



# Re-finding Behaviour in Educational Search

Arif Usta<sup>1(✉)</sup>, Ismail Sengor Altingovde<sup>2</sup>, Rifat Ozcan<sup>3</sup>, and Özgür Ulusoy<sup>1</sup>

<sup>1</sup> Bilkent University, Ankara, Turkey

{arif.usta,oulusoy}@bilkent.edu.tr

<sup>2</sup> Middle East Technical University, Ankara, Turkey

altingovde@ceng.metu.edu.tr

<sup>3</sup> Microsoft, Oslo, Norway

rifatozcan1981@gmail.com

**Abstract.** One of the search tasks in Web search is repeat search behaviour to find out documents that users once visited, which is called *re-finding*. Although there have been several works in the context of general-purpose Web search addressing the latter phenomena, the problem is usually overlooked for vertical search engines. In this work, we report *re-finding* and *newfinding* behaviours of users in an educational search context and compare results with the findings in the literature for general-purpose web search. Our analysis shows that re-finding pattern of students differs from web search drastically as only 26% of all queries indicate *re-finding* behaviour compared to 40% in Web.

**Keywords:** Vertical search · Educational search · Query log analysis · Re-finding

## 1 Introduction

In the literature, analysis of query logs is widely applied to detect search patterns and identify user intents. One of the most common user activities during search is *re-finding*, which constitutes average of 4 out of 5 page visits being to previously seen pages [2]. There have not been many works on analyzing re-finding behavior, until it is noted in [3] that, 17% of the users reports “Not being able to find out a page once visited” as one of the biggest problems to be solved in web search.

One of the large-scale re-finding behavior analysis was done by Teevan et al. [6] through queries issued to Yahoo!. They demonstrate that 40% of query issues lead users into repeat behaviour. Sadeghi et al. [5] analyzed features for detecting a re-finding session in different verticals such as news and movies.

In education domain, which is another popular application area of search, Usta et al. [7] studied general search characteristics of students and compared them with the findings in the literature for general Web. Bilal and Gwizdka [1] analyzed query types and reformulations of students in Google. To the best of our knowledge, none of these works address re-finding behaviour in the context of educational search.

**Table 1.** Categorization of query types in education vertical

All queries: 64078 (100%)	Overlapping click queries - 16676 (26%)			No common clicks 47402 (74%)
	Equal click queries 5010 (8%)		Some common clicks 11666 (18%)	
	Single click 3982 (6%)	Multi click 1028 (1,6%)		
Equal query 22591 (35%)	Navigational queries 2863 (4,4%)	681 (1%)	7355 (11,5%)	11692 (18%)
Different query 41487 (65%)	1119 (1,7%)	347 (<1%)	4311 (7%)	35710 (56%)

In this paper, we explore re-finding behavior of students at K-12 level through analyzing the query logs of a commercial educational search engine, called *Vitamin*. We also report the similarities and differences between educational and general web search in the context of re-finding by comparing our findings to those in [6]. Our findings help understanding re-finding pattern of students in educational search and provide possible new directions to further improve educational verticals.

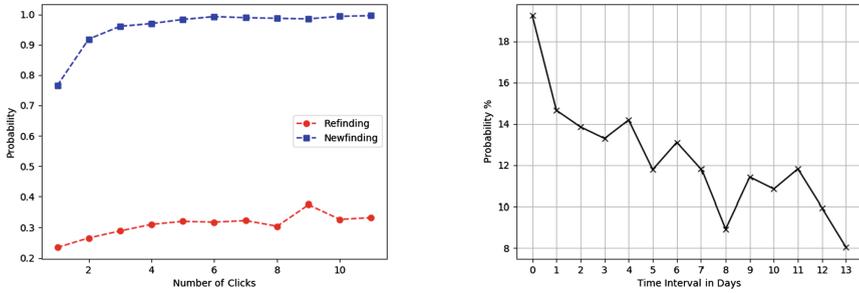
## 2 Re-finding Analysis

For this analysis, we focus on the so-called *re-finding* queries as stated in [6]. Specifically, re-finding query is the query in which user clicks a document that was clicked as a result of another past query issue by again that particular user.

We use a query log consisting of 64,078 queries issued by 18,534 unique users extracted from *Vitamin*. There are also 165,587 learning objects that we refer as documents for the rest of the paper. Other characteristics of query log we use in this paper can be found in [7].

In web, re-finding behavior is observed in 40% of all search sessions, while in our case, only around 26% of queries exhibit re-finding behavior. An analysis of clicked documents also shows differences in comparison to web search. Among all clicked documents, 28% of documents are clicked multiple times by the same user at different search sessions on the web, which indicates Refinding. In contrary, out of all documents clicked at least once in our query log, only 20,594 documents are clicked multiple times by the same user, which roughly corresponds to only 12%.

