Ordering Matters: An Experimental Study of Ranking Influence on Results Selection Behavior during Exploratory Search

Emanuel Felipe Duarte[1], Lucas Pupulin Nanni[1], Ricardo Theis Geraldi[1], Edson OliveiraJr[1], Valéria Delisandra Feltrim[1] and Roberto Pereira[2]

[1]*Informatics Department, State University of Maringá, Maringá-PR, Brazil*

[2]*Institute of Computing, State University of Campinas, Campinas-SP, Brazil*

Keywords: Search User Interfaces, Exploratory Search, Ranking Influence, Relevance Judgement.

Abstract: The design of Exploratory Search tools acquires more importance as the amount of information on the Web grows. Accordingly, informed design decisions concerning the users' behavior during search activities can be used in novel approaches for Exploratory Search tools. With regard to users' behavior on search result selection, literature indicates higher ranked results tend to attract more attention and, therefore, more hits. However, the cause of such behavior is not clear. We experimentally investigate the hypothesis of this behavior being due to ranking. A group of 72 participants was asked to select a search result from a randomly ordered results list. The experiment was carried out on paper to remove specific search engine and digital medium biases. We obtained evidence indicating there is a trend towards choosing higher ranked results even in a different medium and in the context of an Exploratory Search, corroborating to the hypothesis that ranking has influence on users' selection behavior.

1 INTRODUCTION

According to White and Roth (2009), exploration is part of the human nature and, as explorers, we aim to constantly expand our knowledge through exploration. Blandford and Attfield (2010) discuss that continuous advances in information technology over the past decades and, in particular, the advent of the Web have revolutionized the way people interact with information. Consequently, the amount of available information to be explored has become abundant.

However, Blandford and Attfield (2010) discuss how the abundance of information, most of it located on the Web, became a significant problem in recent years. According to White and Roth (2009), there is a growing need for Interactive Information Retrieval systems capable of safeguarding the user's attention through information filtering.

White et al. (2005) argued search technologies available at that time already provided adequate support for users with well-defined information needs, but lacked support for situations where users do not have the knowledge or contextual awareness to formulate queries or navigate complex information spaces (Bates, 1989). According to White et al. (2013), current search technologies still provide insufficient support for this kind of activity, denominated **Exploratory Search**: an activity where the user performs a search with open and abstract goals and need to build knowledge about a particular subject. Saving users' attention is critical in Exploratory Search activities, therefore, there is a growing need for models and tools that support users in Exploratory Search activities.

Current major Web search engines, such as *Google*, *Bing*, *Yahoo!* and *Baidu*, adopt the listing format of ranked search results in their user interface. In this schema, studies reveal the results that appear at first positions of the list tend to receive most of the attention and, therefore, more hits than the results presented at last positions (Granka et al., 2004; Joachims et al., 2005; Pan et al., 2007).

However, it is not clear whether the position itself is enough to influence users to select the first results. The interaction possibilities of the digital medium (*e.g.*, scrolling, clicking, opening in a new tab) along with the additional tricks used by traditional search engines (*e.g.*, graphical signs, additional result information, highlights for sponsored links) might also influence results selection.

Understanding the user's behavior towards search activities may provide basis for informed design deci-

Duarte, E., Nanni, L., Geraldi, R., OliveiraJr, E., Feltrim, V. and Pereira, R.
Ordering Matters: An Experimental Study of Ranking Influence on Results Selection Behavior during Exploratory Search.
In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016) - Volume 2*, pages 427-434
ISBN: 978-989-758-187-8

sions in search user interfaces, including Exploratory Search tools in particular. Among the myriad of involved behaviors, the user's decision of selecting a search result plays an important role in the presentation of results. The way results are presented in the user interface should always be adapted to the user's behavior, promoting access to most relevant documents for a given information need. In Exploratory Search, for instance, this behavior has a stronger influence on the activity outcome, since the first results alone are usually not enough to satisfy the information need. If the first results are chosen with not even regarding their relevance, the listing presentation might not be the best choice for the design of Exploratory Search tools, since low ranked results may not be visited even if they are more relevant to the user's information need. Therefore, it is necessary to investigate and better understand users' behavior at result selection regardless the web search tool in use.

