

# Feedback on the Semantic Relevance of Search Queries

**Herre van Oostendorp**

Department of Information and Computing  
Sciences  
Utrecht University  
Utrecht  
The Netherlands  
h.vanoostendorp@uu.nl

**Tijs Ditvoorst**

Department of Information and Computing  
Sciences  
Utrecht University  
Utrecht  
The Netherlands  
tijsditvoorst@hotmail.com

**Justin van Doorn**

Department of Information and Computing  
Sciences  
Utrecht University  
Utrecht  
The Netherlands  
Justin.vd05@gmail.com

## ABSTRACT

In this paper, we describe research in which visualization of the semantic relevance of queries is examined with younger and older adults when they reformulate their queries. Feedback on the semantic relevance of search queries and search results increased the semantic relevance of search queries as they reformulated for younger as well as for older adults and this applies to more difficult search problems.

## Author Keywords

Search queries, reformulations, semantic relevance, feedback.

## ACM Classification Keywords

- **Information Search and Retrieval** → Search Process
- **User Interfaces** → User/Machine Systems → Human Information Processing

## ACM Reference format:

H. van Oostendorp, T. Ditvoorst, and J. van Doorn. 2018. Feedback on the Semantic Relevance of Search Queries. In *Proceedings of ACM ECCE conference, Utrecht, Netherlands, September 2018 (ECCE 2018)*, 4 pages. <https://doi.org/10.1145/3232078.3232087>

## 1 INTRODUCTION

Searching for information on the Internet can be challenging and involves cognitive processes such as attention, comprehension, memory, problem solving and decision making [1].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
ECCE'18, September 5–7, 2018, Utrecht, Netherlands

© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-6449-2/18/09...\$15.00  
<https://doi.org/10.1145/3232078.3232087>

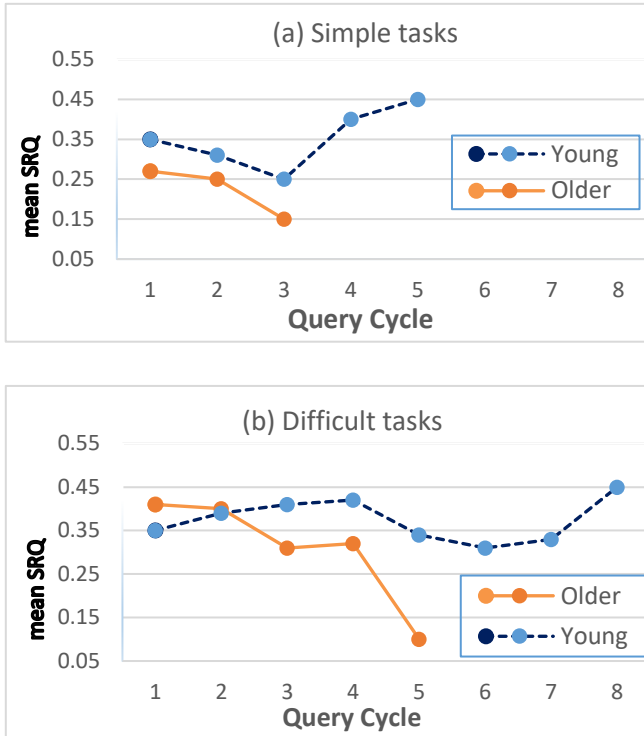
These cognitive processes are known to be affected by aging-related declines in cognitive abilities. It is e.g. known that older adults generate less queries when using a search engine, use less keywords per query, reformulate less, spend more time evaluating the search results, spend more time evaluating the content of websites opened from search engine results pages (SERPs), and switch less between SERPs and websites [2-5]. Information problem solving tasks frequently require users to issue more than one query and to reformulate. Older adults find it often difficult to reformulate their previous unsuccessful query, leading to suboptimal search performance [6, 7]. In empirical literature, there has not been much attention to semantic (knowledge-related) aspects of reformulations and that is the focus of the current study. In next section we will first discuss some studies on the semantics of reformulations, and on differences between younger and older adults. In the third section we present a study that addresses the question whether support of users in reformulating their query by giving them visual feedback on semantic relevance enhances the relevance of their query for finding the answer to the problem. In last sections we draw conclusions and point out limitations of current study.

