Journal of Technology Research (JTR)
Volume 4, Issue 1, (2026), pp 001-011, ISSN 3005-639X

Behavior-Driven Semantic Re-Ranking for Personalized Web Search

Anwar Alhenshiri [1], Alaa Jelwal [*2] Hoda Badesh [*3]

[1] Department of Computer Science, Faculty of Information Technology, Misurata University, Misurata, Libya
*Corresponding author email: alhenshiri@it.misuratau.edu.ly

Received: 03-01-2025 | Accepted: 02-03-2026 | Available online: 13-03-2026| DOI:10.26629/jtr.2026.**

ABSTRACT

This study proposes a behaviour-driven semantic re-ranking system designed to enhance personalized web search without modifying the underlying search engine infrastructure. The approach operates as a lightweight middleware layer over Google Search and integrates BERT-based semantic embeddings, KeyBERT-driven dynamic user profiling, cosine similarity scoring, and an exponential decay mechanism to maintain adaptive interest modelling. A controlled comparative experiment involving 14 participants was conducted to evaluate performance against the baseline Google ranking. Results indicate that the proposed system reduced average search time by 15.3% (from 274 seconds to 232 seconds), decreased the number of required clicks by 21.2% (from 7.71 to 6.07), and lowered query reformulations from 43% to 21%. Furthermore, 79% of participants preferred the personalized system, with statistically significant improvements observed in efficiency and satisfaction ($p < 0.05$). These findings demonstrate that behaviour-driven semantic re-ranking can substantially improve search relevance, user efficiency, and overall experience while maintaining transparency and scalable deployment.

Keywords: Web Search, Intelligent Search, Web User Task, Search Ranking

إعادة ترتيب دلالي مدفوع بالسلوك للبحث المخصص على الويب

أنور الهنشيري[1]، آلاء جلوال[1] ، هدى بادش[1]

[1] قسم علوم الحاسوب ، كلية تقنية المعلومات ، جامعة مصراته ، مصراته، ليبيا

ملخـــــــص البحـــــــث

تقترح هذه الدراسة نظامًا لإعادة ترتيب نتائج البحث الدلالية يعتمد على سلوك المستخدم بهدف تحسين البحث الشخصي على الويب دون تعديل البنية التحتية لمحرك البحث الأساسي. يعمل النظام كطبقة وسيطة خفيفة فوق **Google Search**، ويعتمد على التمثيلات الدلالية باستخدام نموذج **BERT**، وبناء ملف تعريف ديناميكي لاهتمامات المستخدم باستخدام **KeyBERT**، إضافة إلى حساب درجة التشابه باستخدام **Cosine Similarity** وآلية اضمحلال أسي للحفاظ على نموذج اهتمامات متكيف مع مرور الوقت. أُجريت تجربة مقارنة مُتحكَّم بها بمشاركة 14 مستخدمًا لتقييم أداء النظام مقارنة بترتيب **Google** الأساسي. أظهرت النتائج تحسنًا ملحوظًا في الكفاءة، حيث انخفض متوسط وقت البحث بنسبة 15.3% (من 274 إلى 232 ثانية)، وتراجع عدد النقرات المطلوبة بنسبة 21.2% (من 7.71 إلى 6.07)، كما انخفضت إعادة صياغة الاستعلامات من 43% إلى 21%. كما فضّل 79% من المشاركين النظام المخصص. تشير النتائج إلى أن إعادة الترتيب الدلالي المعتمد على السلوك يمكن أن تعزز صلة النتائج وتحسن تجربة البحث مع الحفاظ على الشفافية وقابلية التوسع.

الكلمات الدالة: البحث على الويب ، البحث الذكي، مهام المستخدم، ترتيب النتائج، استرجاع المعلومات.

1. INTRODUCTION

The internet has evolved into a vast and dynamic information repository, hosting billions of web pages across every conceivable domain. As this digital landscape continues to grow exponentially, users face an increasing challenge: how to efficiently locate relevant and trustworthy information. Search engines such as Google and Bing have become the primary gateways for accessing this information, relying on sophisticated ranking algorithms that evaluate factors like keyword matching, page authority, content freshness, and backlink structures [6, 7]. While effective in many cases, these traditional algorithms remain limited by their generic, non-personalized nature.