In this paper we present an experimental study with 72 participants that analyzes the user's decision on which search result to access first during an Exploratory Search activity. This study was conducted to examine whether the behavior of selecting the first results would remain in a situation where the results were presented on other medium than a Web search engine. A paper sheet presenting all the results in the same page for easy comparison and selection is the ideal artifact for this experiment because it is a medium in which users' attention tends to be more equally distributed between its contents. Furthermore, it does not introduce other graphical elements or interaction possibilities that may influence users' perception of relevance and disturb their attention.

The paper is structured as follows: in Section 2 we briefly present and discuss the literature review and some related work; in Section 3 we detail the experimental study considering a structure with definition, planning, operation, results analysis and interpretation, and evaluation of experiment validity; in Section 4 we present and discuss the main results; finally, in Section 5 we present the main conclusion and directions for future research.

2 LITERATURE REVIEW AND RELATED WORK

According to Hearst (2009), users rarely look beyond the first page of search results. If they do not find immediately what they want on the first page, the most common behavior is to give up on those results and reformulate the query. Furthermore, studies suggest that Web users expect the best answer to be in the first or second position, and this expectation has influence on the result accessing behavior.

In a study by Granka et al. (2004) with 26 participants and 397 queries, it was found the first search result was selected approximately 56% of the time, and the second about 14%. It was concluded the first two search results are considerably the most viewed and accessed, accumulating about 70% of all accesses. From the third result onwards there is a dramatic decrease in not only access, but also on visualization (assessed by eye tracking).

In study by Joachims et al. (2005) with 29 participants, it was found the probability of a search result being accessed was 45% for the first one, 17% for the second, 11% for the third and 5% or less for the subsequents, accumulating about 73% of all accesses in the first three results. Such behavior persisted even for 5 participants whose results were presented in reversed order, and also for other 5 participants where only the first two results were reversed. An additional analysis comparing the access frequency of the first and second results with a previous relevance judgment indicated participants still prioritized the first result even though its relevance was previously assigned inferior compared to the second result.

A divergent behavior was observed in a study carried out by Aula et al. (2005) with 42 students and 10 pre-defined queries. They found the presence of two different styles of visual exploration over search results: (1) nearly 46% of the participants were found, in the words of the authors, "economic", visualizing only 3 or 4 results at most in 50% of the tasks; and (2) the other 54% were found, also in the words of the authors, "exhaustive", visualizing more than half the results, and in some cases even all the 10 results, for the most of queries. The authors suggest that more experienced users are more likely to adopt the "economic" profile, while less experienced users are more likely to adopt the "exhaustive" profile. Thereby, they argue the study conducted by Granka et al. (2004) has employed only experienced searchers, which would explain the difference in results between studies.

Another factor to be considered is the effect caused by trust in search engine brand. As shown in studies by Jansen et al. (2009) and Jansen et al. (2012), users usually show a bias toward trusting some search engines. Participants were presented with the same results in different search engines, and attributed more relevance for the results from their favorite search engine. As another example, in a study carried out by Pan et al. (2007), users tended to trust and select the first results from *Google* even though their snippets were considered less relevant than following results.

The above studies identify the users' behavior during the analysis and selection of search results retrieved by a search engine. They show higher ranked results attract more attention and, therefore, have a higher chance of being accessed. In addition, the user's experience with search also contributes to the depth of results exploration, in which less experienced users tend to go further in the list whereas more experienced users tend to stick to the first results. However, such studies do not allow to conclude whether the behavior exhibited by users is a ranking consequence (in the sense of the presentation model, not the trust in the search engine ranking algorithm), a medium consequence (eye tracking analysis by Granka et al. (2004) show not every result receives the same attention) or a combination of both.

In an earlier study with 78 participants separated in two experiments, Eisenberg and Barry (1988) found that the order in which documents are presented has influence on the user relevance judgment. The authors verified that when the documents were presented to participants in a high to low relevance ranking, they consistently underestimated the significance of documents at the end of the list. In a low to high situation, there was overestimation of documents, particularly in the first middle range. These results may have been influenced by the medium: the documents were presented in paper sheets, one document at a time, and participants were not able to return to previous analyzed documents.

In the studies we presented in this section, participants's attention is not distributed equally across the documents, either on account of the media or the manner in which documents are presented. Hence, a similar study where the participants's attention tend to be evenly distributed between the presented search results might provide useful information.

