



Open Research Online

Citation

Ghafourian, Yasin; Hanbury, Allan and Knoth, Petr (2023). Ranking for Learning: Studying Users' Perceptions of Relevance, Understandability, and Engagement. In: Linking Theory and Practice of Digital Libraries. TPDL 2023. (Alonso, Omar; Cousijn, Helena; Silvello, Gianmaria; Marrero, Mónica; Teixeira Lopes, Carla and Marchesin, Stefano eds.), Lecture Notes in Computer Science (LNCS) volume 14241, Springer, Cham, pp. 284–291.

URL

<https://oro.open.ac.uk/94410/>

License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

Ranking for Learning: Studying Users' Perceptions of Relevance, Understandability, and Engagement

Yasin Ghafourian^{1,2}[0000-0001-9683-9748], Allan Hanbury²[0000-0002-7149-5843],
and Petr Knoth³[0000-0003-1161-7359]

¹ The Research Studios Austria FG, Vienna 1090, Austria

² Technische Universität Wien, Vienna 1040, Austria

³ Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom

Abstract. General-purpose search engines are frequently used to retrieve content for learning. However, their ranking strategies are typically optimised for relevance, which means that they do not take into account other criteria important in the learning context, such as the understandability and the degree of engagement of the retrieved resources. We have conducted a user study to assess the extent to which ranking algorithms used by a popular search engine satisfy the expectations of users who are learning by searching. We study the relationships between users' perceptions of topical relevance, engagement, and understandability for retrieved documents with respect to their ranks. While we observe that the perceived user-assigned rank is strongly associated with all dimensions of relevance under study, specifically engagement ($\rho = 0.89$), understandability ($\rho = 0.58$) and topical relevance ($\rho = 0.88$), the relationship between SERP ranks and user-assigned ranks appears unstable, indicating that learners are not necessarily always served well by general-purpose search engines.

Keywords: Information Retrieval · Search as Learning · Background Knowledge · Relevance · Understandability · Engagement

1 Introduction

Search engines have become vital tools for students and researchers [3, 17, 2], offering access to vast academic libraries and up-to-date research, and there is a growing trend of using web searches as a means of acquiring new knowledge [10, 6]. Search engine users interested in learning may have diverse levels of background knowledge in a topic. However, search engines overlook this variation and assume that users' queries adequately represent their information needs [14]. The reason is that search engines are optimised for topical relevance even though relevance is a multidimensional concept consisting of other perspectives such as understandability, novelty, utility, reliability, etc. as well [8, 25, 19, 16]. Additionally, user context such as the user's knowledge state, perceptions, etc., also influence how users interpret the relevance of available online resources [15].

If search engines considered users' background knowledge, they could provide learning resources that match their needs, saving time and effort by avoiding lengthy searches and useless documents. The relationship between user contexts

related to knowledge, such as knowledge gain [10, 26], knowledge state [27], and the evaluation of information retrieval (IR) systems considering users' cognition and knowledge [11, 4, 9, 24] have been examined from different perspectives in the literature. Previous studies have also highlighted the importance of the users' knowledge dimension [1, 5] in creating and evaluating information needs, its effect on user behaviour, and the choice of web pages to read [12, 7, 23].

In this paper, we are building on top of the existing research by conducting an exploratory study to investigate in more detail how users' background knowledge affects their perception of relevance along three dimensions of relevance in the context of learning. We seek to answer the following research question:

Research Question: To what degree do users find the specific order of web pages provided on a Search Engine's Results page (SERP) suitable for learning?

Our methodology is based on a survey design and proceeds by first collecting data focused on participants' preferences for learning about a topic based on a set of web pages from the topic. We asked the participants to re-rank the web pages and at the same time label them along 3 dimensions of relevance: 1) topical relevance, 2) understandability, and 3) engagement.

To the best of our knowledge, our research is the first study that employs explicit individual re-rankings and relevance feedback along three dimensions to look into how users with varying levels of topic-specific knowledge exhibit differences in their perceived understandability, relevance, and engagement towards SERP web pages and how these differences are reflected in different ranking preferences compared to an algorithm's ranking.

The rest of the paper is structured as follows: Section 2 outlines our methodology, including a description of how our study is formulated, and how the data is collected. In Section 3, we present our findings after processing and analysing the collected data. Section 4 reflects on the study and concludes the paper.