## 2 STUDIES ON THE SEMANTIC ASPECT OF REFORMULATIONS

A number of interfaces are developed recently to assist users during the information search process [e.g. 8-10]. However, these interfaces do not directly examine the phase of issuing a query and its reformulation. An exception is Scentbar [developed by 11]. Scentbar does focus on the queries by visualizing the amount of missed, relevant information. However it does not provide an estimate of how semantically relevant a user's query is, which is exactly what our proposed interface plans to do. Furthermore, our interface also enables the user to assess quickly the semantic relevance of search results, where with Scentbar users still have to spend time analyzing these.

Research on search strategies has so far overlooked age-related differences in the actual content of the queries during reformulations, and analysis of the semantic aspects of reformulations forms the starting point of our research. We briefly describe two studies that did examine effects of age and task difficulty on reformulating queries. The focus in these studies [7, 12] was whether the relevance of the reformulated queries

semantically increased compared to the preceding queries. For this purpose, the semantic relevance of queries (SRQ) using Latent Semantic Analysis (LSA) was computed [13]. LSA is mathematical technique that provides a vector representation for each word (or text) in a high dimensional semantic space, and provides a similarity measure between any two words in that space using the cosine between the vectors [14]. That space is in our case based on a corpus of 70,000 Dutch documents (consisting of 60% newspaper articles and 40% medical and health related articles) used to create a semantic space in Dutch, applying Gallito [15]. The LSA metric estimates how close in semantic similarity the queries generated by the participants are to the target information by computing the



**Figure 1:** Mean semantic relevance of queries (SRQ) for each reformulation cycle for (a) simple and (b) difficult tasks. The first cycle corresponds to the 1st query for a specific task, and the 2nd cycle stands for the 2nd reformulated query in that task and so forth.

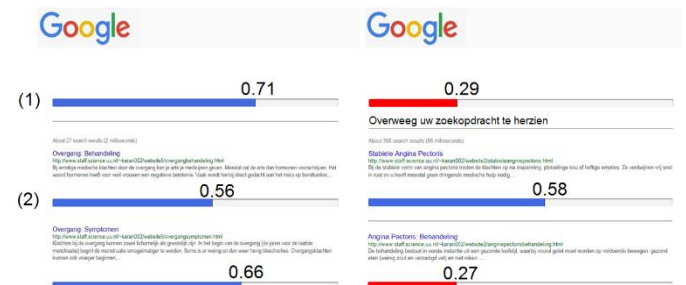
cosine value between vectors of the query and target information, respectively. A high LSA value indicates a high semantic overlap between the query and the target information. Therefore, in general, the higher the SRQ is, the more relevant the query is, as demonstrated in [7, 12]. It appeared in these studies that the semantic relevance (SRQ) generated by younger adults remained constant across reformulations, or even increased, while that of the queries generated by older adults showed a decreasing trend as they reformulated further [7, 12]. That was particularly the case with difficult tasks (Figure 1). Difficult tasks were in these studies tasks that could not directly be answered by looking at the search results on the SERP, but needed some further examination

of search results and reasoning. One example to illustrate a sequence of search terms with increasing LSA values, given a difficult task (translated from Dutch) such as “*Fieke, 6 year old, wants to drink a lot of water and has to urinate frequently. Often very exhausted. The physician diagnoses a high glucose value. What could be the problem, be specific, and what treatment would the physician start?*” A participant issued as first search terms “*child of 6, thirsty, urinates frequently. What could be the problem?*” (LSA = .11), 2nd cycle of search terms “*6 year old, exhausted, high glucose value. What could be the problem?*” (LSA = .57) and finally as third series of search terms “*Type 1 diabetes of 6 years old, treatment?*” (LSA = .95). The correct answer was here Diabetes Type 1 and treatment with insulin.