One of the most pressing issues in modern information retrieval is the inability of traditional search engines to adapt to individual user preferences and contexts. These systems typically serve the same ranked list of results to all users issuing the same query, regardless of their distinct backgrounds, goals, or prior knowledge. This lack of personalization results in search experiences that often require users to reformulate queries, sift through irrelevant results, and invest additional time and effort to locate suitable content. This problem is particularly evident in ambiguous or broad queries, where user intent is not explicitly stated and must be inferred [5].

To bridge this gap, researchers have explored personalized search systems that learn from user behavior—such as previous search queries, clicked links, and browsing patterns—to refine and reorder search results dynamically. Notably, the integration of re-ranking mechanisms has emerged as a promising strategy for improving result relevance without altering the core search engine. These systems work by intercepting the top results returned by a search engine and reordering them based on user-specific criteria, effectively augmenting rather than replacing the existing infrastructure [13, 14].

Among the various personalization strategies, content-based filtering has shown particular promise. This technique analyzes the characteristics of content with which a user previously interacted and recommends new content that shares similar attributes [8, 9, 10]. By creating and maintaining a user profile, typically as a weighted list of keywords or topics, the system can compare incoming search results to this profile and prioritize those that are semantically aligned with the user's interests. Collaborative filtering [11] and hybrid filtering models [12] further enhance personalization by combining content-based analysis with patterns derived from similar users.

Recent advances in Natural Language Processing (NLP) have further boosted the capabilities of personalized systems. Deep learning models like BERT (Bidirectional Encoder Representations from Transformers) enable semantic text understanding at a much deeper level, allowing for more accurate similarity measurements between user queries, document content, and profile information. These embeddings, when paired with similarity metrics like Cosine Similarity, form a powerful basis for re-ranking algorithms [12, 13, 15].

Although prior work has explored ontology-based, hybrid, and socially-informed personalization models, many existing systems require complex infrastructure changes, large-scale data integration, or browser-level modifications. Few approaches provide real-time personalization through a lightweight middleware layer that operates transparently on top of existing search engines without modifying their internal ranking mechanisms.

This research builds on these advances by developing a personalized re-ranking system that operates on top of Google Search. The system dynamically constructs a user profile using KeyBERT [15], extracts semantic features from search queries and results using BERT embeddings, and computes relevance scores that reflect both the current query and historical user interests. The results are then reordered accordingly to deliver a more tailored search experience.

Our work is inspired by several studies that have demonstrated the value of personalized re-ranking. For instance, the study in [1] proposed a dynamic model using ontologies and multi-agent systems to adapt result rankings to user preferences. In addition, the research in [2] introduced contextual adaptation via the DROPT algorithm, emphasizing the role of user context in information retrieval. More recent work by [3] and [5] highlighted the effectiveness of hybrid and socially-informed re-ranking systems.

In this study, we propose a practical, user-centric re-ranking system and evaluate its impact on search efficiency and satisfaction through a comparative experiment involving 14 participants. We analyze metrics such as search time, number of clicks, and perceived relevance of results to assess how well the personalized system meets user needs compared to the standard Google ranking.

This study proposes a behavior-driven semantic re-ranking framework that:

1. Operates as lightweight middleware over Google Search,
2. Uses dynamic BERT-based semantic embeddings,
3. Applies exponential decay for adaptive user modeling,
4. Demonstrates statistically validated improvements in search efficiency and satisfaction.

The results show that personalized re-ranking indicates promising improvements. Though, future large-scale validation (***N>40***) is required for broader generalization. Given the exploratory nature of the evaluation and the limited sample size, this study should be considered a pilot investigation aimed at assessing feasibility and identifying behavioral trends rather than establishing population-level generalizations.

2. MATERIALS AND METHODS

Personalized search and result re-ranking have been extensively studied over the past decade as effective approaches to mitigate the limitations of traditional search engines. While general-purpose search engines such as Google employ sophisticated algorithms to rank search results using hundreds of factors—including page relevance, authority, and freshness—they often lack the capability to adapt to individual user preferences and contextual intent [7].

2.1 Personalization through User Profiling

A common technique in personalizing web search involves the use of user profiles, which serve as digital representations of a user's interests, preferences, and behavior over time. One early yet influential contribution in this area came from [1].

who developed a dynamic re-ranking system that builds a user profile using semantic ontologies and a multi-agent framework. Their model analyzes historical queries and search behavior to infer user interests and aligns future results accordingly. The key innovation in this work was the use of ontological concepts to bridge lexical gaps and improve semantic understanding.

