Observations and Lessons Learnt from Non Health Professionals Evaluating a Health Search Engine

Natalia PLETNEVA[[1]](#footnote-1)a,, Zdenka URESOVAb, Jean-Jacques ALTMAN c, Nicolas POSTEL VINAY c, Patrice DEGOULETc, Jan HAJICb, Celia BOYER a

a Health on the net Foundation, Chene-Bourg, Switzerland

b Charles University in Prague, Czech Republic

c Georges Pompidou European Hospital, Paris, France

**Abstract.** This article presents the results of one of the stages of the user-centered evaluation conducted in a framework of the EU project Khresmoi. In a controlled environment, users were asked to perform health-related tasks using a search engine specifically developed for trustworthy online health information. Twenty seven participants from largely the Czech Republic and France took part in the evaluation. All reported overall a positive experience, while some features caused some criticism. Learning points are summed up regarding running such types of evaluations with the general public and specifically with patients.

**Keywords.** Internet, Search engine, User evaluations, Patients

Introduction

In the era of Dr Google, EU project Khresmoi (2010-2014, project No. 257528 (FP7/2007-2013)) aims at developing trustworthy search engines targeting three different audiences: the general public, physicians and radiologists. Khresmoi for everyone (hereinafter K4E, <http://everyone.khresmoi.eu>) was specifically developed for the first (and the widest) group users. This group is also known to be heterogeneous in terms of age, gender, educational and income level [1] and consequently of IT and health literacy [2,3] what is eventually linked to health status [4]. This type of users is often not aware about variability of the quality of online health information and lacks skills to judge it [5,6]. K4E aims at targeting this problem and offering an access to reliable online health resources based on 8.300 HONcode certified websites and some others manually selected by experts..

By the third year of the project the first integrated platform featuring various functionalities was developed [7] and tested in real-life settings. This article describes the results of one of the types of evaluation tests conducted in 2013 [8]. Health on the net Foundation (HON) was in charge of defining the protocols, coordinating and conducting the evaluations in collaboration with Charles University in Prague (CUNI, Czech Republic) and Evaluation and Language Resources Distribution Agency SA (ELDA, France) in different locations. The main goal of this type of evaluation was to understand users’ acceptance and use of the prototype, specifically which tools were considered to be the most useful. Consequently, these findings were analysed and integrated in the next prototype. Although we overall followed a standard usability testing protocol, some peculiarities of health domain had to be taken into consideration while designing the study.

# Methods

A qualitative research method was chosen in order to have in-depth understanding of the users actions and behaviour and the reasons inducing these actions. The method consisted, during an average of one hour, in observing and recording users’ interaction with the prototype K4E using the Morae software by TechSmith [9] while completing three health-related online searches. The Morae software allowed capturing the screen of the computer, video (participant face), and click-through data. Set up usually included one participant, one observer - who was adding notes during the test - and one facilitator, whose aim was to assist the participant on site in case of questions or problems and to collect feedback at the end of each interview.

Several tasks covering various health topics were elaborated in English and piloted with eleven participants in two rounds in Sofia (Bulgaria) and Geneva (Switzerland). The final setup includes same three simple tasks and several questionnaires.

The main inclusion criteria were not being a health/medical professional and at least occasionally searching for online health information. Upon agreeing to participate, all participants filled out a consent form. One page presentation of the project and a quick demonstration of the prototype were shown to the participants. Once started, participants filled in a demographic and Internet usage questionnaire. Then they had about 10 minutes to explore the search engine by themselves. Afterwards, the three tasks were given. The first task was to find a formula and calculate a Body Mass Index (BMI). The second task aimed at finding the information about treatment options for liver cancer. The third task required participants to find other patients’ opinions on a drug called “Metformin”. After each task, the participants had to answer two questions: whether they already had a personal experience with a problem described in the task and whether they were able to solve it (detailed in tab 1). At the end of the evaluation, an extended System Usability Scale (SUS) questionnaire was filled out. In addition to the 10 standard questions [10], we developed 15 specific to K4E questions. Participants had to grade each statement using the Likert scale from one to five, from strongly disagree to strongly agree, accordingly.

