Queries without Clicks: Successful or Failed Searches?

Sofia Stamou
Computer Engineering and Informatics Department
Patras University, GREECE

stamou@ceid.upatras.gr

Efthimis N. Efthimiadis
Information School, University of Washington
Seattle, WA, USA

efthimis@u.washington.edu

ABSTRACT

The critical aspect in the evaluation of retrieval effectiveness is the satisfaction of the user needs in the retrieved results. Current efforts for evaluating retrieval performance rely either on explicit user feedback or on the analysis of the search transaction logs in order to elicit the user needs and thus be able to infer their satisfaction in the retrieved results. In this paper, we propose a method for evaluating the user satisfaction from searches not followed by clickthrough activity on the retrieved results. To that end, we carried out a user study in order to identify the search intentions of queries without follow-up clicks. Our findings indicate that queries without clicks may pursue specific search goals that can be satisfied in the list of retrieved results the user views rather than in the contents of the documents the user visits for the query.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: search process

General Terms

Performance, Experimentation, Human Factors.

Keywords

Task-oriented search, queries without clickthrough.

1. INTRODUCTION

With the advent of the web and the proliferation of both information sources and information seekers, there has been a shift of interest from the retrieval of query-relevant documents to the retrieval of information that is relevant to the user needs. Automatically identifying the user needs is a challenging task that has mainly focused on the analysis of the user activity on the query results [6] [7]. Although clickthrough data can be perceived as an indicator of implicit user feedback on the relevance of retrieved results [5], it might generate biased relevance judgments unless we consider that users make click decisions based on a limited set of options, i.e. the displayed information on the results page [8].

Recently, researchers proposed that context of search, i.e. the task the user is trying to accomplish, should be the driving force in the quest for effective retrieval evaluation [4] [15]. In this respect, there have been proposed user-centric approaches to the evaluation of retrieval performance [13][14]. The commonality in the above approaches is that they rely on the analysis of the user interaction with the retrieved results for judging their usefulness in satisfying the user search intentions. One aspect that existing IR evaluation techniques do not systematically address is the user's perception of the usefulness of the results retrieved but not visited. Despite the acknowledgment that some queries are not followed by result clicks because the desired information is presented in the snippets (abstracts) of the results [12], to our knowledge no effort has been reported that investigates the contribution of retrieved but un-visited results in relation to users' tasks. In this paper, we investigate the impact that retrieved but not visited results might have on user satisfaction from retrieval effectiveness and examine whether and how these should be accounted in the retrieval evaluation process. First, we present the findings of a survey we carried out in order to identify the context of searches without clicks. In Section 3 we propose a model for evaluating the effectiveness for contextual searches not followed by result clicks.

2. SEARCHES WITHOUT CLICKS

The goals that lead people to engage in information seeking behavior affect their judgments of usefulness of the retrieved results [2]. This, coupled with the observation that nearly 50% of the searches do not result on a single click on the results [4], motivated our study on how to evaluate retrieval

effectiveness for queries not followed by result clicks. To that end, we carried out a survey in order to identify the intentions associated with queries not followed by clicks. We recruited 38 postgraduate computer science students and asked them to answer four questions per search performed on their preferred search engine(s) in a single day. The questions, presented to our subjects via an online questionnaire, asked if they did or did not click on results and the reasons for it. Specifically, we instructed our participants to open the questionnaire in a new browser window while conducting their searches and answer the questions for each of their queries right after the submission of the query and the review of the retrieved results. Before conducting our survey we familiarized our subjects with the questions by giving them verbal explanations for every question. The collected user feedback was anonymous in the sense that neither the user identities nor their issued queries or preferred search engines were recorded. Table 1 reports selected results of our survey.

Table 1: Queries without clicks - Survey Results

Examined queries	908
Queries with clicks	87.22%
Queries w/o clicks (intentional-cause)	6.06%
Queries w/o clicks (unintentional-cause)	6.72%
Classification of unintentional queries w/o clicks	
No results retrieved	14.78%
Displayed results seemed irrelevant	62.29%
I have already seen these results for the query	13.11%
Search was interrupted	9.82%
Classification of intentional queries w/o clicks	
Check spelling/syntax of query term(s)	30.91%
See if there's a new page retrieved from the last time I issued the query	32.73%
Find out what the query is about by looking at the retrieved abstracts	21.82%
See if there's a web site about my query	14.54%