2.2 Context-Aware and Adaptive Search

Building on the need for semantic depth, the research in [2] introduced a context-aware re-ranking model using the DROPT algorithm, which integrates both user features and contextual parameters such as time and location to personalize search results dynamically.

Their work demonstrated that context, when combined with behavioral data, can significantly enhance search precision. The dynamic reweighting of content based on real-time factors aligned well with the shift toward adaptive search interfaces.

2.3 Hybrid and Multi-Signal Models

Recognizing the limitations of using a single information source, several studies proposed hybrid models that combine content-based and collaborative filtering methods. For instance, the work in [2] proposed a hybrid recommender system for academic search that analyzes both content similarity and browsing history to recommend relevant research papers. Their study demonstrated a 13% improvement in precision over baseline models [15], emphasizing the benefit of integrating multiple signals into the re-ranking process.

Similarly, the recommendation system proposed in [4] leveraged Local Evidence Densities (LED) to selectively adjust the user profile based on previous clicks and search sessions. Their findings suggested a 9% gain in ranking performance, further supporting the use of dynamic, evolving user models.

2.4 Social and Implicit Data Integration

Recent efforts have expanded personalization beyond the search engine itself, incorporating social behavior and implicit feedback. For example, the research in [5] presented a novel approach that mines user interactions on social media to identify shared interests and preferences, which are then used to personalize web search results. Their model used machine learning to infer patterns of similarity between users and significantly outperformed non-personalized baselines in search accuracy.

2.5 Clustering and Semantic Filtering

In line with the growing adoption of NLP in search systems, the study in [6] introduced the use of user profile clusters to improve personalized search. These profiles were built using dynamically extracted keyword clusters from browsing sessions and organized into thematic groups. This clustering enabled finer-grained understanding of user intent and improved ranking relevance.

2.6 Summary and Connection to This Research

The body of prior work clearly shows a progression from static keyword-based models to adaptive, semantically enriched systems that integrate both explicit (e.g., queries, ratings) and implicit (e.g., clicks, dwell time) user behavior. Several approaches have proven effective: ontology-based modeling, context adaptation, hybrid filtering, social signal integration, and deep learning-based embeddings. However, many of these solutions remain complex, resource-intensive, or limited to specific domains.

Building on these foundations, the present research proposes a lightweight, content-based re-ranking system that leverages BERT embeddings, cosine similarity, and KeyBERT-based keyword extraction to dynamically construct and update a user profile, and to reorder search results retrieved from Google. Unlike prior work, our system is designed to operate transparently on top of an existing search engine via API calls, offering real-time personalization without modifying the core engine. By focusing on dynamic user profiles and semantic similarity scoring, this study contributes a scalable and user-friendly solution to improving search relevance and experience.

3. METHODOLOGY

To evaluate the effectiveness of personalized re-ranking in improving the web search experience, we designed a comparative experimental study involving a prototype system developed for this purpose. This section describes the overall research methodology, including system design, user modeling techniques, re-ranking strategy, data collection procedures, and evaluation metrics. The main methodology adopted in the work on this research is depicted in Figure 1.

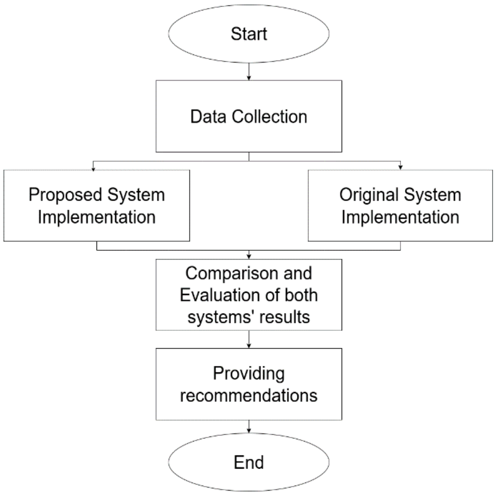

Fig 1. Research Methodology

3.1 Research Design

The research adopts a comparative experimental methodology to assess the difference in user experience between traditional search results retrieved directly from Google and the same results reordered using the proposed personalized re-ranking system. Fourteen participants were recruited to interact with both systems under identical search conditions. Each participant was asked to perform a series of predefined information-seeking tasks using both:

1. *Baseline system:* Google's original ranking (via Google Programmable Search Engine API).
2. *Experimental system:* The same Google results, re-ranked using the user's behavioral profile and semantic similarity techniques.

