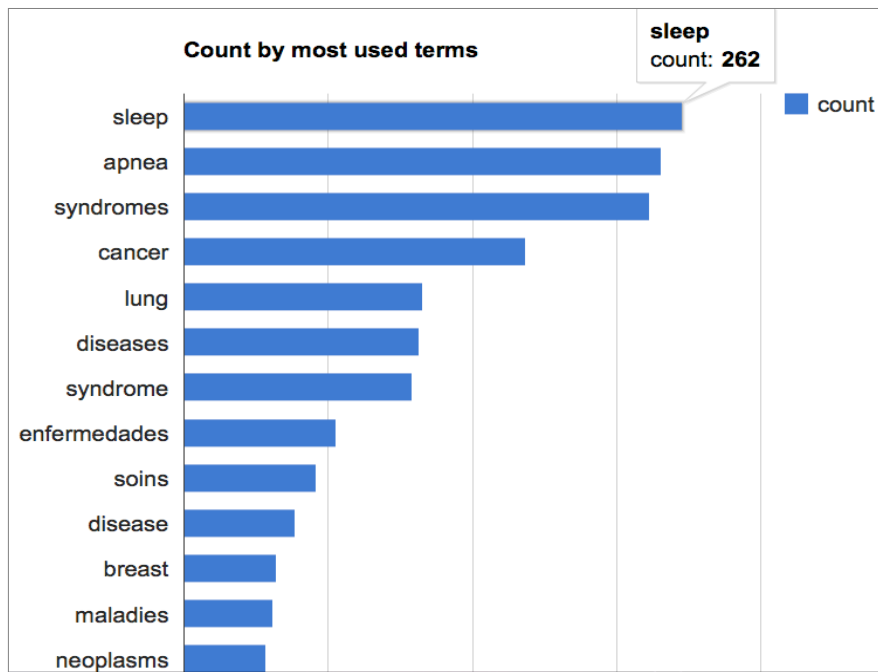


Fig. 2. TOP 13 frequent terms within query log

We can observe that *sleep* and *apnea* are the most used terms typed in HON search engine, with respectively 262 and 252 hits. *antibioticos* and *quirurgica* are the less used with 25 hits.

The graphic below (Fig. 3) displays term co-occurrence which measures the occurrence of term pairs within queries in the entire query log.