The “full user tests” were conducted during a few days from May to July of 2013 in Prague, Geneva and Paris. All the materials, originally available in English, were translated into Czech and French. However, the interface of the Morae system remained in English.

# Results

Thirty people took part in the evaluation tests and for the final analysis 27 recordings were kept. One of the 27 participants encountered a technical problem with Morae, and his answers to the demographic questionnaire were not recorded, however some, though not all of them were recovered by watching the video, hence in some cases a sample size is 26 (N=26).

Participants were mostly from the Czech Republic (14) and France (9), while 3 were from Switzerland (N=26). There were 14 females and 13 males of various age groups: the most prevalent were aged 30-39 (37%) and 60-69 (22%). The majority of participants hold a Master degree (41%), while all education levels were present. All participants were active online, 25 indicated they were connected to the Internet on a daily basis, and two did so several times a week. All participants also conducted online searches: 23 every day, and 4 several times a week. Everyone indicated Google as the main search engine being used (27), two Czech participants also mentioned Seznam.cz, one added Yahoo and another one added DuckDuckGo. Almost half of them (48%) said they were very confident with web search and considered themselves as expert users (N=27). 44% had problems in finding information from time to time.

As for online health search, 42% of the participants did it once a month and 16% at least once a week (N=26). All participants gave preference also to Google for looking for health content (N=27). All participants sought for information about a specific disease or medical problem. Out of 26 respondents, 9 (35%) reported having been diagnosed with various diseases, mostly (7 out of 9) for more than three years. 6 out of 9 participants reported that their Internet use to search for online health information did not change after the diagnosis, while for 2 participants it increased and for 1 it decreased.

All of the participants (after free search engine use) have completed three tasks, described in Methods section, and responded to two questions per task (Tab. 1).

**Table 1.** Summary of experience with three tasks made by all participants (N=27).

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Previous experience with the issue** | **Task success rate** | **Average time to complete the task** |
| **A question asked:** | **Yes** |
| 1 | Have you already been familiar with body mass index before doing the task? | 70% | 100% | 2.50 min |
| 2 | Have you (or someone you know) been diagnosed with liver cancer? | 8% | 81% | 5.09 min |
| 3 | Have you previously searched for medication information on the Internet? | 74% | 48% | 5.22 min |

We wanted to understand whether the use of K4E helped with solving the tasks. In case of the first task, it was not clear as all of the participants had solved it. However, K4E did help for the second task: 80% of the participants who had no prior experience and all participants who did have prior experience with liver cancer were able to solve the task. The remaining 20% of those who did not have prior experience with liver cancer were not able to find a satisfactory answer. As for the third task, taking into account the data we had, K4E did not help. There was no significant difference in task success rate between Czech and Francophone participants.

For the analysis of the final satisfaction questionnaire we split all statements into “happy” (17) and “not happy” (8). “Happy” statements reflected positive user feedback; for such statements, the higher the grade was the more satisfied the users were with the prototype and their experience. Overall, participants agreed with the happy statements (between three and four on a five-point Likert scale). Several statements received more than four points out of five (i.e. “strongly agree”) - the following system characteristics were most appreciated: (1) results classification (4.15), (2) quick learning of how to use the system (4.07), (3) feeling of confidence (4.04), (4) ease of use (4.04). On the contrary, “not happy” statements reflected disappointment and higher grade for them meant higher dissatisfaction. All of “not happy” statements received less than 3 out of 5 which indicates that on average participants disagreed with all of them (statements like: “I thought there was too much inconsistency in this system”, “I found the system unnecessarily complex” etc).

Francophone participants reported having more positive experience with the tools which would help them to understand the topic, for example, disease definition, i.e. tools targeting a health domain, while Czech users tended to appreciate more “common” search engine tools such as automatic translation and filtering.