Both systems received identical initial result sets retrieved via the same Google API call to ensure fair comparison and eliminate variability in returned results.

To reduce bias and ensure consistency, all users searched for the same queries, with system order randomized to avoid sequence effects, and without knowing which system was Google and which was the personalized re-ranking system.

3.2 System Architecture Overview

The designed experimental system operates as a lightweight middleware layer on top of the Google Search API. The system architecture consists of the following components:

1. User Interaction Interface: A web-based frontend built using Flask, allowing users to log in, input search queries, and view results.
2. Query Dispatcher: Sends queries to the Google Custom Search API and retrieves the top N results (title, snippet, and link).
3. Re-ranking Engine. Measures semantic similarity between the query and each result, as well as between the user profile and each result, then reorders the results accordingly.
4. User Profile Manager: Dynamically constructs and updates the user profile using previous queries and clicked results.
5. Database Layer: Stores user accounts, search logs, click logs, and profiles using MySQL.

This modular design not only supports extensibility, allowing new features and components to be integrated with minimal changes to the existing system, but also facilitates detailed logging of user interactions. Such logging provides a structured record of user behavior, which can be systematically analyzed to evaluate system performance, measure user engagement, and identify opportunities for further improvement.

3.3 User Profiling and Interest Modeling

A core feature of the system is the construction of a dynamic user profile, which reflects the evolving interests of each user. The profile is represented as a JSON-based structure containing weighted keywords extracted from search queries entered by the user in addition to titles and snippets of search results the user clicks on. To extract keywords, the system uses KeyBERT, a keyword extraction model that leverages BERT embeddings to identify semantically meaningful phrases. Keywords are assigned weights based on frequency, recency, and context relevance.

To prioritize recent behavior, newly added keywords receive higher initial weights, ensuring that current interactions have a stronger influence on the user model. Meanwhile, the system employs an exponential decay function to progressively lower the weights of older keywords, enabling the profile to evolve over time and preventing it from becoming stagnant. The decay mechanism is formally defined as:

$$W_t = W_0 \cdot e^{-\lambda t}$$

Where:

- W_t is the updated keyword weight at time t,
- W_0 is the initial keyword weight,
- λ is the decay constant controlling the rate of forgetting,
- t is the elapsed time since the last interaction.

This formulation ensures that older interests gradually lose influence while preserving long-term behavioral patterns. The profile is updated in real time after each interaction and it plays a central role in the re-ranking process by capturing the user current interests.

For new users with no prior interaction history, the profile is initially empty and the system relies solely on query–result semantic similarity scoring. As behavioral data (queries and clicks) accumulate, the dynamic user profile is progressively constructed and begins influencing the re-ranking process. This ensures stable system performance during early interactions while enabling personalization over time.

While KeyBERT effectively extracts explicit semantic keywords from queries and clicked content, it may not fully capture latent or implicit user interests. Alternative approaches such as topic modeling (e.g., Latent Dirichlet Allocation or BERTopic) could potentially model deeper thematic structures and will be explored in future work.

3.4 Semantic Similarity and Re-Ranking

For each new query, the system computes a semantic similarity score between the user query and each search result and also the user profile and each search result. Both query and profile, along with each result (title + snippet), are converted into embedding vectors using the all-MiniLM-L6-v2 variant of the BERT model via the *SentenceTransformers* library.

The all-MiniLM-L6-v2 model was selected due to its favorable trade-off between computational efficiency and semantic performance. Compared to larger transformer models (e.g., all-mpnet-base-v2), it enables real-time embedding generation with lower latency and reduced resource consumption, making it suitable for lightweight middleware deployment.

Similarity is computed using cosine similarity, a metric that measures the angle between two vectors, ignoring their magnitude. The final relevance score for each result is calculated using the weighted formula:

$$\textbf{Final Score} = \alpha \cdot Sim(Q,R) + \beta \cdot Sim(P,R)$$

Where:

- Q: is the embedding of the query,
- P: is the embedding of the user profile,
- R: is the embedding of the search result,
- α and β are weights controlling the influence of the query vs. user profile (empirically set to $\alpha = 0.6$, $\beta = 0.4$).

