Considering subjects and scenarios in large-scale user-centered evaluation of a multilingual multimodal medical search system

Liadh Kelly[1,*], Lorraine Goeuriot[1], *Gareth J.F. Jones*[1], *Allan Hanbury*[2]
1. Centre for Next Generation Localisation, School of Computing, Dublin City University, Dublin 9, Ireland
2. Department of Software Technology and Interactive Systems, Vienna University of Technology, 1040 Vienna, Austria
** Address for Correspondence: lkelly@computing.dcu.ie*

Introduction

Medical search applications can be required to service the differing information needs of multiple classes of users with varying medical knowledge levels, and language skills, as well as varying querying behaviours. The precise nature of these users' needs has to be understood to develop effective applications. Evaluation of developed search applications requires creation of holistic user-centred evaluation approaches which allow for comprehensive evaluation while being mindful of the diversity of users.

This paper describes plans for evaluation of the effectiveness of the Khresmoi system, a large scale multilingual eHealth system being developed across 12 institutions in the EU (http://khresmoi.eu/). Khresmoi aims to provide a multilingual and multimodal search and access system for biomedical information and documents, which targets three classes of users. Two of these are groups with general medical interests: the general public and general medical practitioners. The other group is an example clinician group with a specific expertise: radiologists. The system seeks to provide these groups with innovative and effective services for searching through the very large amount of medical information available. Relevant information may often not be available in the searcher's native language, as scientific content is assumed to be more often available in English. Thus one of the major features of our system is translation support to provide cross-lingual access to the English language medical information. A key feature here is to provide support appropriate to the varying language skills of the users. The system gathers several innovative technologies linked to medical information (e.g. text and image information extraction and retrieval, machine translation) to provide a comprehensive tool adapted to all users.

The overall aim of our evaluations is to enable an assessment of the efficiency and effectiveness of the developed technologies and of the quality of the support provided to the target users. The results of these evaluations will guide further research and development in the Khresmoi project. These evaluation plans consist of both empirical and user-centred evaluations. In this paper we focus on elements of the design of the user-centred evaluations.

Materials and Methods.

Users and their requirements are key to creating useful medical information retrieval systems. While user behaviour and satisfaction has been studied for general web search [1] and health-related search [2], further investigations have to be carried out to define a proper evaluation strategy within multifunctional medical search systems such as Khresmoi. Surveys and interviews with representative potential users have been conducted and guided us in understanding the requirements of the end users of the Khresmoi system. The surveys were conducted with 385 members of the general public [3], 556 physicians [4] and 34 radiologists [5]. Questions asked included their use of search systems to get medical information, as well as features they would like in a new medical information system. Based on the results of these requirements, two prototype systems have been developed: one for the general public and general practitioners; and the other for radiologists. Detailed evaluation plans are being developed for these prototypes. The first steps in this strategy relate to the experimental subjects and evaluation scenarios. In this section we highlight the considerations of these elements in our large scale user-centered evaluation strategy.

Firstly, we define demographic features of each class of user in order to get representative groups. This will allow us to recruit suitable subjects, as well as to explore any relationships between the

demographics and users satisfaction regarding the system features. In order to conduct thorough user-centered evaluation of our multilingual medical system servicing the different categories of users, each of these user categories should be evenly represented in the evaluations (general public, general practitioners and radiologists). Gaining subjects of this nature can be problematic, especially when professionals are required. Members of the general public will be recruited through patients' organisations, general practitioners through a Society of Physicians, and radiologists through two hospitals. Using real subjects of this nature raises issues such as whether subjects should be paid; the extent to which evaluations can be conducted online; and practical issues related to the geographic distance between subjects - practicality and cost associated with subjects travelling to the investigator, or for the investigator to travel to many distant locations. Since our system is intended for use by individuals with different native languages (English, German, French, Spanish and Czech as test cases) and different skill levels in the English language (ranging from none to fluent) for which translation support is provided, the subjects used in the experiments should have varying levels of English. This will enable us to evaluate the utility of the translation support functionality for individuals with different English levels. Gaining a good spread of subjects with different language skills, or indeed determining the language skills of the subjects one gains, is nontrivial. Lessons on how to approach this topic could be taken from Marlow et al's explorations into the effects of language skills on multilingual web search [6]. Similar to the language skills, other varying characteristics of the users may have a strong impact on their use of the system and their satisfaction, e.g. medical knowledge level, computer skills. To determine the demographic spread of subjects, standard questionnaires should be completed by subjects to determine their age, gender, medical condition, prior use of medical search engines, computer skill, etc.

Our second consideration for a large scale user-centered evaluation strategy is development of a robust methodology for scenario creation. The user studies mentioned earlier in this section gave us an insight into the types of search tasks that users are performing and would perform in the medical space. A method to model this task information will be defined for use in development of experimental search scenarios. This scenario creation process is complicated by the diversity of users. In creating scenarios for the general public for example, we need to think in standardised form about the different characteristics of users within this population, e.g. different English language skills. In doing this, we also need to be mindful that scenarios should include situations examining both monolingual and multilingual search.

Conclusion

In this paper we described the importance of subject and scenario considerations for user-centered evaluation strategy development of the Khresmoi multilingual multimodal biomedical information system, being developed for three target user groups: general public; general practitioners; and radiologists.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°257528 (KHRESMOI).

References

1. D. Rose and D. Levinson (2004). *Understanding User Goals in Web Search.* In *WWW*, p.13-19.
2. R. Cline and K. Haynes (2001). *Consumer health information seeking on the Internet: the state of the art.* In *Health Education Research* 16(6).
3. N. Pletneva and A. Vargas (2011). *D8.1.1. Requirements for the general public health search. Technical report, Khresmoi Project.*
4. M. Gschwandtner, M. Kritz, and C. Boyer (2011). *D8.1.2: Requirements of the health professional search. Technical report, Khresmoi Project.*
5. H. Müller (2011). D9.*1: Report on image use behaviour and requirements. Technical report, Khresmoi Project.*
6. J. Marlow, P. Clough, J. C. Recuero, and J. Artiles (2008). *Exploring the Effects of Language Skills on Multilingual Web Search.* In *ECIR 2008.*