Accepted for publication in 15th International Conference on Human-Computer Interaction, Las Vegas (July 2013)

Author's Copy

Search Engine Accessibility for Low-Literate Users

Débora Maurmo Modesto, Simone Bacellar Leal Ferreira and Aline Silva Alves

Departamento de Informática Aplicada, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
{debora.modesto, simone, aline.alves}@uniriotec.br

Abstract. Search engines are often used to retrieve content on the Web, but it is not a simple activity for low-literate users since they have to know the technology and create strategies to query and navigate. Their interaction with search engines differ from high-literate users on strategies used, perception, communication and performance. In order to improve search engines and create solutions, we need to understand these users' needs. This research aimed to identify how search engine features influence the interaction of low-literate users. We analyzed the interaction of ten users through user tests that were part of a case study. Based on a limited set of features of a specific search engine, we identified what features were used, the perception about them and some barriers faced by these users. This study led to a list of recommendations for the development of search interfaces focused on low-literate users.

Keywords: Low-Literate Users, Search Engine Accessibility, Guidelines.

1 Introduction

The available tools on the Web are part of everyday life [1] and perform a social role especially for users with disabilities [2]. Therefore, there should be no barriers to access the Internet. There are users who have limitations related to literacy that can jeopardize the interaction mainly because the available content on the Web is mostly textual [3].

To help users to retrieve this content, search engines are often used: 80% of the access to Web pages comes from these tools [4]. Search involves analyzing different types of media, so it is a mentally exhausting activity that requires focus and attention [5]. Low-literate users have some limitations, particularly related to the strategies they use to conduct a search and the perception of interfaces [6-7]. Besides that,

high-literate users interact with such tools in a different manner than low-literate users [3], [8].

In order to improve low-literate users experience on the Web, search engines should accomplish their mental models. Designers and developers should know how they interact with this kind of tools, what features are used and how they influence on user experience.

This research aimed to identify how search engine features influence the interaction of low-literate users. We observed how the interaction takes place and identified some barriers faced by these users. In order to do this, a case study was conducted leading us to a list of recommendations for the development of search interfaces focused on low-literate users.

Besides this introduction, this article is structured as follows: section 2 presents literacy concepts; section 3 presents related work; section 4 presents methodology; section 5 explains how the study was conducted; finally, section 6 presents conclusions.

2 Literacy in Brazil

According to United Nations Educational, Scientific and Cultural Organization (UNESCO), literacy can be analyzed by different perspectives and conceptualized in different ways. A common understanding about literacy involves oral, reading and writing skills, besides abilities with logic, mathematics, symbolic analysis (images and sounds) and text interpretation [9]. Nowadays it is a common approach to consider that these skills must be contextualized and they are not developed equally among different individuals. Besides, literacy concept also considers functional aspect that means the ability to apply oral, reading and writing on different areas of daily life, as in computing, ecology, health and other areas [9].

The concept of functional illiteracy varies from region to region. In Brazil, federal agencies as Brazilian Institute of Geography and Statistics (IBGE) adopt the same criteria as UNESCO to define function illiterates: people between 15 and 64 years old, which lack mastery of skills in reading, writing, calculations and science, corresponding to an education of less than four years of study [9-10]. In Brazil, it corresponds to incomplete 4th grade of elementary education. A research

performed by IBGE showed that 29 million people in Brazil are functional illiterate that means more than 20% of the total population [10].

In this work, we adopt this concept but we call the participants as "*low literate users*", once one of the protocols used to guide the study recommended not to call them functional illiterates [8].

There are other criteria that could be used on the study instead of considering only age and years of study on formal education. There are institutions that developed specific literacy and numeracy tests, with levels of difficulty and punctuation, to evaluate people's skills in reading, writing, calculations and science. However, these tests are private, so there are no public tools that allow us to classify users this way [8].

3 Related Work

3.1 Taxonomy of Web Search

Web search can be categorized in three groups: (a) informational, when users want to find more information about a topic; (b) transactional, when users want to perform operations after searching, as downloading a file; (c) navigational, when users want to find a site from some organization they already know [4], [11-12].

3.2 General Users' Behavior

Some studies address general users' behavior when using search engines: On informational or transactional queries, users usually focus on the title and description of a search result. Images are also an expected type of media, but videos can distract users once it is not possible to comprehend its full meaning quickly [4]. On navigational searches, users tend to ignore results from the fourth position on and they focus also on the URL [4].

If users do not find what they want on first results, they tend to perform another search [13]. Users are usually influenced by suggested results' relevance, identified through positioning [14-15].

Another study identified that experience influence users' behavior. Novice users adopt search strategies less flexible than expert ones, have difficulty to formulate a query and do not know how search engines

work [16]. Another study stated that only a few users know how to use advanced features [17].