The weighting parameters ($\alpha = 0.6$, $\beta = 0.4$) were empirically selected to prioritize query relevance while maintaining personalization influence. A preliminary sensitivity analysis showed that performance remained stable within ±0.1 variation of these parameters, indicating robustness to moderate weight adjustments. At the end, the results are sorted by the final score and presented to the user in a descending order of relevance.

3.5 Data Collection

The experiment was conducted over four days at the Faculty of Information Technology, Misurata University. Each of the 14 participants completed a series of search tasks using both systems. During this process, the following data were collected. All data were anonymized and stored securely.

1. Quantitative: the number of queries, number of result clicks, and the time to reach the desired task goal.
2. Qualitative: The user feedback through structured questionnaires on satisfaction, perceived relevance, and ease of use.
3. Behavioral: Logs of user interactions including queries, clicked links, and response times.

3.6 Participant Demographics

Table 1 summarizes the demographic characteristics of the **14** participants involved in the experimental evaluation. Reporting these attributes ensures transparency and reduces concerns about sampling bias.

Table 1. Demographics of Participants

Variable	Distribution
Total Participants	14
Age Range	20–28
Gender	9 Male, 5 Female
Academic Level	Undergraduate (10), Graduate (4)
Technical Proficiency	High (8), Moderate (6)

The participant pool consisted primarily of university students with moderate to high technical proficiency, ensuring familiarity with web search tools while maintaining diversity in search behavior.

3.7 Evaluation Metrics

The system was evaluated using both objective and subjective criteria. The metrics used included the time (in seconds) required to locate relevant information, the number of result clicks needed to reach a satisfactory answer (click efficiency), the frequency of query re-wording or clarification (query reformulation), the user satisfaction collected via questionnaires, and the perceived relevance. The latter was collected by asking users to rate the relevance of the top results. Statistical analysis was performed to compare user behavior between the two systems.

To control for the increased risk of Type I error due to multiple statistical comparisons, Bonferroni correction was applied. Given four independent hypothesis tests, the adjusted significance threshold was set to $\alpha = 0.0125$. Statistical significance was interpreted accordingly.

3.8 Limitations

While the study demonstrates promising results, several limitations should be acknowledged. First, the participant pool was relatively small (14 users), which may limit the statistical generalizability of the findings. However, this sample size is consistent with exploratory and usability-focused evaluations, where the objective is to assess feasibility and identify behavioral trends rather than achieve population-level inference. Second, the prototype currently supports only English-language queries, restricting applicability in multilingual contexts. Finally, the system depends on Google's Search API and is therefore subject to query limits and potential variability in API responses. Despite these constraints, the study provides a valid proof of concept and establishes a solid foundation for larger-scale and longitudinal evaluations.

3.9 Ethical Considerations

The study protocol was reviewed and approved by the Faculty Research Committee at the Faculty of Information Technology in Misurata University. All participants provided informed consent prior to participation. Data were anonymized and stored securely, and no personally identifiable information was retained after analysis.

4. RESULTS

This section presents the outcomes of the comparative experiment conducted to evaluate the effectiveness of the personalized re-ranking system against the traditional Google search interface. Both quantitative and qualitative results were analyzed to determine the extent to which the proposed system improved user experience, efficiency, and satisfaction.

4.1 Quantitative Analysis

4.1.1 Time to Information

One of the primary goals of the system was to reduce the time required for users to locate relevant information. The average time taken by participants to find satisfactory results was measured across both systems. In the case of the baseline system, the time was 274 seconds on average. However, it was only 232 seconds in the case of the experimental system. This represents a reduction in search time, indicating that personalized re-ranking enhances search efficiency by placing more relevant results higher in the list. The results are depicted in Figure 2.

The student t-test results show a significant difference between the two systems at the 0.05 level (***t* = 2.3, *p* < 0.02**). However, after Bonferroni adjustment (***α* = 0.0125**), this result did not remain statistically significant. The effect size, calculated using ***Cohen's d*** for paired samples, was ***d* = 0.61**, indicating a moderate practical effect

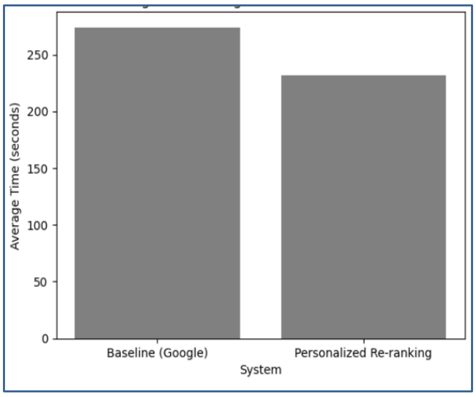

Fig. 2. Average Time to Information (Baseline vs. Personalized Re-ranking)