Other feedback collected from the observation and free comments of participants can be summed up as follows. Overall, all participants were able to effectively interact with the prototype. The interface was found to be simple enough to use, but the translation, health topic cloud categorization and several other features were mentioned as needing improvement. The availability of the contents was judged as poor by most users, especially in the Czech language. Indeed, not many websites were HONcode certified in Czech and only few websites were added during the first two years of the project. While Czech participants had often turned to translated results, French-speaking participants could complete the tasks without automatic translation of results into English, though few of them used automatic translation from English. The quality of the snippet translation was deemed sufficient but requiring some improvement. A particular problem French-speaking participants encountered was with accents while typing a query. Having forums appeared to be a contradictory issue as opposite opinions exist on the matter: while some of the users considered forums and blogs a valuable source of information, others did not and would prefer to filter them.

# Discussion

In our tests, we had Internet users of both genders, all ages, with different web search experience and skills and personal health experience. A high percentage of participants held at least a Master degree, it is nevertheless a common tendency for online health seekers to have higher education compared to the average web surfer.

The main conclusion is that the perception of the users with different profiles was overall positive and the system and its tools were helpful to solve the proposed tasks while other tools have been found in the need of further improvement.

    Having analyzed users’ feedback and experience we proposed the following steps to improve the prototype:

* Simplify and localize interface across the languages (including titles, filters presentation and definition of search terms),
* Improve query assistance and automatic translation,
* Expand the index, especially with Czech resources,
* Increase system speed.

Regarding the setup of the evaluation itself, users wished to have more time to explore the system on their own.

Main learning points are the following. As the main goal was to provide a roadmap for further prototype development, such evaluations are deemed to be very appropriate to truly understand user behaviour and retrieve unique and profound feedback about the system. Overall, for such kinds of evaluation, a pilot test is never superfluous, and it indeed seems appropriate to organize both pilots and tests in many small-sample steps - for example, 2-3 persons per each step. After each step, both the setup and the prototype should evolve to allow new participants to contribute with a new input on the next step in a two-to-three month period. The last, but not the least point is directed towards patient organizations and facilities providing healthcare, to encourage them to promote such studies amongst themselves and their patients to ensure the final product meets their needs.

Follow up evaluations are being conducted in 2014 to evaluate whether the progress made in the meantime facilitates users interactions with the prototype.

HON expresses its gratitude to our partners who helped to run this evaluation: Priscille Schneller and Jérémy Leixa from ELDA and Dr. Line Kleinebreil who has contributed to the protocol development. HON thanks its team members Rafael Ruiz de Castaneda, Frederic Baroz and Samia Chahlal. We thank all the volunteers for their time and participation.

References

1. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet Use and eHealth Literacy of Low-Income Parents Whose Children Have Special Health Care Needs. J Med Internet Res, 2011, p. 13(3):e75.
2. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, Schillinger D. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun, 2010, pp. 15 Suppl 2:183-96.
3. Lam K.M., Lam L.T. Health Information-seeking Behaviour on the Internet and Health Literacy among Older Australians. Electornic Journal on Health Informatics, 2012, pp. Vol 7, No 2.
4. Fox S., Duggan M. Health Online 2013. Washington, D.C. : Pew Research Center’s Internet & American Life Project, 2013.
5. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth Literacy With Colorectal Cancer Knowledge and Screening Practice Among Internet Users in Japan. J Med Internet Res, 2012, p. 2012;14(6):e153.
6. Feufel MA, Stahl SF. What do Web-Use Skill Differences Imply for Online Health Information Searches? J Med Internet Res , 2012, p. 14(3):e87.
7. Pletneva N, Kriewel S, Kritz M. Prototype of a seconf search system based on feedback. 2013.
8. Baroz F., Boyer C., Gschwandtner M., Goeuriot L., Hajic J., Hanbury A., Kritz M., Leixa J., Palotti J., Pletneva N., Ruiz de Castañeda R., Sachs A., Samwald M., Schneller P., Stefanov V., Uresova Z. D10.1 Report on user tests with initial search system. 2013
9. Brooke, J. SUS: a "quick and dirty" usability scale. B. Thomas, B. A. Weerdmeester, & A. L. McClelland P. W. Jordan. Usability Evaluation in Industry. London : Taylor and Francis, 1996.
10. Krug S. Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability (3rd Edition) (Voices That Matter) . s.l. : New Riders, 2013.
1. Nataly.Pletneva@HealthOnNet.org [↑](#footnote-ref-1)