A study about terms suggestion showed that most users preferred to refine the query manually [18]. On other study some people stated that this feature was helpful but it was also a distraction. However, after a week of log analysis, users started to use it in an iterative way [19].

Users usually do not know how the search engine works. It creates incorrect expectations about outcomes. Explanations presented about how search engine works helps users to understand outcomes [20].

Some studies showed that grouping results by categories is a good way to present them [21-22]. Present a good description of the outcomes is also helpful. Some studies showed that summarization of the page and highlighting terms were good to improve result analysis [23-24].

3.3 Low-Literate Users' Behavior

Low literate users usually do not check correctness of information [6]. These users usually cannot scan results as high literate users do [4], so they need to read every word to understand information [6]. They tend to have more difficulty recovering from errors or changing search strategy. They also become confused when navigating in pages full of information and links. Their performance is usually worst when compared to high literate users, since they take longer to finish tasks, spend more time on the same page and visit more pages [6].

A study stated that they need features that support decision about results such as in which link they should click or if information is relevant or not. The same study mentions that this kind of feature is more important than the ones to support them to define query [25].

No study was found about how search features influence low-literate users. This paper shows how search features affect low-literate users' experience, mainly Brazilians whose first language is Portuguese.

4 Methodology

This was a qualitative and exploratory research. The research took place in Rio de Janeiro, Brazil, and consisted of a case study to analyze the interaction of ten low-literate users with Google search engine

through user tests. The participants were between 15 and 64 years old and had less than four years of study on formal education [9]. They were all Brazilians from different regions of the country.

This approach can be considered limited because it does not consider extracurricular capabilities and years of study are related to a feasible educational goal by Brazilian government [4]. If other criteria were used to select participants, maybe other results could be found.

Tests were operationalized through "Protocol for conducting usability testing with a focus on accessibility" [26], that defined steps to accomplish planning, preparing, conducting and reporting results, and "Protocols for Web accessibility evaluation involving functional illiterates" [8], that details the approach and execution of the user tests.

People were recruited in schools with youth and adults education classes, churches and residential buildings. As recommended by the protocol [8], a portable usability lab was set and used on the tests that were performed in various locations to facilitate transportation for participants.

A questionnaire was applied in order to find more information about users' profile. Based on this, Google was chosen for observation once it was the main search engine used for all participants. Besides that, search volume on this search engine corresponds about to 66% of all searches on the Web [27-29]. In 2011, Google Brazil (www.google.com.br) reached more than 92% of searches performed by users in Brazil [30].

The search engine was explored on its default state and all features were available with no customization. A limited set of features related to activities such as writing, reading and formulating search, guidance, navigation and feedback were selected for analysis. This set consisted of: auto complete, spell checker, related searches, advanced search, filters, layout (header, search bar, advertisements, pagination and footer), search results, keyboard navigation, "I'm Feeling Lucky" button, and Google Instant features (as page and result preview). Some aspects such as simplicity of the text, amount of terms used, how they elaborate a query, perception and orientation were also analyzed.

The case study consisted of two units of analysis. On the first unit five users interacted with Google search engine to perform five tasks that varied by difficulty level. The first three tasks were considered easy and it was expected that all participants concluded all of them to

finish the test. The fourth task was considered a medium task since it involved notions of magnitude, as maximum and minimum and advanced vocabulary. The fifth one was considered a difficult task since it involved notions of history, current events and interpretation skills. Tasks also varied by search goals, since three of them were informational tasks, one was navigational and one transactional. Test could finish after five tasks concluded or thirty minutes, what have happened earlier. After the test, each user was interviewed and answered questions about perception of the features´ utility.

On the second unit, five other users interacted with two resources that were not used on the first unit of analysis by anyone, but were considered useful: filters and advanced search. Before performing each task of this unit, a video was shown teaching how to use each feature.

Textual material used in the study, including the questionnaire, tasks and informed consent, were prepared with the assistance of a checklist for plain writing for Web [31]. Data was analyzed following Four-Phase Framework for Search [32]. This framework states that every search consists of four main phases that are formulation, action, review of results and refinement. Features were grouped on these categories and analyzed according to each phase´s goals.

5 Case Study

Three men and two women participated of each unit of analysis. Data collected on questionnaires indicated that eight users were less than two years of experience with internet and two were less than five years. Despite the wide range of age adopted as criteria, we selected users who had similar education and experience with computers in order to minimize a possible bias caused by age difference. Five participants said they usually ask for help when using search engines. Eight participants indicated that the main difficulty is to know whether a word is spelled correctly. The result analysis was also considered a difficult task by six participants. On the other hand, seven participants stated they were comfortable when they need to formulate a query. Regarding to search topics, products were the most mentioned (cited by eight participants), followed by music and videos (both mentioned by seven participants).