The study showed that the reasons for not clicking on the query results fall into two categories: intentional-cause and unintentional-cause. The unintentional cause for not clicking is encountered when the user submits a query, but the retrieved results are unexpected to the user, hence they decide not to click. These reasons (Table 1) are: nothing retrieved, seemed irrelevant, already seen, interrupted search. Conversely, the intentional cause for not clicking is encountered when the user issues a query with a predetermined intention to look for answers in the results' snippets and without following any link. According to our participants, searches without clicks are encountered when they want to accomplish the following types of tasks: (i) get an update or (ii) obtain instant information about the query. In particular, the information goal of users engaging in an *update*[1] search is to find out if there is *new* information retrieved since their last submission of the query. On the other hand, the goal of users performing an *instant* search is to obtain information about the query from the title or the snippets of the displayed results. In both cases, the information need of the user engaging in update or instant searches can be satisfied by the contents of the result list displayed (i.e. the snippets) without the need to follow any results per se. Therefore, retrieval effectiveness for update and instant searches that do not generate clickthrough activity could also be evaluated based on the results displayed to the user. We recognize that this is rather difficult; however a model that attempts this is discussed below.

---

[1] Update searches as determined by our users could be perceived as an instance of repeat searches [16] since they both concern queries the user has issued in the past.

3. DISCUSSION

Given the findings of our study, we propose a retrieval evaluation framework for queries without clicks. Our evaluation relies on the observable user activity following a query submission in order to infer the user perception of the displayed results' usefulness. The idea of utilizing the searcher activity on the returned results as an indicator of implicit relevance judgments is not new. There exists a large body of work on how the different post-query activities can be interpreted as implicit feedback signals (for an overview see [11]). The searchers' behavior that researchers observed as implicit measures of interest are: time spent on a page combined with the amount of scrolling on a page [3], duration of search and number of result sets returned [5], click data on and beyond the search results [9], use of eye-tracking methods to capture the user's visual attention on the results [10], repetition of result clicks across user sessions [16]. Although, the above measures have been applied for inferring the user satisfaction from the results visited for some query, we propose their utilization towards capturing the user satisfaction from the results displayed for queries not followed by clickthrough events. From the above measures, we obviously exclude click data since we are dealing with searches not followed by result clicks.

Our proposed model examines the post-query user activity in order to firstly identify the user goals for queries without clicks and then based on the identified goals to infer the user satisfaction from search results. Our model first examines whether a query without clicks returned any results. If the query retrieved no documents, then it concludes that search failed to satisfy the user needs. On the other hand, if the query retrieved results that the user did not visit, our model tries to deduce the user satisfaction from retrieval effectiveness based on the examination of the following features (partially based on the proposal of [1]): (i) time spent on a results page combined with the amount of scrolling on the page (ii) terminological overlap between the query term(s) and the displayed result titles and/or snippets, (iii) terminological overlap between two consecutive queries, (iv) repetition of the query and (v) type of user activity on the displayed results (e.g. read, copy text from snippet, move to the next results page). The idea is that the features characterizing the post query user activity are valuable indicators of the query intentions. Thus, if the intention of the query is to obtain information in the snippets of the displayed results, then evaluation of retrieval performance should concentrate on the usefulness of the result snippets. The features that characterize the intentional cause of queries without clicks and which imply the user satisfaction from the search results can only be determined explicitly via user studies. Next, we discuss a probabilistic approach for capturing the query intention and the user satisfaction from searches not followed by clicks. Our approach relies on the combination of the following measures that are presented below: (a) query refinement probability, (b) query results usefulness, and (c) update search probability.

Query refinement probability, i.e. the probability that a query q which did not yield result clicks was refined in the search (q_i) that immediately followed. Formally, $p(q|q_i)$ can be determined proportionally to the number of overlapping terms between q and q_i. If $p(q|q_i)$ exceeds a threshold (to be empirically determined via user studies), then q was refined in its succeeding search (q_i) and we may conclude that the user did not satisfy her information needs in the results displayed for q. If $p(q|q_i)$ is below the threshold, i.e. q is not refined in the next search, we examine the following:

Query-results usefulness, i.e. the probability that q was not followed by result clicks because it was satisfied in the list of displayed results. To derive such probability, we rely on the terminological overlap between the query term(s) q_t and the terms s_t in the result titles and/or snippet, given by: $O(q, r) = |q_t \cap s_t| / |s_t|$. In addition, we estimate the amount of time the user spent on the results list as well as the type of the demonstrated user activity on the results. The combination of the above measures can serve as an approximation of the displayed results' usefulness to the query intention. Again, threshold weights should be empirically set via user studies before the deployment of our approach to a retrieval evaluation setting. Another factor we should examine is the:

Update search probability, i.e. the probability that the user intention is to obtain new information about a previous search. The probability $p(q)$ that the query has been submitted before can be determined based on the frequency of q in the observable user searches. If $p(q)$ exceeds a given threshold, then q probably represents an update search. User satisfaction from the

results retrieved for an update query can be estimated based on: $p(r_n) \cdot O(q, r_n)$ where $p(r_n)$ is the probability that r_n is a new result not previously retrieved for q and $O(q, r_n)$ is the probability that q is satisfied in the information displayed for r_n. This probability combined with the amount of time spent on the results and the type of user activity on the results can give rough indications of the user satisfaction from update searches. Again, user studies need to be carried out in order to set the threshold values upon which conclusions can be drawn.

For queries without clicks that are not refined in their succeeding searches and do not represent update requests, as well as for queries without clicks that have low probability of being satisfied in the result snippets or they are the last searches in the user session, the only way to assess user satisfaction from displayed results is in terms of explicit user feedback.

4. CONCLUDING REMARKS

We have proposed the utilization of implicit feedback measures for inferring the user satisfaction from searches not followed by result clicks. The parameters of our approach need to be validated and fine-tuned via additional user studies. We hope that our approach will contribute towards the design of IR evaluation frameworks where search is seen holistically and incorporate multiple features for measuring retrieval quality.

5. REFERENCES

[1] Agichtein, E., Brill, E., Dumais, S. and Rango, R. 2006. Learning user interaction models for predicting search result preferences. In the 29th ACM SIGIR Conference.

[2] Belkin, N. 2008. Some(what) grand challenges for information retrieval. In ACM SIGIR Forum, 42 (1): 47-54.

[3] Claypool, M. Le, P., Waseda, M. and Brown, D. 2001. Implicit interest indicators. In Intl. Conference on Intelligent User Interfaces, pp. 33-40.

[4] Callan, J., Allan, J., Clarke, Ch.L.A., Dumais, S., Evans, D.A., Sanderson, M., and Zhai, Ch. 2007. Meeting of the MINDS: an information retrieval research agenda. In ACM SIGIR Forum, 41(2): 25-34.

[5] Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T. 2005. Evaluating implicit measures to improve web search. ACM Transactions on Information Systems, 23(2): 147-168.

[6] Jansen, B.J. and Spink, A. 2006. How are we searching the www: a comparison of nine search engine transaction logs. Information Processing & Management 42(1):248-263.

[7] Jansen, B.J., Booth, D.L. and Spink, A. 2008. Determining the informational, navigational and transactional intent of web queries. Information Processing & Management 44:1251-1266.

[8] Joachims, T., Granka L., Pan, B., Hembrooke H., Padlinski, F. and Gay, G. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. ACM Transactions on Information Systems, 25(2):1-26.

[9] Jung, S., Herlocker, J.L. and Webster, J. 2007. Click data as implicit relevance feedback in web search. Information Processing & Management, 43(3):791-807.

[10] Granka, L.A., Joachims, T. and Gay, G, 2004. Eye-tracking analysis of user behaviour in www results. In ACM SIGIR Conference, pp. 478-479.

[11] Kelly, D. and Teevan, J. 2003. Implicit feedback for inferring user preference: a bibliography. In ACM SIGIR Forum, 37(2):18-28.

[12] Radlinski, F., Kurup, M. and Joachnims, T. 2008. How does clickthrough data reflect retrieval quality. In CIKM Conf.

[13] Sharma, H., and Jansen, B.J. 2005. Automated evaluation of search engine performance via implicit user feedback. In the 28th ACM SIGIR Conference, pp. 649-650.

[14] Spink, A. 2002. A user centered approach to evaluating human interaction with web search engines: an exploratory study. Information Processing & Management, 38(3):401-426.

[15] Taksa, I., Spink, A. and Goldberg, R. 2008. A task-oriented approach to search engine usability studies. Journal of Software, 3(1): 63-73.

[16] Teevan, J., Adar, E., Jones, R. and Potts, M. 2007. Information re-retrieval: repeat queries in Yahoo's logs. In the 30th ACM SIGIR Conference.