3 EXPERIMENTAL STUDY

We (1) scoped, (2) planned, (3) operated, (4) analyzed and interpreted the results and (5) evaluated the experiment validity according to the experiment structure suggested by Wohlin et al. (2012). This structure was chosen because it describes systematically how to conduct and evaluate experiments not only in Software Engineering (its original purpose), but also in Computer Science in general. Each step is described in this section.

3.1 Experiment Scoping

We applied the Goal Question Metric (GQM) template by Basili et al. (1994) and expanded by Van Solingen and Berghout (1999) to structure, conduct and summarize the experiment's nature as well as the main aspects involved.

The following description about the experiment objective was elaborated: we are interested in **observing and analyzing** users' behavior when selecting a search result from a list, removing possible additional influences conventional search engines may exert; **for the purpose of** understanding relevant aspects of user behavior during Exploratory Search activities, which can be applied in informed design decisions for Exploratory Search tools; **with respect to** ranking influence in search result selection behavior shown by users; **from the viewpoint of** the design of Exploratory Search tools; and **in the context of** undergraduate and graduate students of Computer Science and Informatics performing Exploratory Search activities.

3.2 Experiment Planning

Experiment planning was composed of five main steps. In general terms, we: (1) elaborated the hypotheses to be tested; (2) planned the search results selection; (3) selected the experiment participants; (4) defined the experiment variables; and (5) elaborated the experiment instrumentation. Each step is described as follows.

3.2.1 Hypotheses Formulation

The following hypotheses were elaborated and tested:

- **Null Hypothesis** (H_0)**:** Users are not influenced by the search results ranking when selecting results to access.

$$Corr(SampleDist, StandardDist) = 0 \quad (1)$$

- **Alternative Hypothesis** (H_1)**:** Users are negatively influenced by the search results ranking when selecting results to access.

$$Corr(SampleDist, StandardDist) < 0 \quad (2)$$

- **Alternative Hypothesis** (H_2)**:** Users are positively influenced by the search results ranking when selecting results to access.

$$Corr(SampleDist, StandardDist) > 0 \quad (3)$$

In Equations 1, 2 and 3, *Corr* is a correlation function. *SampleDist* is the access distribution observed for each position in the search results list and *StandardDist* is the access probability for each position in the results list, according to the related work of Granka et al. (2004) and Joachims et al. (2005). These two studies are used since the data can be compared with the data collected in this experiment.

3.2.2 Planning of Search Results Selection

To perform the experiment in a consistent manner with an Exploratory Search activity, we defined an information need compatible with this type of activity and specified it as a query. Such compatibility is grounded on Exploratory Search attributes listed by Kules and Capra (2009) and expanded by Wildemuth and Freund (2012). The information need and query were presented to the participants so that they could identify the search context and associate it to the presented search results. The information need and the query employed in this study are (translated from Portuguese):

- **Information Need:** Suppose you want to enter to the job market with a good job. A large, well-known and praised company in your area is hiring promising professionals and invited you for a job interview. As you would like so much to get the job, but have never taken part in a interview before, you want to know more about how job interviews are conducted and how to get prepared to succeed and get the job.
- **Query:** "how to get prepared for a job interview".

We issued the query to *Google*. This information was omitted from the participant to avoid a possible confidence bias on the search engine brand. We retrieved a list consisting of 100 search results and preselected an aleatory sample of 10 results. We evaluated the relevance of these 10 pre-selected results by their snippets (title, URL and description), and they all were considered relevant to the information need. Then a set of 200 random configurations of the 10 pre-selected results was created from its randomization. The generation of random configurations allowed to neutralize the search engine intrinsic ranking factor, disassociating the result potential quality from its ranking order. Therefore, the behavior of results selection could be evaluated only according to is position in the ranking.

Participants were asked to select the search result likely to be the first one accessed by them in a natural search condition. Only one search result must be selected since, at this time, we were interested only in the behavior of the first selection, which could be compared with the literature.

3.2.3 Participant Selection

Participant selection included students engaged daily in Exploratory Search activities. For viability reasons, we selected undergraduate and graduate students of Computer Science and Informatics during class with the permission of their professor. Every student in class was invited to participate and there was no rejection. In addition, everyone authorized the use of their experimental data and there was no exclusion of participants after the data was collected. This process resulted in a sample of $N = 72$. Participation was voluntary, unpaid, with the possibility to leave the study at any time with no justification. The participants were not randomly selected due to viability reasons. However, the assignment of search results lists was randomized among the participants, characterizing this study, according to Wohlin et al. (2012), as a quasi-experiment.