5.1 Recommendations for Search Engine Development Focusing on Low-Literate Users

Below are presented the recommendations developed based on observation of low-literate participants during user tests:

1. Provide features that help writing and problem formulation (like "spellchecker", "autocomplete" and "related searches"). Show changes made by the "spellchecker" as soon as the results are presented, positioned above the results, so they are quickly visualized. Present results corrected by "spellchecker" and provide feedback, indicating that terms were changed. Show terms suggestion to complete the query while user types next to the search field to enable quick viewing. Show the feature "related searches" below the results for easier query refinement.
2. Provide "filters" not only for refinement, but also for formulating the query. In both cases, filters should be formatted as categories or as a menu. Place "filters" in areas with less emphasis, such as the header of the page. Present results related to "filters" or categories in a different format from conventional results.
3. Provide a large text box to write the query. The terms that the user typed should not be hidden so he does not forget the words he used and do not get confused about the research problem.
4. Low-literate users cannot handle too much information at the same time and get confused with lots of text, so display around seven main results at a time (at least five and at most nine). In general, people feel more comfortable to handle this amount of results [33]. Secondary outcomes that lead to internal pages of a website can be displayed once they not hinder the understanding of users, but should not be excessive.
5. Present the results divided into pages and display a paging feature to navigate between them, positioned at the end of the results page. Also use the expression "see more" beyond the page numbers, because this term is more familiar to users.
6. Provide a visual indication of results' relevance, which is not only showed by the page rank and positioning on the page. Positioning has not a clear meaning for these users.

7. Do not present other media formats or filtered results among the conventional results because it confuses users. Allocate a page area to present this kind of outcome.
8. Display the title and description for a result emphasizing the first one. Show other information as the URL on demand, only if user requests. Generally, low-literate users do not visualize this information and do not use it to decide whether to click on a result. Search terms should be highlighted on the description, in order to keep user focused on the subject of task. Distinguish snippets extracted from different parts of a website through background colors, for example. Avoid using suspension points for that.
9. Features that provide instant feedbacks are recommended, however, should be prominently displayed so they can be readily seen.
10. Provide features to help users to decide whether or not to select a result. These inputs should be showed only on demand. For example, "page preview" feature provides inputs to the user to decide whether to select a result, but it's not used by low-literate users once they cannot comprehend what is in the page without reading carefully its content.
11. Advanced features such as "keyboard navigation", can be available if the interface is also used by advanced users. However, low-literate users do not make use of these resources.
12. The footer area of the results page is less visualized. Provide information that does not need emphasis in this area.
13. Use tips and directions about the use of the interface so that it does not distract users nor overload the page with lots of information.

6 Conclusion

During this research, a set of recommendations was developed, considering low-literate users' preferences, such as writing instead of reading and features that don´t impact performance. They also addressed issues related to user's needs such as: (1) features to assist them in writing and formulating search problem, (2) minimize distractions, (3) provide clear and visible feedback, (4) provide instructions about search engine behavior, (5) facilitate recovery from errors, (6) provide ways to

stay focused on one activity at a time, (7) decrease the amount of text and results, (8) organize search results by categories.

These recommendations still need a validation since it was not addressed on this study. Despite the similarity with other search engines, generalization should be carried carefully once no tests on other tools were performed to validate the guidelines at this time. These recommendations can assist developers in creating interfaces for search engines or search features within websites. It is expected that interaction of low-literate users is enhanced on this kind of tools and they find information more easily.