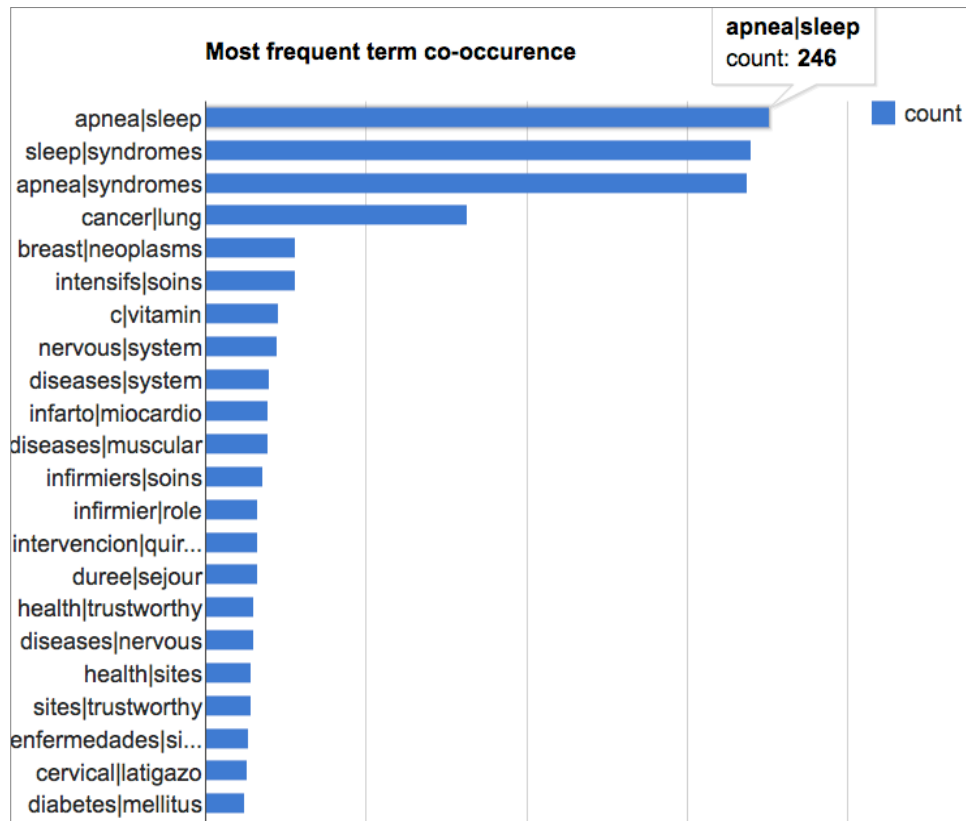


Fig. 3. TOP 22 frequent term pairs within the query log

We can observe that the apnea|sleep and sleep|syndromes pairs are the most used together in user query formulations.

2.4.2 Query analysis

This analysis uses the query as the base metric. A query is a list of terms submitted to a search engine.

Total nb queries: 9660

D8.2 Use case definition including concrete data requirements

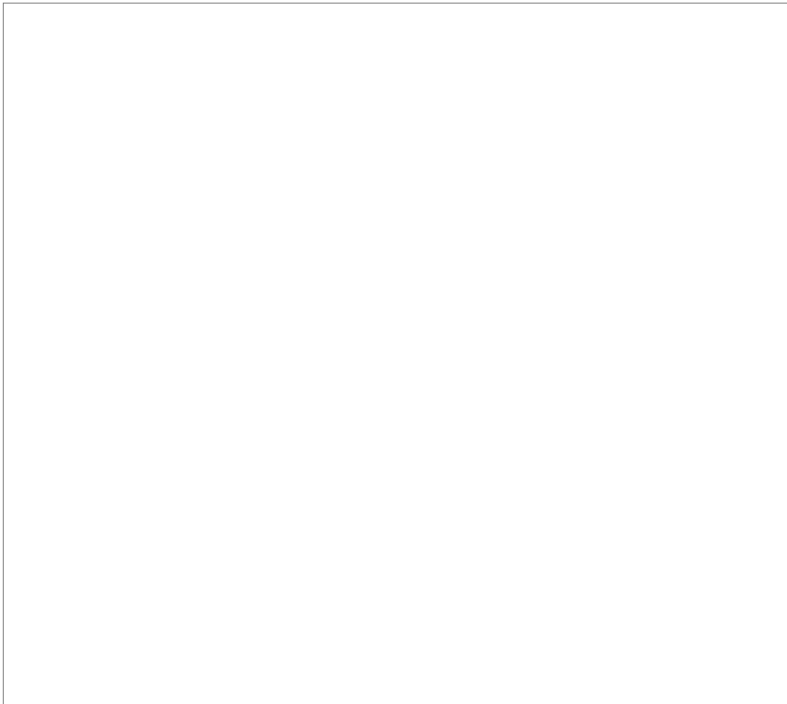


Fig. 4: A count of the occurrences of query lengths

Queries with one term are mostly spread, and ranked at the top of the list with a count of 2649 queries. Only 10 queries contained 6 terms.

2.4.3 Analysis by country origin

Nb. of IP addresses: 2558

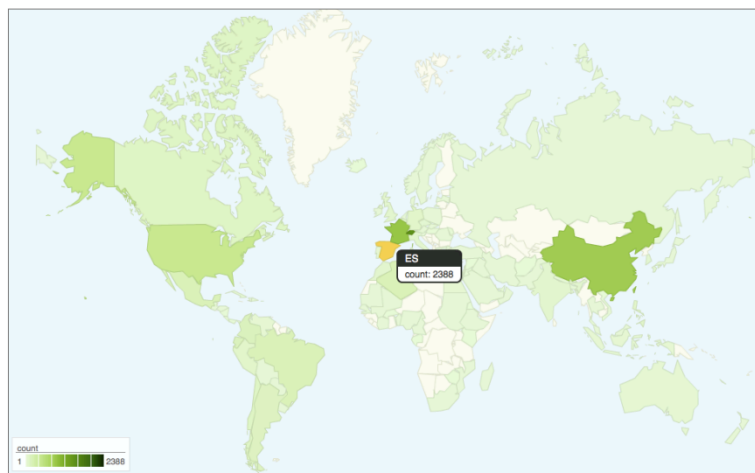


Fig. 5. Queries by the country of origin

There have been identified 2558 distinct user ip addresses as recorded in query log, which can give some relevant information concerning the location of users using the HON search engine.