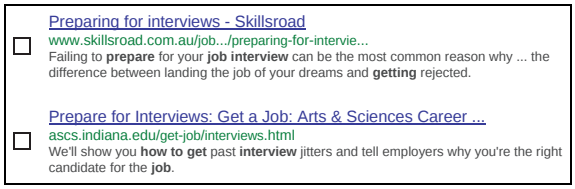

Figure 1: A fragment exemplifying how the results are presented to the participants.

3.2.4 Variable Definition

The following variables were defined: (1) a list of 10 search results randomly sampled (independent, as it represents the input, the possible cause for the analized effect); and (2) the ranking influence rate on users' search result selection behavior (dependent, as it represents the output, the analized effect).

3.2.5 Instrumentation

The objects used in the experimentation process were:

- a document describing the experiment, the general context of the proposed information need and the query issued to obtain the search results, along with a consent form;
- a form containing a checklist with 10 search results in random order, in which, checking a result means the participant would access it. This instrument does not contain any search user interface element other than the results themselves and selection checkboxes, as shown in Figure 1; and
- a participant characterization questionnaire asking educational level and experience with search. Regarding experience with search, those were the categorization options: (i) **very experienced:** I know how to use advanced search features such as logical operators and filters, and I can find anything I search for; (ii) **experienced:** I am not familiar with advanced search features, but I do perform searches daily and I easily find what I need; (iii) **intermediate:** I perform searches daily, but I

do not always find what I need; and (iv) **inexperienced:** I only perform searches casually, and I have difficulty knowing how to search properly.

The independent variable (list of 10 search results randomly sampled) was measured from the participant's selected item in the checklist. The dependent variable (ranking influence rate on users' search result selection behavior) was measured from the correlation between the user's behavior observed in related studies and the curve generated from the selection made by the participants.

3.3 Operation

We prepared and presented the experiment instruments (Section 3.2.5) to the participants, who became aware of the information need and proceeded to selected the first search result they potentially would access in the given context. Participants were instructed to check only one search result in a way as natural as possible. To promote a natural behavior they were told there was not a "correct" answer. The experiment was performed individually by every participant, with no external aid and time limit. Since we did not detect any kind of cheating or improper answering during the experiment, the obtained results were considered valid for the purpose of this study.

3.4 Analysis and Interpretation

We collected and summarized the data to identify the total number of hits for each search result position in the list. Data concerning the participants' characterization, consisting of their education and their experience with search was also accounted for. In a sample of $N = 72$ participants, 65 of them (90.28%) are undergraduate students and 7 (9.72%) are graduate students. Regarding the experience with search, 8 (11.11%) participants claimed to be intermediate, 46 (63.89%) experienced and 18 (25.00%) very experienced. No participant claimed to be inexperienced.

For purposes of comparison, the data analysis requires a distribution to be used as a referential. Therefore, we established an equation based on data obtained by Granka et al. (2004) and as well by Joachims et al. (2005). The equation is based on the decay of access probability according to the search result position. A curve that represents this behavior can be described as $StandardDist = a/x$, such that the area under the curve in the interval $[1;10]$ is equal to 1, the accumulated probability for the search result selection. This constraint can be written as:

$$\int_1^{10} \frac{a}{x}\,\mathrm{d}x = 1 \tag{4}$$

Solving Equation 4 we find $a = 1/\ln 10$. Thus, the probability curve can be rewritten as:

$$StandardDist = \frac{1}{x \ln 10}, \tag{5}$$

where x is the position in the search results list and $x \in [1;10]$.