References

1. Nielsen, J., Loranger, H.: Usabilidade na Web: Projetando Websites com Qualidade. Elsevier, Rio de Janeiro (2007)
2. Ferreira, S.B.L., Nunes, R.R.: e-Usabilidade. LTC, Rio de Janeiro (2008)
3. Kodagoda, N., Wong, B.: Effects of Low & High Literacy on User Performance in Information Search and Retrieval. In: Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, pp.173-181. The British Computer Society, Swinton, UK (2008)
4. Thurow, S., Musica, N.: When Search Meets Web Usability. New Riders, Berkeley (2009)
5. Hearst, M.A.: Search User Interfaces. Cambridge University Press, New York (2009)
6. Kodagoda, N., Kahan, N., Wong, W.: Identifying Information Seeking Behaviors of Low and High Literacy Users: Combined Cognitive Task Analysis. In: Proceedings of NDM9, the 9th International Conference on Naturalistic Decision Making, pp. 347-354. The British Computer Society, London (2009)
7. Gupta, N.K., Rosé, C.P.: Understanding Instructional Support Needs of Emerging Internet Users for Web-based Information Seeking. Journal of Educational Data Mining 2, pp. 38-82 (2010)
8. Capra, E.P.: Protocolos para Avaliação da Acessibilidade Web com a Participação de Analfabetos Funcionais. Master Dissertation. Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro (2011)
9. United Nations Educational, Scientific and Cultural Organization.: Understandings of Literacy. In: The Education For All Global Monitoring Report: Literacy for Life. pp. 149-159 (2006). http://www.unesco.org/education/GMR2006/full/chapt6_eng.pdf
10. Instituto Brasileiro de Geografia e Estatística: Síntese de Indicadores Sociais: Uma Análise das Condições de Vida da População Brasileira. (2010). http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/sinteseindicsociais2010/SIS_2010.pdf
11. Broder, A.A.: Taxonomy of Web Search. ACM SIGIR Forum 36(2), 3-10 (2002)
12. Manning, C., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Stanford (2008)
13. Spink, A., Jansen, B.J.: A Study of Web Search Trends. Webology 1, (2004). http://webology.ir/2004/v1n2/a4.html

14. Keane, M.T., O'Brien, M., Smyth, B.: Are People Biased in Their Use of Search Engines?. Communications of the ACM 51, pp. 49-52 (2008)
15. Bar-Ilan, J., Keenoy, K., Levene, M., et al.: Presentation Bias is Significant in Determining User Preference for Search Results: A User Study. Journal of the American Society for Information Science and Technology 60, 135-149 (2009)
16. Hölscher, C., Gerhard, S.: Web Search Behavior of Internet Experts and Newbies. Computer Networks 33(1–6), 337-346 (2000)
17. Machill, M., Neuberger, C., Schweiger, W., Wirth, W.: Navigating the Internet: A Study of German Language Search Engines. European Journal of Communication 19, 321-347 (2004)
18. Anick, P.: Using Terminological Feedback for Web Search Refinement: A Log-based Study. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (SIGIR '03). Toronto, Canada, pp. 88-95 (2003)
19. Anick, P., Kantamneni, R. G.: A Longitudinal Study of Real-time Search Assistance Adoption. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08). Singapore, pp. 701-702 (2008)
20. Muramatsu, J., Pratt, W.: Transparent Queries: Investigating Users' Mental Models of Search Engines. In: Proceedings of the Twenty-fourth International ACM Conference on Research and Development in Information Retrieval. New Orleans, LA, (2001)
21. Käki, M.: Enhancing Web Search Result Access with Automatic Categorization. Ph.D. dissertation, University Of Tampere, Finland (2005)
22. Käki, M.: Findex: Search Result Categories Help Users When Document Ranking Fails. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05), pp. 131-140. Portland, Oregon, USA (2005)
23. Tombros, A., Sanderson, M.: Advantages of Query Biased Summaries in Information Retrieval. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98), pp. 2-10, New York (1998)
24. White, R.W., Jose, J., Ruthven, I.: A Task-oriented Study on the Influencing Effects of Query-biased Summarization in Web Searching. Information Processing and Management 39(5), 707-733 (2003).
25. Gupta, N.K., Rosé, C.P.: A Foray into Understanding the Next Billion Search Users. 2010b. http://www.cs.cmu.edu/~nkgupta/papers/chi2010a_submitted.pdf
26. Henry, S.: Just Ask: Integrating Accessibility Throughout Design. 2007. http://www.uiaccess.com/accessucd/
27. Experian. Search Engine Analysis. http://www.hitwise.com/us/datacenter/main/dashboard-23984.html
28. Comscore. comScore Releases January 2012 U.S. Search Engine Rankings. http://www.comscore.com/Press_Events/Press_Releases/2012/2/comScore_Releases_January_2012_U.S._Search_Engine_Rankings
29. Netmarketshare. Desktop Search Engine Market Share. http://netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0
30. Serasa Experian. Share Mensal de Buscas do Google Atinge 92,15% em Janeiro. http://www.serasaexperian.com.br/release/noticias/2011/noticia_00381.htm

Accepted for publication in 15th International Conference on Human-Computer Interaction, Las Vegas (July 2013)

Author's Copy

31. Barboza, E., Nunes, E.: A Inteligibilidade dos Websites Governamentais Brasileiros e o Acesso para Usuários com Baixo Nível de Escolaridade. Inclusão Social 2(2), 19-33 (2007)
32. Shneiderman, B., Byrd, D., Croft, W.B.: Clarifying Search: a User-Interface Framework for Text Searches. D-Lib Magazine, 1-15 (1997) http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html#formulation
33. Miller, G. A.: The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review 101(2), 343-352 (1955).