D8.2 Use case definition including concrete data requirements

The countries origin of searching interactions displayed in darker colour represent the countries with highest number of searching interaction recorded in query log. 2388 IP coming from Spain were identified.

2.4.4 Number of queries per day

From a total number of 9660 query log as recorded by the search engine, only 8181 logs contains non-empty queries. The other recorded log which don't contain query terms may be results of simple button clicks on the empty search engine textbox. An average of 315 queries per day have been identified.

Total nb of query log: 9660

Total nb. of queries: 8181

Averaged nb. of queries per day: 315

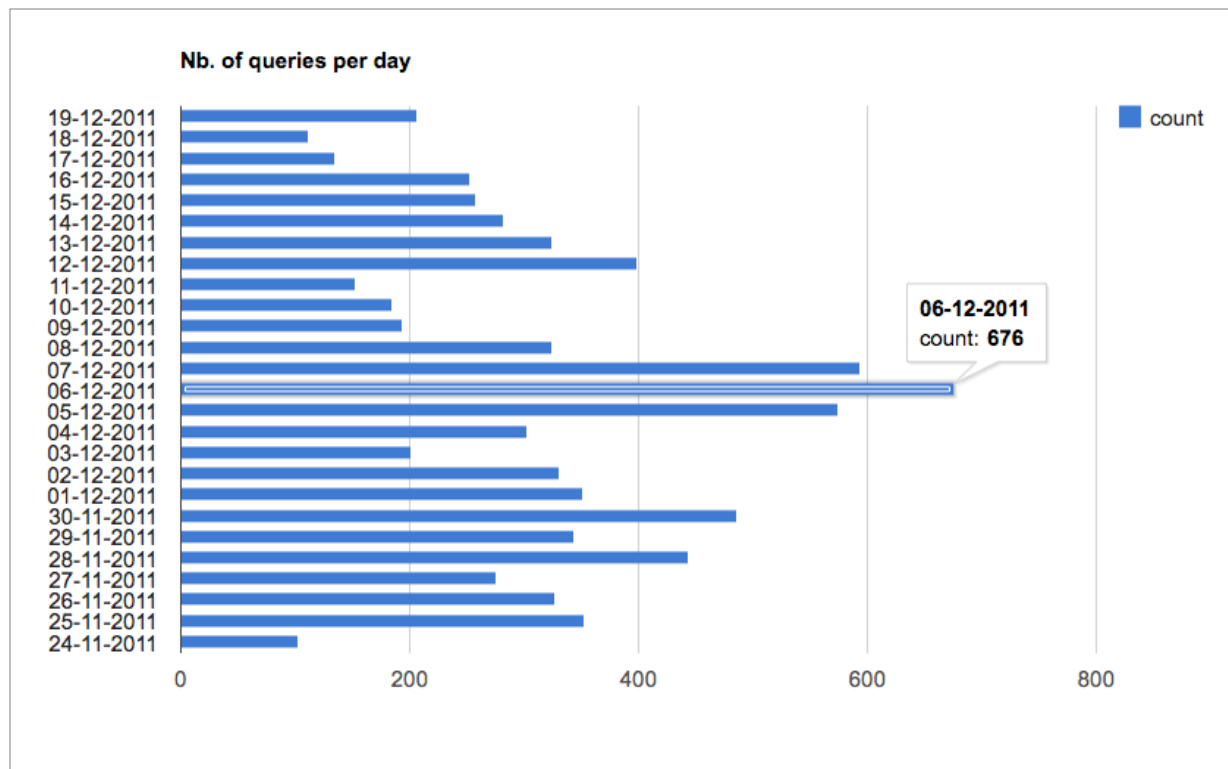


Fig. 6. Number of queries per day

The results in the figure above show the number of submitted queries per day, during the period 24 Nov- 19 Dec 2011. The flux of submitted queries varies during the selected period under observation, with a peak of 676 queries recorded on 6th of December, while an average of 315 queries per day has been identified.