We validated this curve by performing the correlation between its points (*StandardDist*) and the distribution obtained in the experiments by Granka et al. (2004) (which we call *GrankaDist*) and Joachims et al. (2005) (which we call *JoachimsDist*). The correlation statistic was applied because it is able to directly measure the relationship significance between two ranked distributions, which can be interpreted by a predefined scale. As the three sets of values are non-normal distributions (according to Shapiro-Wilk test with a significance level of 1%), the Spearman correlation test was applied with a correlation range passing through -1.0 (perfect negative correlation), 0.0 (no correlation) and 1.0 (perfect positive correlation). Values obtained for the Spearman correlation for the distributions were, both with significance level equal to 0.1%:

$$Corr(GrankaDist, StandardDist) = 0.8842$$
$$Corr(JoachimsDist, StandardDist) = 0.9232$$

These values indicate the *StandardDist* curve is consistent with previous studies, showing a strong positive correlation. Therefore, the curve can be used as a reference in correlation with the results sampled in this study.

In Figures 2, 3 and 4 we illustrate the graphics generated for the performed correlation. In Figures 2 and 3 the data proximity to the curve *StandardDist* is clearly visible, which is plausible since the positive correlation is strong. Although the correlation illustrated in Figure 4 is not clearly visible, the Spearman correlation (employed since the study sample comes from a normal distribution and *StandardDist* is a non-normal distribution, according to the Shapiro-Wilk test with significance level of 1%) shows there is a strong positive correlation, with significance level equal to 1%:

$$Corr(SampleDist, StandardDist) = 0.7693$$

Therefore, this correlation value allows the null hypothesis (H_0) and the alternative hypothesis (H_1) to be rejected with a reliability of 99%, and the alternative hypothesis (H_2) to be accepted, indicating users are positively influenced by the search results ranking during selection since:

$$Corr(SampleDist, StandardDist) > 0$$

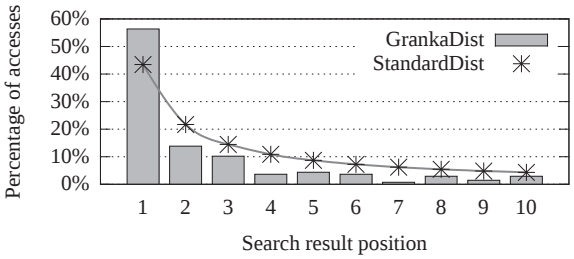

Figure 2: Correlation between *GrankaDist* and *StandardDist* distributions.

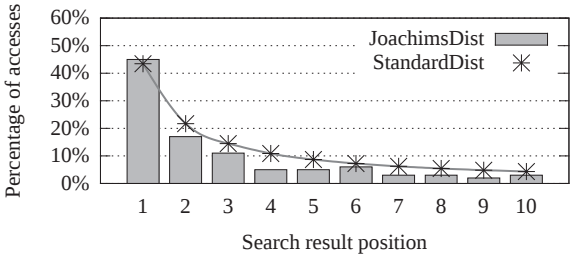

Figure 3: Correlation between *JoachimsDist* and *StandardDist* distributions.

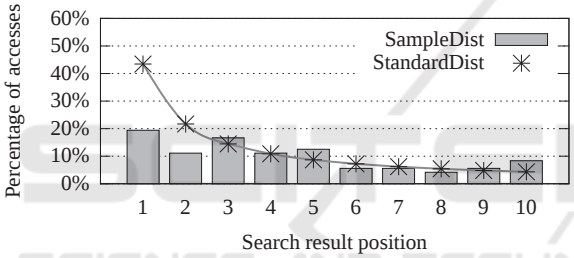

Figure 4: Correlation between *SampleDist* and *StandardDist* distributions.

3.5 Validity Evaluation

Regarding the conclusion validity, the sample size ($N = 72$), although larger than those of related work, may not be sufficient to represent the population of students who perform Exploratory Search activities. The main risk for the conclusion generalization is that all the participants are students from the IT area in the same institution. Replications of this experiment should consider participants from other areas of knowledge, from different institutions and cultures.

Regarding the construct validity, there was a great care for the participants to understand correctly what should be done in the experiment. The information need was described and explained clearly and succinctly, and the task to be performed by the participants was simple enough given their education. No participant had difficulties in understanding and carrying out the experiment.