### 3 STUDY ON SEMANTIC RELEVANCE FEEDBACK

#### 3.1 Visual Feedback on Semantic Relevance

Based on the above studies and outcomes, we performed some studies with a new and modified interface for the SERPs. This new interface visually displays the semantic relevance of a query and search results (Figure 2), in contrast to Scentbar [11]. Having available this information could be helpful in generating new query. As we saw above particularly older participants have difficulties especially with difficult tasks in progress of the semantic relevance of a new query. A reason could be that they do not have information available to evaluate easily the relevance of a query or search results. The central idea of current study is to keep users informed about the relevance of their queries and search results, which might help in formulating better new queries. We hypothesize that monitoring the relevance of a query and warning when it falls below a threshold also would help when the search results are monitored. Two LSA values are examined: a lenient criterion, i.e. the LSA threshold was set on 0.3 and a strict one, i.e. the threshold at 0.5, that is, if the LSA values are below these values, participants were asked to reformulate their query, so many more search results will become red when the strict criterion is applied. Because the LSA threshold values vary between 0 and 1, they have to be predefined and we will examine the optimal value for two values empirically. Based on previous research [7] we took the mean LSA value of SRQ of successful tasks and not-successful tasks resp., and that provided the values of 0.5 and 0.3 used in current study.



**Figure 2: The feedback on the interface: (1) is the value corresponding to the search query and (2) to the relevance values of the search results. In the right image, the bars become red after they fall below the pre-defined threshold value of 0.3.**

Note that both support aspects (feedback on relevance of query and search results) are simultaneously examined.

### 3.2 Method

*Participants* were randomly assigned to conditions, and individually tested at home. In a first study we tested 48 university students (mean age 23 years), in a second study 48 older adults without clear cognitive or physical problems (mean age 72).

*Materials.* We used 12 simulated information search tasks [16] from the domain of health, divided into 6 low precise (= difficult) tasks and 6 high precise (=easy) tasks. Task preciseness is based on the semantic similarity between the task description and the content of the target page(s) containing the answer to the task from the (5) available, realistic mockup websites (ranging 37-194 pages) with Dutch medical and health information. We checked with LSA the difference in semantic similarity. Tasks with a high LSA value of task preciseness provide accurate contextual information directly pointing to the target information, while low precise tasks require users to engage in higher-level cognitive activities. Compared to the previous difficulty index, this measure provides an automatic, objective and validated way of calculating overlap in meaning [7, 12].

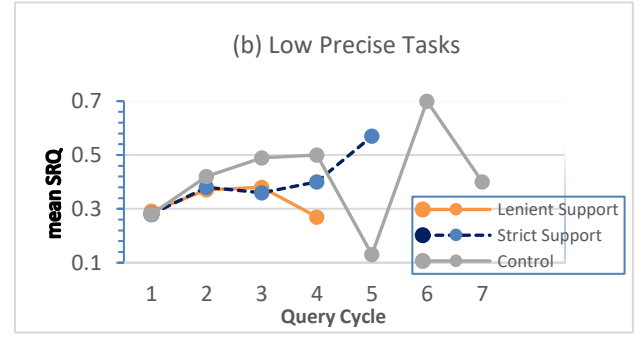
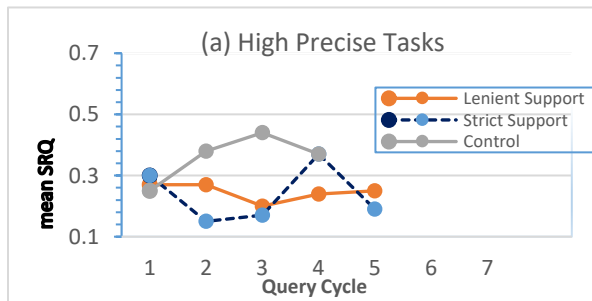
*Design & Procedure.* A mixed design was used, with control, lenient and strict support condition as between-subjects variable and task preciseness as within-subjects variable. Presentation order of tasks was counter-balanced, and time limit per task was 8 min.