2.4.5 Countries by term

D8.2 Use case definition including concrete data requirements

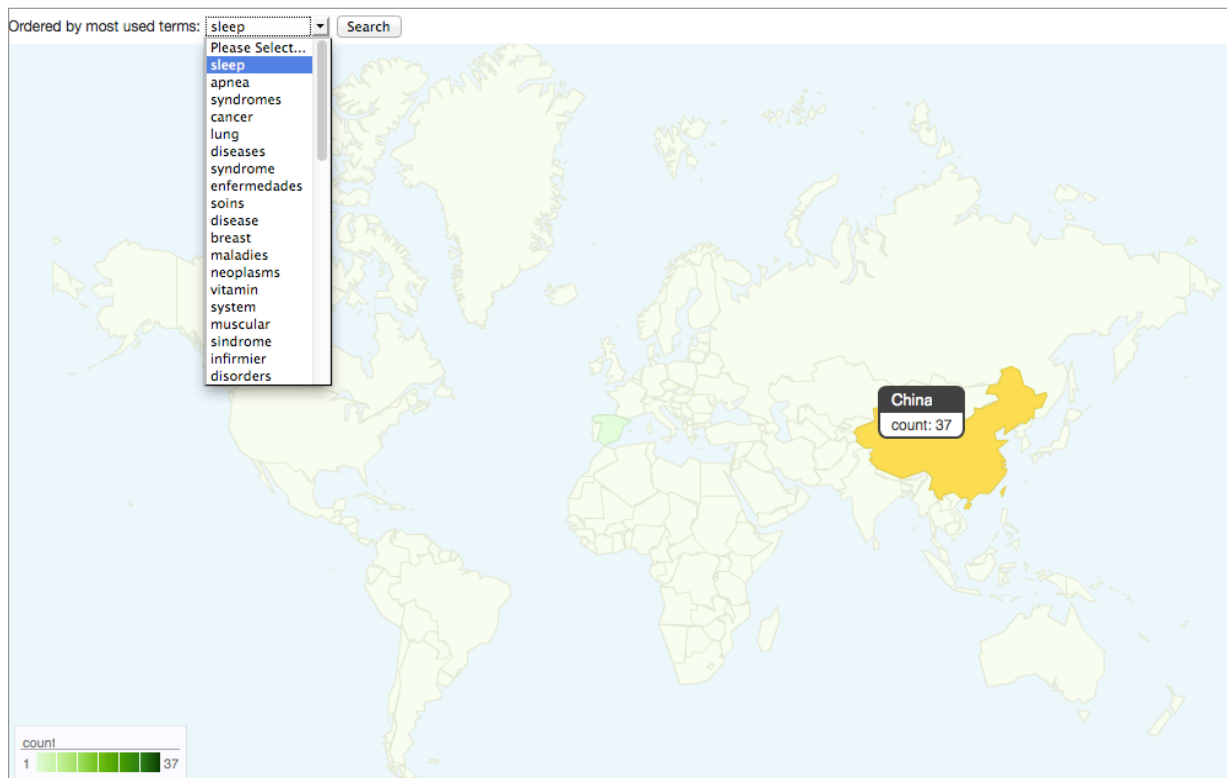


Fig. 7. Geographical representation of the countries the most used terms were searched

The map above displays the countries origins for each of most used terms. We can observe that the maximal number of hits for the term 'sleep' (which is the most used term in queries collected for the period 24 Nov-19 Dec 2011) comes from China, from 37 distinct user IPs identified followed by Spain with one hit.