Regarding the experiment internal validity, the following threats were considered:

- **Answers Accuracy:** each participant selected only one search result. They were requested to select the first result they would access on a real search situation; it was a simple activity that everyone easily understood, so the responses were considered accurate.
- **Fatigue Effects:** the experiment, including explanation, lasted approximately 15 minutes and the requested task did not require any form of excessive cognitive load. Thus, fatigue effects were not a problem.
- **Influence Among Participants:** the experiment was supervised by the authors and there was no exchange of information between participants during the study. Therefore, the influence among participants was not a problem.
- **Participants Diversity:** the study sample can not guarantee this behavior also holds for people from other areas of knowledge and other cultures. As pointed out by Weinschenk (2011), culture can affect how people think. Western cultures (the experiment case) have a tendency to focus on individual elements, whereas eastern cultures have a tendency to focus on context and relationships. Furthermore, the predominant amount of users who consider themselves very experienced or experienced (88.89% in total) may have influenced the observed behavior due to the "economic" profile commonly adopted by this kind of user (Aula et al., 2005).

Regarding the experiment external validity, the following threats were considered:

- **Artificial Task:** Russell and Grimes (2007) argue that artificial search activities do not accurately reflect the behavior of natural search activities, thus, to "impose" a search activity for the participant is not an ideal approach. The authors, however, admit that artificial search activities are necessary due to research viability, and that they can provide useful information.
- **Atypical Media:** in this study, the digital media on which searches actually occur was replaced by the printed medium. As a consequence, the behavior of comparison and selection was different in terms of scan and selection time. However, although the printed medium is not the natural context of a search activity, the choice was deliberate so that participants could distribute their attention more evenly between the search results, and were not influenced by graphical elements and interactions possibilities offered by traditional web search engines.

4 RESULTS DISCUSSION

The correlation found in the present study (0.7693) was not as high as the correlation found in the other two distributions from Granka et al. (2004) (0.8842) and from Joachims et al. (2005) (0.9232). However, it is still a strong positive correlation, and hence it still corroborates the evidence that the search result selection behavior is influenced by ranking, with users showing a tendency to select higher ranked results.

As a possible reason for the relatively lower correlation value, it can be noted that users, on average, selected lower ranked search results when compared to related work. In the present study, the selected result average position was $\mu = 4.37$ with a standard deviation of $\sigma = 2.87$, whereas in the related studies, users where mostly concentrated in the first two results. Although this average may appear as an homogeneous distribution, approximately 71% of the selections were concentrated in the five higher ranked results, corroborating that higher ranked results are selected more often. The selection of lower ranked results in comparison with related work can be attributed to factors such as:

- the participants' experience with search combined with their education and area of knowledge: 88.89% of the participants declared themselves very experienced or experienced, hence an "economic" profile would be expected, however, their level of education and their area of knowledge may lead to other behaviors;
- the experiment was carried out in paper, evoking a different behavior from a real search activity. While people rarely look at everything in Web pages, not even seeing most of the search results, with paper sheets they tend to give at least a look at every result. At the same time, they can't go beyond the 10 results presented to them, as they can on the Web; and
- the participants had performed a suggested search task, and not a natural search activity, which may have implied in an unnatural search behavior, and thereafter, an unnatural search result selection behavior.

Overall, the evidence acquired through this study can be useful during informed design decisions for Exploratory Search tools. Since the observed behavior can represent a risk to the success of an Exploratory Search activity, Exploratory Search tools should try to encourage users to take a different approach by exploring more deeply the results. This could be approached, for example, with novel user interfaces or recommendation systems.

5 CONCLUSION

Although Exploratory Search has been discussed in several studies in recent years, current search engines still lack support for this kind of activity. Therefore, there is still room for models and tools to support the user in Exploratory Search activities, and the design of such models and tools can benefit from understanding users' behavior during this kind of activity. Accordingly, information about the search result selection behavior can be particularly useful for informed design decisions in Exploratory Search tools.

Studies indicate users give more attention and, therefore, more hits to higher ranked search results. However, such studies do not allow to conclude whether this behavior is only due to the ranking schema, whereas there are other factors possibly influencing user behavior. Therefore, we investigated whether such behavior remains in a medium which the user's attention tends to be distributed more evenly between the presented results and with no additional graphical tricks and interaction possibilities.

We defined a standard distribution of access probability based on a decay curve according to the search result position. This standard distribution was evaluated against two other distributions from the literature to be used as a baseline for the experiment. The sample distribution was then compared to the baseline and, by means of a strong positive correlation, we provided evidence the users' behavior on search result selection is indeed influenced by the ranking schema.