We categorized query types using the same methodology in [6]. Queries are categorized according to their texts (*Equal vs Different*) and click sets. Overlapping click query results in Table 1 represent re-finding behaviour. Considering query texts, the ratio between *Equal-Query* and *Different-Query* is almost the



**Fig. 1.** Probability of a query session being Refinding and Newfinding given number of clicks (left plot) and Probability of Repeat Queries Having Common Click Depending on Time Intervals (right plot).

same with Web, which is two to one. However, the behavior is quite different than Web for click characteristics. In Web, 29% of queries have equal click sets. In our query log, we only have 8% of queries categorized as *Equal-Click*. One of the reasons for different behavior in education context is that students tend to click more documents [7] than users in Web, which decreases the chance of click-through sets of queries being exactly same even if query texts are the same. Due to this search behavior, the results for navigational queries also differ, constituting only 4% of all queries compared to 24% in Web. Other radical difference between education context and Web is that although there are much less documents available in our query log, the volume of queries having no common clicks is higher than the Web. The result is mostly due to having *Equal-Queries* with no common clicks as 18% of all queries compared to 4% in Web.

### 2.1 Student Click Patterns for Re-finding

Students have different click characteristics than users in Web [7]. They tend to click more which eventually alters re-finding behavior. We examine repeat queries that have overlapping clicks in terms of time intervals between pairs. Unlike Web, highest probability for a repeat query to have common clicks is when queries are issued in the same day, which can be seen in Fig. 1. The probability decreases drastically as time interval between issued queries gets bigger.

Another analysis made in the paper [6] is whether re-finding behavior depends on the number of clicks user made. Among all query issues made, 29% of the search issues which result in single click on a document have re-finding behavior. The results for this analysis on our query log are similar to Web. Students prefer re-finding document 23% of time in a search with a single click.

We also examine queries with multiple clicks for comparison. In web, among all query instances including multiple clicks on documents, only 5.3% of them include re-finding behavior. 30% of queries including multiple clicks lead students to click on a document they once visited. We believe that there are two reasons for this behavior. First, the educational vertical considered in our analy-

sis includes a significantly smaller number of documents in comparison to Web. Hence, students who tend to click more documents in the educational context, eventually end up with clicking some of the documents they have visited before. The second reason would be that students are less likely to remember the documents they visited before in a successful search [4], therefore for a re-finding intent, they have to click more documents than general Web users.

Apart from single or multi-click query instances, we also explore the correlation between the number of clicks in a search and the probability of the query instance being re-finding or newfinding. The results can be seen in Fig. 1. As the number of clicks increases, the probability of query being re-finding increases until certain click number. Newfinding probability is lowest for single click queries as expected.

### 3 Conclusion

In this work, we presented an in-depth analysis of re-finding pattern of students at K-12 level using a query log extracted through a commercial educational search engine. We also compared our findings with Web and reported similarities and differences. Our analysis shows that educational search differs from the Web in terms of re-finding Behavior. We believe there are two different aspects to consider when explaining difference. First, search characteristics of students differ from users in the Web as they may fail in expressing their search intents clearly and tend to click more documents in result lists [7]. Second, searching for an educational document for learning purpose definitely changes user behavior. In educational search environment, students periodically study different materials related to the subjects listed in curriculum. For re-finding, it is understandable to access the same learning object in order to reinforce their knowledge. On the other hand, in terms of learning aspect, it might be rational and wise to choose a document they have never visited to explore what that particular material can offer for them to learn the subject better.

**Acknowledgments.** This research is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under the grant no 113E065. I.S. Altıngövdü is supported by Turkish Academy of Sciences Distinguished Young Scientist Award (TUBA-GEBIP 2016).

### References

1. Bilal, D., Gwizdka, J.: Children's query types and reformulations in google search. *Inf. Process. Manag.* **54**(6), 1022–1041 (2018)
2. Cockburn, A., Greenberg, S., Jones, S., McKenzie, B., Moyle, M.: *Improving web page revisitation: analysis, design, and evaluation* (2003)
3. iProspect: Search engine user behavior study. [http://district4.extension.ifas.ufl.edu/Tech/TechPubs/WhitePaper\\_2006\\_SearchEngineUserBehavior.pdf](http://district4.extension.ifas.ufl.edu/Tech/TechPubs/WhitePaper_2006_SearchEngineUserBehavior.pdf). Accessed 20 May 2019

4. Rutter, S., Ford, N., Clough, P.: How do children reformulate their search queries? *Inf. Res. Int. Electron. J.* **20** (2015)
5. Sadeghi, S.S., Blanco, R., Mika, P., Sanderson, M., Scholer, F., Vallet, D.: Re-finding behaviour in vertical domains. *ACM Trans. Inf. Syst.* **35**(3), 21:1–21:30 (2017)
6. Teevan, J., Adar, E., Jones, R., Potts, M.A.S.: Information re-retrieval: repeat queries in Yahoo’s logs. In: *Proceedings of SIGIR*, pp. 151–158 (2007)
7. Usta, A., Altingovde, I.S., Vidinli, I.B., Ozcan, R., Ulusoy, O.: How k-12 students search for learning?: analysis of an educational search engine log. In: *Proceedings of SIGIR*, pp. 1151–1154 (2014)