4.1.2 Number of Clicks

The number of clicks made before finding relevant information reflects the effort required by users to explore results. In the case of the baseline system, the number of clicks was 7.71 on average. It was only 6.07 clicks on average in the case of the experimental system. Hence, users needed fewer clicks with the personalized system, supporting the idea that the system helps surface more relevant content early in the result list. The student t-test results indicate a significant difference between the baseline system and the experimental system (***t*=3.03**, ***p*<0.005**). The effect size for click reduction was ***d* = 0.81**, indicating a large practical effect.

4.1.3 Query Reformulations

The number of query reformulations—a sign of dissatisfaction or ambiguity—was recorded. Users on the baseline system reformulated their queries in **43%** of the tasks. However, they only reformulated their queries in **21%** of the tasks in the case of the experimental system. This finding suggests that the experimental system better interpreted the intent of user queries, reducing the need for manual adjustments.

4.2 Qualitative Feedback

Following the search tasks, participants completed a questionnaire assessing their subjective experience with both systems. The questions attempted to measure the following aspects.

4.2.1 User Satisfaction

When asked whether the system met their expectations, **79%** of the participants rated the experimental system as more satisfying. Only **21%** preferred the default Google ordering for relevance. Users reported feeling more in control and better supported when their previous interests were reflected in the results using the experimental system. The difference between the two percentages is significant according to the results of the *z-test* for difference between two proportions **(z=3.02,**

p<0.003).

4.2.2 Relevance of Top Results

Participants were asked whether the first three results contained what they intended to find. Although a higher percentage of participants found relevant results within the top three positions in the experimental system (**79%**) compared to the baseline (**51%**), the difference did not reach statistical significance (***z = 1.58, p = 0.1***). Therefore, this finding should be interpreted as a trend suggesting potential improvement, requiring validation through a larger-scale study.

4.3 Behavioral Observations

Through direct observation, researchers noted that users appeared more confident and made decisions faster using the experimental system. Moreover, users exploring academic or technical topics benefitted the most, as their profiles built up with domain-specific keywords over time. It is worth mentioning that the system's performance improved with continued use, as the profile had more data to rely on—an indication of learning over time.

5. DISCUSSION

The results of this study provide encouraging evidence that the personalized re-ranking (experimental) system meaningfully improves user search experience compared to the baseline Google-based interface. Both the quantitative metrics and qualitative feedback converge on the conclusion that personalization leads to greater efficiency, reduced user effort, and higher satisfaction. This section interprets these findings, discusses the underlying reasons behind the observed improvements, and reflects on the implications for personalized information retrieval systems.

5.1 Efficiency and Reduced Cognitive Load

The significant reduction in time to information and number of clicks can be attributed to the system's ability to surface results that are contextually aligned with users' prior search behavior. By prioritizing documents that match previously expressed interests, the re-ranking mechanism reduces the need for trial-and-error navigation. This means that users encounter relevant items earlier in the ranked list, lowering both the time and cognitive effort required to locate satisfactory results. The fact that query reformulations also dropped substantially supports this interpretation: users did not feel compelled to adjust their queries repeatedly because the system was already interpreting intent more accurately.

5.2 User Satisfaction and Perceived Control

The higher user satisfaction ratings are likely a consequence of the system's personalization aligning closely with users' sense of agency and control. Participants reported that their prior interests were reflected in the results, creating a sense of continuity across sessions. This is particularly important because searchers often evaluate not only the final outcome but also how well the system "understands" them. The trust-building element is also noteworthy: while some users initially expressed concern over personalization, their attitudes shifted once they recognized that the re-ranking was based only on their own prior searches. This suggests that transparency in personalization mechanisms plays a critical role in building user trust.

5.3 Relevance of Top Results and Behavioral Confidence

Although the difference in the perceived relevance of the top three results was not statistically significant, the upward trend in favor of the experimental system reinforces the idea that even modest improvements at the top of the ranked list can positively influence the search process. Behavioral observations showed that participants interacted with the experimental system more confidently and decisively, particularly when engaging with complex academic or technical queries. This indicates that the personalization model may be especially effective in domains where users have persistent or evolving information needs, as the system can progressively refine its ranking strategy with accumulated interaction data.