Such finding is complementary to the aforementioned related work, with the addition that the influence occurred even though the results were randomly reordered, and participants tended to distribute their attention more equally between the results. Since giving more attention and, therefore, more hits to higher ranked results is not ideal for Exploratory Search activities, tools for these kind of activities should encourage users to explore more deeply the results. Thereby, the listing presentation may not be the best choice for the design of Exploratory Search tools, once it is not guaranteed the most relevant results will always be high ranked.

For future work we aim to (i) replicate the experiment aiming to achieve a more distributed sample; (ii) carry out a new experiment containing varied information needs and, consequently, different results lists; (iii) conduct another experiment allowing the participants to select as many results as they want; and (iv) carry out both qualitative and quantitative studies to analyze different behaviors of users engaged in Exploratory Search activities from users engaged in simpler search activities.

ACKNOWLEDGEMENTS

This study was financially supported by Coordenao de Aperfeioamento de Pessoal de Nvel Superior (CAPES) and Conselho Nacional de Desenvolvimento Cientfico e Tecnolgico (CNPq). The authors would like to dedicate this work to Professor Srgio Roberto Pereira da Silva who is dearly missed.

REFERENCES

Aula, A., Majaranta, P., and Räihä, K.-J. (2005). Eye-tracking reveals the personal styles for search result evaluation. In *Proc. Int. Conf. IFIP TC13 on Human-Computer Interaction*, pages 1058–1061. Springer-Verlag.

Basili, V., Caldiera, G., and Rombach, H. D. (1994). *Encyclopedia of Software Engineering*, chapter The Goal Question Metric Approach, pages 528–532. John Wiley & Sons, Inc.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424.

Blandford, A. and Attfield, S. (2010). *Interacting with Information*, volume 3 of *Synthesis Lectures on Human-Centered Informatics*. Morgan & Claypool Publishers.

Eisenberg, M. and Barry, C. (1988). Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *Journal of the American Society for Information Science*, 39(5):293–300.

Granka, L. A., Joachims, T., and Gay, G. (2004). Eye-tracking analysis of user behavior in www search. In *Proc. Int. Conf. ACM SIGIR on Research and Development in Information Retrieval*, pages 478–479. ACM.

Hearst, M. (2009). *Search User Interfaces*. Search User Interfaces. Cambridge University Press.

Jansen, B. J., Zhang, L., and Mattila, A. S. (2012). User reactions to search engines logos: investigating brand knowledge of web search engines. *Electronic Commerce Research*, 12(4):429–454.

Jansen, B. J., Zhang, M., and Schultz, C. D. (2009). Brand and its effect on user perception of search engine performance. *Journal of the American Society for Information Science and Technology*, 60(8):1572–1595.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *Proc. Int. Conf. ACM SIGIR on Research and Development in Information Retrieval*, pages 154–161. ACM.

Kules, B. and Capra, R. (2009). Designing exploratory search tasks for user studies of information seeking support systems. In *Proc. ACM/IEEE-CS Joint Conf. on Digital libraries*, pages 419–420.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., and Granka, L. (2007). In google we trust: Users decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3):801–823.

Russell, D. and Grimes, C. (2007). Assigned tasks are not the same as self-chosen web search tasks. In *Proc. Int. Conf. HICSS on System Sciences*, pages 83–83.

Van Solingen, R. and Berghout, E. (1999). *The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development*. McGraw-Hill Higher Education.

Weinschenk, S. (2011). *100 Things Every Designer Needs to Know About People*. New Riders Publishing.

White, R. W., Capra, R., Golovchinsky, G., Kules, B., Smith, C., and Tunkelang, D. (2013). Introduction to special issue on humancomputer information retrieval. *Information Processing & Management*, 49(5):1053 – 1057.

White, R. W., Kules, B., and Bederson, B. (2005). Exploratory search interfaces: categorization, clustering and beyond: report on the XSI workshop at the human-computer interaction laboratory, university of maryland. *SIGIR Forum*, 39(2):52–56.

White, R. W. and Roth, R. A. (2009). *Exploratory Search: Beyond the Query-Response Paradigm*, volume 1 of *Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool Publishers.

Wildemuth, B. M. and Freund, L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. In *Proc. Symp. on Human-Computer Interaction and Information Retrieval*, pages 4:1–4:10. ACM.

Wohlin, C., Höst, M., Runeson, P., Ohlsson, M., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Computer Science. Springer.