2 Methodology

Our methodology employs a survey design to gather explicit relevance feedback from online participants' preferences for learning about a specific topic. To ensure the success and accuracy of our experiments, we had initiated a pilot study with a group of 14 Ph.D. students to identify any areas that required revision and confirm the effectiveness of our research methods. Subsequently, we launched the revised survey with a larger sample of participants.

We selected a set of four topics for our research, and for each topic, following the established method in the literature [10, 26, 20, 6], we prepared a multichoice question knowledge test with 10 questions using available online quizzes in a survey to assess the participants' current knowledge on that topic (Urgo et al. [22] summarises various types of knowledge assessment used in the literature). Before the quiz, participants were also asked to self-assess their topical knowledge on a 5-point Likert scale. The topics have been chosen in a way that maintains a balance between being specific enough to be informative and not too narrow to limit their relevance to the general population. The topics are: 1) Covid-19, 2) World War 2, 3) Financial Literacy, and 4) Theory of General Relativity.

Subsequently, we sampled 10 web pages from the web for each of the topics. The participants were then instructed to re-rank the given web pages in descending order of how suitable they found the web pages for learning about the topic. Pages were presented as clickable links with randomized order for each user to mitigate user rank selection bias [18] (AKA trust bias [13]). Simultaneously, they were asked to provide three labels for each web page on a 5/7 point Likert scale: 1) topical relevance, 2) understandability, and 3) level of engagement offered by the web page meaning its motivational value for learning about the topic. The decision to limit the number of web pages to 10 in our study was based on insights gained from the pilot study, as it was found to be the largest number of articles that users could effectively sort within the available time.

We recruited participants from Prolific⁴, specifically selecting those with a high level of English proficiency from continental Europe and the UK. A total of 207 participants were included, each topic having 50-56 exclusive participants.

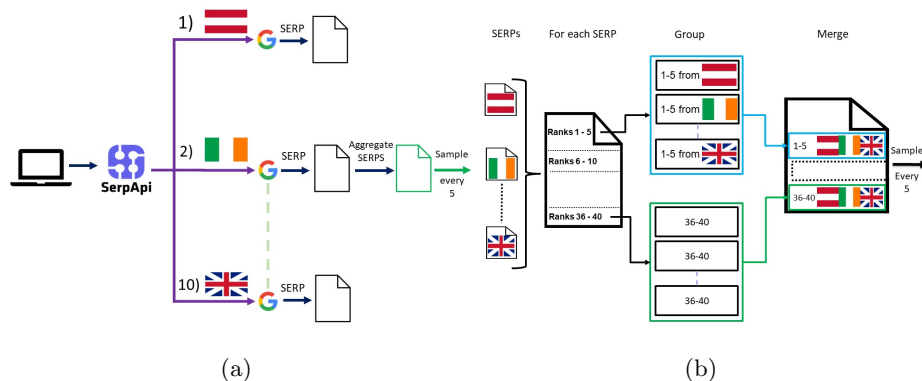


Fig. 1: An overview of our sampling approach to collect web pages for re-ranking and labeling. (a) Collecting SERPs from different locations. (b) Merging SERPs.

To find topic-related web pages, we used keywords and phrases from knowledge tests to form a representative query for the topic. We selected the top 10 phrases and keywords with the highest TF-IDF score [21] calculated against Wikipedia dump⁵ as a background corpus. To sample web pages, we used Google as the search engine and SerpAPI⁶ as the tool to retrieve results in a machine-readable format (In our case XML). To ensure consistent search engine results pages (SERPs) among survey participants, we submitted each query 10 times at 10-minute intervals from various locations in our participant pool. The 10 SERPs obtained per query were merged on the page level with 5 links per page. Finally, a webpage was sampled from each page on the merged SERP, resulting in a total of 10 web pages. Higher weighting was given to pages appearing in multiple SERPs and having higher ranks. We excluded links to books, PDFs,

⁴ <https://www.prolific.co/>

⁵ <https://dumps.wikimedia.org/enwiki/20221220>

⁶ <https://serpapi.com>

and video links to focus on multimedia web pages consisting primarily of textual content. The sampling approach is depicted in Figure 1.