2.4.6 Number of queries and details by user

Once identifying each user by his unique Id as recorded in query log, it's possible to construct a user searching profile based on the set of his submitted queries on HON search engine. The table 2 below, gives a statistic overview of total number of queries typed by user in descending order. It needs to be underlined the fact that the total number of queries per user refers to all the searching interactions of each user during an interval of several days as recorded by server software. Furthermore, we can access more detailed information for each user in order to understand behaviours in their searching experience.

D8.2 Use case definition including concrete data requirements

User	Nb. of queries	Details
200.117.211.37.1322510296876473	92	View Details
62.39.11.146.1322212510124391	70	View Details
200.54.91.180.1322326182236754	50	View Details
79.147.119.76.1322327857851713	38	View Details
195.176.237.21.1323070918307667	37	View Details
95.124.71.55.1324229632271423	34	View Details
195.176.237.20.1323070931225716	33	View Details
89.107.243.1.1323677849647026	33	View Details
195.176.237.20.1323157662118356	29	View Details
195.176.237.20.1323157681156291	29	View Details
88.8.128.19.1322313448444986	28	View Details
193.144.61.240.1322733175862626	28	View Details
193.144.61.240.1324033621051131	27	View Details
195.176.237.21.1323070907237541	25	View Details
195.176.237.21.1323157680777231	25	View Details
79.152.7.48.1322396704126373	24	View Details
195.176.237.21.1323070922437517	24	View Details
190.42.53.194.1323499116267025	24	View Details
193.144.61.240.1323702215224226	24	View Details
195.176.237.21.1323070922352362	23	View Details
195.176.237.21.1323070923153236	23	View Details

Tab. 2. Number of queries by user id

We can observe in the Table 2, the whole searching experience of a specific user, by displaying his country origin, the list of query terms he types in each query, and measuring the time elapsed before the user types another query.

Details of the user: 90.15.8.81.1322501368845854	
Country: France	
Time elapsed:	Query Terms:
00"	[mortalite]
09' 11"	[mortalite, maternelle]
03' 22"	[consultation]
29"	[prenatal]
12"	[prenataux]
34"	[maternite]
11"	[grossesse]
50"	[grossesse]
02' 44"	[prenatale]
14d 14h 52' 52"	[psychotherapie]
28"	[psychologie, clinique]
07h 39' 48"	[deviation]
09"	[cloison]
26"	[stenose]
06'	[echarde]
33"	[corps, etranger]

Tab 3. Search performed by a specific user

From the results of the table above, we can see that a user identified by his unique ID who lives in France has started the searching interactions with HON search engine by typing the first query composed by a single term 'mortalite'. After 9 minutes 11 seconds, this user details his query by typing "mortalite, maternelle". After 3 min 22 sec, the user makes a different query 'consultation' etc.

3 Conclusions

Query log are important means for collecting data concerning the user–system interactions. Using these data, one can conduct different analysis for better understanding users behaviour in searching information in search engines. The results from different levels of analysis presented in this paper such as most used terms, most used term pairs, occurrences of query length etc provides us a diversity of information regarding our data to be further investigated. It serves as the basis for further exploitations in term relationships, advanced searching assistance etc.

4 References

- [1] Chau, M., Fang, X., and Sheng, O.R.L. Analyzing the query logs of a Web site search engine. Journal of the American Society for Information Science and Technology.
- [2] Jansen B.J, Search log analysis: What it is, what's been done, how to do it/ Library & Information Science Research 28 (2006) 407–432

Annex 5 Useful data sources for physicians

The list below provides examples of useful data sources for physicians. The aim is to provide an overview of what type of concrete data resources are required by physicians with respect to the proposed categorizations. The focus of the proposed resources is on secondary resources within Austria to emphasize the need of local, secondary resources. In the course of the project the list will be extended to include worldwide resources.