### 3.3 Results

We will discuss the impact of feedback on the reformulations per cycle of young and older adults. In Figure 3 and 4 the mean semantic relevance values of generated queries are depicted for Younger Adults and Older Adults, respectively, for each reformulation cycle.

The main statistical significant effect with Younger Adults was that for low precise tasks, there is an increasing linear trend in the strict support condition ( $F(4,55)=5.98, p<.05$ ). Other effects were not significant. See Figure 3.

Study 1: Younger Adults

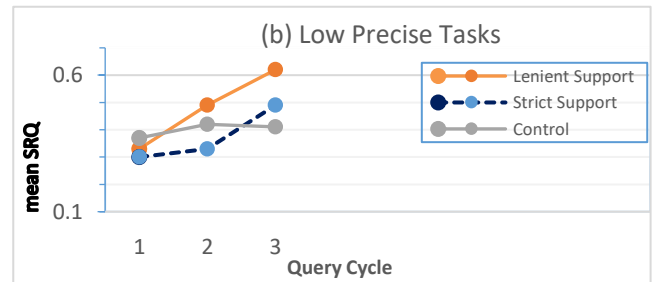
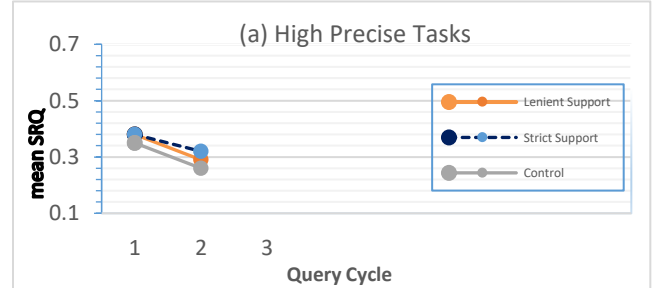


**Figure 3: Mean semantic relevance of queries at each reformulation cycle for (a) high precise and (b) low precise tasks for Younger Adults.**

Regarding the Older Adults (Study 2) we found again only for low precise tasks significant effects: for the lenient support condition as well as for the strict support condition, there is a significant increasing linear trend ( $F(2,31)=10.44, p<.01$  and  $F(2,33)=5.02, p<.05$  resp.). See Figure 4.

These patterns of younger and older adults indicate, thus, that as participants reformulated, they produced queries whose semantic relevance was significantly increasing, at least for difficult tasks when they were supported. For both age groups a strict LSA criterion appeared to be effective.

Study 2: Older Adults



**Figure 4: Mean semantic relevance of queries at each reformulation cycle for (a) high precise and (b) low precise tasks for Older Adults.**

## 4 CONCLUSIONS

For both older adults and younger adults an increasing trend was found in semantic relevance of the queries when they received feedback on difficult tasks. We assume that providing feedback on the semantic relevance of a search query and the search results, enables participants to make an informed decision to reformulate or not, as well as on the semantic direction of the reformulation. When looking at search performance itself (not reported here), that is, the time needed to solve the task and the number of clicks needed, we found that younger adults who got the feedback were also faster in their search and clicked significantly less, while for older adults there was no such difference in search performance. Therefore, even though the feedback does indeed help both groups to formulate better queries, older adults probably need more time and practice to adjust to the new search interface before they improve their search performance, whereas younger adults easily adapt to the new interface and thus we do see with them an improvement in their search performance.

## 5 LIMITATIONS

In current study both means of support (feedback on the relevance of the query and relevance of search results) were presented simultaneously, so it is impossible based on the current data, to make a difference in effectiveness of the two support means. Further empirical research has to clarify this issue.