3 Results

3.1 Participants’ Demographics, Self-declared vs Demonstrated Knowledge

Table 1 provides an overview of the participants who participated in each of the four topics, including their demographic distribution, the average declared knowledge on the topic, and the average score obtained after taking the knowledge test. The participants were well-balanced in terms of their gender across the four topics. Most participants were in the age group 25-44.

Table 1: An overview of participants’ demographics and characteristics. The average declared knowledge is reported using a 5-point Likert scale and the attained average knowledge test scores are mapped to the same scale to allow comparison.

Topic Name	Number of Participants	Time Spent on the Survey (Minutes)		Gender Distribution			Age Distribution						Average Declared Knowledge (1-5)	Average Knowledge Test Score	Difference between Declared Knowledge Score and Test Score
		Mean	Standard Deviation	Female	Male	Other	18-24	25-34	35-44	45-54	55+				
World War 2	56	21.25	10.37	44%	56%	0%	5%	51%	22%	15%	7%	3.4	3.48 (62%)	0.08	
Financial Literacy	51	22.10	10.12	45%	55%	0%	12%	27%	37%	20%	4%	3.1	3.20 (55%)	0.10	
Covid-19	50	18.8	7.41	58%	40%	2%	12%	52%	18%	14%	4%	3.82	2.84 (46%)	0.98	
Theory of General Relativity	50	26.62	14.12	44%	54%	2%	4%	38%	28%	24%	6%	2.26	2.32 (33%)	0.06	

3.2 SERP Rank vs User-assigned Rank

We evaluated the alignment between user-assigned ranks and SERP ranks by comparing the average assigned rank for each web page with its corresponding SERP rank. The relationship between the two ranks varied significantly across the four topics (Figure 2). We found low-moderate positive correlations for two topics where users were more knowledgeable (WW2 & Finance) but strong negative and no correlations for the other two topics. This result was surprising, and it suggests two possible explanations: (1) Search engines may not reliably retrieve the most suitable documents for learners’ needs across different topics. (2) Learners with limited knowledge of a topic may struggle to evaluate and rank documents appropriately. Both phenomena could contribute to this outcome.

3.3 Engagement, Topical Relevance, and Understandability as a function of User-assigned Rank

We conducted an analysis of the Likert-scale labels that the participants assigned to each of the web pages for topical relevance, engagement, and understandability. We wanted to understand to what extent they are associated with the user-assigned ranks. This is important as it will help us to understand the extent to which these dimensions of relevance should be taken into account when building, and potentially personalising, search engine for learners.

As these labels were assessed using a Likert scale, we converted the indicated perceptions into numeric values. To facilitate a comparison between ranks and

these labels, we converted the Likert scale values so that a lower value indicated a higher preference. For example, “Very Engaging” in the engagement label was assigned the value of 1, while “Not Engaging at all” received the value of 5. Figure 3, reports on the relationship between the average user-assigned rank of

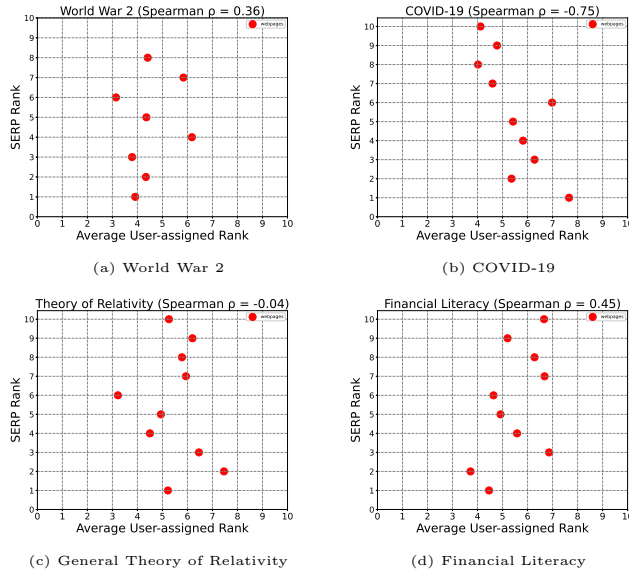


Fig. 2: Comparative scatter plots of web pages’ actual rank vs their average user-assigned rank across the four analysed topics.

the web pages and the average-assigned value for engagement, topical relevance, and understandability. We observe that user-assigned rank is strongly correlated with engagement ($\rho = 0.89$) and topical relevance ($\rho = 0.88$) while also medium-strongly correlated with understandability ($\rho = 0.58$). This confirms that both engagement and understandability, are two characteristics of retrieved content that are important to learners as well as topical relevance.