General sources

Pubmed (Primary resource)	http://www.ncbi.nlm.nih.gov/pubmed/
Cochrane	http://www.cochrane.org/

Secondary Resources

General information/Medical updating	
Springer medizin	http://www.springermedizin.de
AVMF	http://www.awmf.org
Medizin Medien	http://www.medizin-medien.at/
Universimed	http://universimed.com/
Univadis	http://www.univadis.at/
UptoDate	http://www.uptodate.com
Best practice	http://bestpractice.bmj.com
Docguide	http://www.docguide.com/general-practice/popular/30days
WebMD	http://www.webmd.com/Im Cache
Medscape	http://www.medscape.com
Wikipedia	http://en.wikipedia.org/
Austrian Magazine for physicians	http://www.aerztemagazin.at
Austrian physician newspaper	http://www.aerztezeitung-oesterreich.at/

Drugs

Drug Database- Austria Codex	http://www.univadis.at/medical_and_more/de_AT_AustriaCodex_Detail?Refresh=true
Safety of drugs	http://www.fda.gov/

Guidelines:

D8.2 Use case definition including concrete data requirements

AWMF (Die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) Diabetes	http://www.awmf.org/leitlinien/leitlinien-suche.html
OEDG Guidelines German	http://www.oedg.org/oedg_leitlinien.html
OEDG Guidelines English	http://www.deutsche-diabetes-gesellschaft.de/redaktion/mitteilungen/leitlinien/Uebersicht_engl_LL.php
International Diabetes Federation - Guidelines	http://www.idf.org/guidelines/guidelines-development

App Resources

Tapforms (App Database)	http://www.tapforms.com/
Itunes Medical Apps	http://itunes.apple.com/app/dg-apps/id445912798?ign-mpt=uo%3D5

Medical conferences /events

Veranstaltungen der GAW	http://www.billrothhaus.at/index.php?option=com_vf_veranstaltungskalender&Itemid=139
Berufsverband Deutscher Internisten	http://www.bdi.de/fortbildungen/kurse-kongresse.html
Fortbildungskalender	http://www.der-Fortbildungskalender.de/index.php?bundesland=&typ=normal&datum=&datum_max=&Buchungscode=&action=search

Diabetes

German Diabetes Society	http://www.deutsche-diabetes-gesellschaft.de/redaktion/veranstaltungen/jahres_herbsttagungen.php
-------------------------	---

Pulmonology

Austrian pulmonology society	http://www.ogp.at/aus_und_fortbildung/veranstaltung/en/index.php
------------------------------	---

Online CME

DFP Gesellschaft der Ärzte Wien	http://www.billrothhaus.at/index.php?option=com_vf_elearning&task=dfp
Medizin Medien	http://www.medizin-medien.at/dynasite.cfm?dsmid=69263

D8.2 Use case definition including concrete data requirements

Universimed- DFP	http://www.med-diplom.at/
Mein DFP	www.meindfp.at
Fortbildungskalender	http://www.der-fortbildungskalender.de/index.php?bundesland=&typ=online&datum=&datum_max=&Buchungscode=&action=search
OEDG Online Videos	http://www.oedg.org/onlinevortraege.html

Medical news

Univadis	http://www.univadis.at/medical_and_more/News#
Universimed	http://universimed.com/latest-news

Medical Forum

Medical forum for physicians in training	http://www.assistenzaerzte.at/forum/turnusarztcom/allgemeines
Medical forum for physicians	http://www.hippokranet.com/de/forum

Further Education

Univadis- 3D Anatomy	http://www.univadis.at/medical_and_more/Anatomy_Intermediate
----------------------	---

Social insurance companies in Austria (presenting a list of governmental hospitals in Austria)

Vienna	http://www.wienkav.at/kav/spital.asp
Oberösterreich	http://www.gespag.at/37388.php
Kärnten	http://www.kabeg.at/
Steiermark	http://www.kages.at/cms/ziel/5646146/DE/
Burgenland	http://www.krages.at/Standorte.5.0.html?&L=%00ttkawe
Salzburg	http://www.salk.at/
NÖ	http://www.holding.lknoe.at/kliniken.html

Physician Directory

German Diabetic Society	http://www.deutsche-diabetes-gesellschaft.de/redaktion/organisation/diabetologen/suchform.php
List of hospitals/Institutions in Germany Specialised Diabetes	http://www.deutsche-diabetes-gesellschaft.de/redaktion/einrichtungen/

[Anerkannte Einrichtungen DDG.php](#)

Job Opportunities

http://www.univadis.at/medical_and_more/News#

<http://universimed.com/latest-news>