One of the main limitations of our work so far is that the semantic relevance of queries measurement can only be used for those types of tasks for which there is a known target answer page(s). It is necessary to know the target page(s) in advance to be able to compute the semantic relevance. Though this limits the applicability of the metric in real environments where neither the user intent nor the target answer are known beforehand, it can be very useful in providing training and support to users with low information search skills such as older adults. We suggest to use the task description as an alternative to the target answer, in real environments, at least in those situations where a clear task description is available. Whether the semantic similarity values computed using the task description (instead of the target answer) give the same outcomes or not as in this study, has to be empirically verified.

Finally, further studies could explore the comparative effectiveness of different ways of visualizing the feedback or reordering of search results, but even before such work to optimize the interface, the results of this work show that providing feedback on, and making users aware of, semantic relevance fosters formulation of better queries.

## ACKNOWLEDGMENTS

We thank Saras Karanam for his valuable input to this study. This research was supported by the Netherlands Organization for Scientific Research (NWO) Open Research Area Plus project MISSION (464-13-043).

## REFERENCES

- [1] Wildemuth, B., Freund, L., & Toms, E.G. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*, 70(6), 1118-1140.
- [2] Chin, J., Anderson, E., Chin, C., & Fu, W.-T. (2015). Age differences in information search: An exploration-exploitation tradeoff model. In *Proceedings of the Human Factors and Ergonomic Society (HFES 2015)*, pp. 85-89.
- [3] Dommès, A., Chevalier, A., & Liao, S. (2011). The role of cognitive flexibility and vocabulary abilities of younger and older users in searching for information on the web. *Applied Cognitive Psychology*, 25, 5, 717-726.
- [4] Pak, R., & Price, M. (2008). Designing an information search interface for younger and older adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50, 4, 614-628.
- [5] Queen, T.L., Hess, T.M., Ennis, G.E., Dowd, K., & Gruhn, D. (2012). Information search and decision making: Effects of age and complexity on strategy use. *Psychology and Aging*, 27(4), 817-824.
- [6] Chevalier, A., Dommès, A., & Marquié, J.C. (2015). Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, 53, 305-315.
- [7] Karanam, S., & van Oostendorp, H. (2016). Age-related differences in the content of search queries when reformulating. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5720-5730). ACM.
- [8] Gomez-Nieto, E., San Roman, F., Pagliosa, P., Casaca, W., Helou, E. S., de Oliveira, M. C. F., & Nonato, L. G. (2014). Similarity preserving snippet-based visualization of web search results. *IEEE transactions on visualization and computer graphics*, 20(3), 457-47.
- [9] Peltonen, J., Belorustceva, K., & Ruotsalo, T. (2017). Topic-Relevance Map: Visualization for Improving Search Result Comprehension. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 611-622). ACM.
- [10] Iwata, M., Sakai, T., Yamamoto, T., Chen, Y., Liu, Y., Wen, J. R., & Nishio, S. (2012). AspecTiles: Tile-based visualization of diversified web search results. In *Proceedings of the 35th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 85-94). ACM.
- [11] Umamoto, K., Yamamoto, T., & Tanaka, K. (2016). ScentBar: A Query Suggestion Interface Visualizing the Amount of Missed Relevant Information for Intrinsically Diverse Search. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 405-414). New York: ACM.
- [12] Karanam, S., & van Oostendorp, H. (2017). Age-related effects of task difficulty on the semantic relevance of query reformulations. In *Proceedings of the 16th IFIP TC 13 International Conference on Human-Computer Interaction INTERACT 2017*.
- [13] Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- [14] McNamara, D.S., Cai, Z., & Louwerse, M.M. (2007). Optimizing LSA measures of cohesion. In Landauer, T.K., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.), *Latent Semantic Analysis* (pp. 379-400). Mahwah: Erlbaum Ass.
- [15] Jorge-Botana, G., Olmos, R., & Barroso, A. (2013). Gallito 2.0: A natural language processing tool to support research on discourse. In *Proceedings of the 13th Annual Meeting of the Society for Text and Discourse*. Valencia, Spain.
- [16] Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), 225-250.