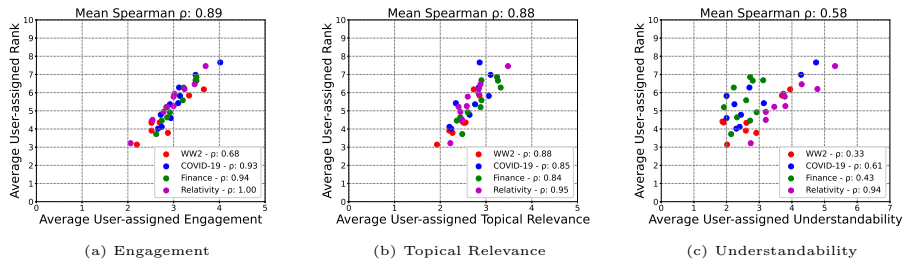


Fig. 3: Spearman correlations between average rankings and average values of (a) engagement, (b) relevance, and (c) understandability of web pages across four topics.

Furthermore, we noticed that in topics “World War 2” and “Financial Literacy”, the average user-assigned ranks have a weaker correlation with average user-assigned understandability value compared to engagement and topical relevance. This observation suggests that in topics where our users were more knowledgeable, the understandability of the texts played a relatively less important role compared to perceived topical relevance and engagement. Additionally, we noted a higher correlation between the average assigned rank and the perceived understandability of web pages for the two topics where users possessed a lower average knowledge level, namely “General Theory of Relativity” and “COVID-19,” compared to the other two topics. This implies that for topics that users know less about, the understandability of the web pages becomes a more crucial factor to be considered by search engines. Reflecting on the alignments observed between user ranks and SERP ranks across all topics, our data suggest that the original SERP ranking may not have adequately accounted for the understandability of web pages, as an increased emphasis on understandability in users’ specified rankings leads to decreased agreement with the SERP ranking. The findings of this analysis highlight the role of perceived engagement, understandability, and topical relevance as factors that determine users’ preferred rankings.

4 Discussion and Conclusion

In this paper, we explored in more detail how search engine users perceive the relevance of the SERP results rankings along different dimensions when viewing the results for learning purposes. We have conducted a user study in which we asked online users to re-rank a set of web pages from SERP and asked them to provide labels for these web pages along three dimensions of relevance as well.

Our work has a notable strength in its unique approach of organizing web pages for all participants within each topic, ensuring comparability of rankings and allowing focus on variables like background knowledge and perceived relevance. We also obtained explicit feedback from users, facilitating informed conclusions. However, a limitation is that users couldn’t participate in multiple topics, limiting result comparability and exploration of knowledge influence. Additionally, while we intentionally included a range of topics to have diversity, the topics also vary in their complexity, making generalisation across topics challenging. Furthermore, our evaluation relied on a limited set of questions, potentially not capturing users’ full knowledge depth and breadth.

Our analysis showed that the relationship between the users’ specified ranks and the SERP ranks ranges from low-moderate to strongly correlated, suggesting that delivered search engine ranks were found on average unsuitable by users for learning. Moreover, our findings showed that along with topical relevance, engagement, and understandability are also other aspects of the web pages that are significant to learners having moderate-strong correlations with user ranks.

Acknowledgements This work was supported by the EU Horizon 2020 ITN/ETN on Domain Specific Systems for Information Extraction and Retrieval – DoSSIER (H2020- EU.1.3.1., ID: 860721).

References

1. Anderson, L.W., Bloom, B.S., et al.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, (2001)
2. Apuke, O.D., Iyendo, T.O.: University students' usage of the internet resources for research and learning: forms of access and perceptions of utility. *Heliyon* **4**(12), e01052 (2018)
3. Aroyo, L., Denaux, R., Dimitrova, V., Pye, M.: Interactive ontology-based user knowledge acquisition: A case study. In: European Semantic Web Conference. pp. 560–574. Springer (2006)
4. Bailey, P., Chen, L., Grosenick, S., Jiang, L., Li, Y., Reinholdtsen, P., Salada, C., Wang, H., Wong, S.: User task understanding: a web search engine perspective. In: NII shonan meeting on whole-session evaluation of interactive information retrieval systems, Kanagawa, Japan (2012)
5. Bloom, B.: Taxonomy of educational objectives, handbook i: The cognitive domain. new york: David mckay co inc. as cited in file. D:/bloom. html (1956)
6. Cãmara, A., Roy, N., Maxwell, D., Hauff, C.: Searching to learn with instructional scaffolding. In: Proceedings of the 2021 conference on human information interaction and retrieval. pp. 209–218 (2021)
7. Cole, M.J., Zhang, X., Liu, C., Belkin, N.J., Gwizdka, J.: Knowledge effects on document selection in search results pages. In: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval. pp. 1219–1220 (2011)
8. Cosijn, E., Ingwersen, P.: Dimensions of relevance. *Information Processing & Management* **36**(4), 533–550 (2000)
9. El Zein, D., da Costa Pereira, C.: User's knowledge and information needs in information retrieval evaluation. In: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization. pp. 170–178 (2022)
10. Gadiraju, U., Yu, R., Dietze, S., Holtz, P.: Analyzing knowledge gain of users in informational search sessions on the web. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. pp. 2–11 (2018)
11. Ghafourian, Y., Knoth, P., Hanbury, A.: Information retrieval evaluation in knowledge acquisition tasks. Proceedings of the Third Workshop on Evaluation of Personalisation in Information Retrieval (WEPIR) @ CHIIR pp. 88–95 (2021)
12. Jansen, B.J., Booth, D., Smith, B.: Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management* **45**(6), 643–663 (2009)
13. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., Gay, G.: Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)* **25**(2), 7–es (2007)
14. Kelly, D., Belkin, N.J.: A user modeling system for personalized interaction and tailored retrieval in interactive ir. Proceedings of the American Society for Information Science and Technology **39**(1), 316–325 (2002)
15. Liu, J., Liu, C., Belkin, N.J.: Personalization in text information retrieval: A survey. *Journal of the Association for Information Science and Technology* **71**(3), 349–369 (2020)
16. Mao, J., Liu, Y., Zhou, K., Nie, J.Y., Song, J., Zhang, M., Ma, S., Sun, J., Luo, H.: When does relevance mean usefulness and user satisfaction in web search? In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 463–472 (2016)

17. Marchionini, G.: Exploratory search: from finding to understanding. *Communications of the ACM* **49**(4), 41–46 (2006)
18. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L.: In google we trust: Users' decisions on rank, position, and relevance. *Journal of computer-mediated communication* **12**(3), 801–823 (2007)
19. Pasi, G.: Contextual search: issues and challenges. In: *Symposium of the Austrian HCI and Usability Engineering Group*. pp. 23–30. Springer (2011)
20. Roy, N., Moraes, F., Hauff, C.: Exploring users' learning gains within search sessions. In: *Proceedings of the 2020 conference on human information interaction and retrieval*. pp. 432–436 (2020)
21. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Information processing & management* **24**(5), 513–523 (1988)
22. Urgo, K., Arguello, J.: Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management* **59**(2), 102821 (2022)
23. Vakkari, P.: Searching as learning: A systematization based on literature. *Journal of Information Science* **42**(1), 7–18 (2016)
24. Vakkari, P., Völske, M., Potthast, M., Hagen, M., Stein, B.: Modeling the usefulness of search results as measured by information use. *Information Processing & Management* **56**(3), 879–894 (2019)
25. Xu, Y., Chen, Z.: Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology* **57**(7), 961–973 (2006)
26. Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P., Dietze, S.: Predicting user knowledge gain in informational search sessions. In: *The 41st international ACM SIGIR conference on research & development in information retrieval*. pp. 75–84 (2018)
27. Yu, R., Tang, R., Rokicki, M., Gadiraju, U., Dietze, S.: Topic-independent modeling of user knowledge in informational search sessions. *Information Retrieval Journal* **24**, 240–268 (2021)