5.4 Why These Results Emerged

Taken together, the improvements can be explained by the dynamic nature of the re-ranking model. Unlike static ranking approaches, the personalized system integrates a learning mechanism that adapts to user preferences over time. This adaptation minimizes redundant effort, decreases frustration, and makes the interaction smoother. Furthermore, the exponential decay function used in the system ensures that older, less relevant preferences gradually lose influence, allowing the profile to stay current and preventing outdated interests from distorting the results. Thus, the balance between memory (long-term user preferences) and adaptability (updating based on recent behavior) likely drove the observed performance gains.

5.5 Broader Implications

These findings underscore the potential of lightweight, user-centric personalization approaches to enhance search systems without requiring invasive data collection. They also highlight that user trust can be maintained—and even strengthened—when personalization is made transparent and grounded in the individual's own interaction history rather than opaque third-party data. The results suggest that future search technologies should focus not only on improving relevance metrics but also on supporting user confidence, efficiency, and control through adaptive learning models.

6. FUTURE WORK

While the personalized re-ranking system demonstrated clear improvements in efficiency, satisfaction, and perceived relevance, several directions for future work remain open.

6.1 Larger-Scale Evaluation.

This study was conducted with a relatively small participant pool. A larger and more diverse user base would provide stronger statistical power and help determine whether or not the observed improvements generalize across different demographics, domains, and search contexts.

6.2 Longitudinal Studies.

Because the system adapts to user behavior over time, a longer-term study would shed light on how personalization evolves with extended use. Such studies could reveal whether efficiency gains plateau or continue to improve as the user profile becomes richer.

6.3 Richer Personalization Features.

Currently, the system relies primarily on keyword-based profiling. Future work could explore semantic embeddings, topic modeling, or hybrid approaches that capture deeper aspects of user intent. Integrating contextual signals such as task type, device, or time of day may also yield more accurate re-ranking.

6.4 Balancing Personalization and Diversity.

Although personalization increases efficiency, it risks narrowing the range of information shown to users. Future research should investigate mechanisms to balance personalization with result diversity, ensuring that users are not trapped in "filter bubbles" and still have opportunities for serendipitous discovery.

6.5 Transparency and User Control.

Findings showed that user trust improved once personalization mechanisms were explained. Future versions of the system could incorporate user-facing controls that allow individuals to view, edit, or even reset their profiles, thereby improving transparency and autonomy.

6.6 Integration with Existing Search Engines.

Exploring how such a system could be deployed as a browser extension, plugin, or middleware layered on top of mainstream search engines would be a practical next step. This would test the system's real-world applicability and scalability.

7. CONCLUSION

This study demonstrated that a personalized re-ranking system can significantly enhance the web search experience by reducing search time, lowering the number of clicks and query reformulations, and improving overall user satisfaction. Both quantitative and qualitative findings confirmed that personalization enables users to locate relevant information more efficiently while fostering greater trust and a stronger sense of control. Importantly, the system's adaptive learning mechanism allowed it to remain responsive to evolving user interests, highlighting the value of personalization grounded in user history rather than external data. Overall, the results suggest that lightweight, transparent personalization approaches can meaningfully complement traditional search engines and offer a promising direction for improving information retrieval in diverse contexts.

REFERENCES

[1] Tchantchou Samen, S.; Ezin, E.C.; Onana, F.S.M. A dynamic personalized web search based on user profile and semantic ontology. *Int. J. Web Inf. Syst.*, 2017, **13**, 1–19.

[2] Kehinde, A.; Adebayo, O.S.; Odetunmibi, O.A. Context-aware personalized web search using DROPT algorithm. *J. Web Eng.*, 2018, **17**, 257–276.

[3] Zhang, X.; Cheng, X.; Liu, Y. A hybrid recommendation model combining user behavior and content for academic paper retrieval. *Scientometrics*, 2019, **120**, 53–72.

[4] Mohammadi, S.; Ahmadi, A.; Adibi, P. Improving personalized web search using local evidence densities in dynamic profiles. *Expert Syst. Appl.*, 2020, **140**, 112885.

[5] Lee, K.; Jeong, H.; Lee, C. Personalized web search via social signals and behavioral patterns. *Inf. Process. Manag.*, 2022, **59**, 102953.

[6] Chen, L.; Pu, P. Dynamic user clustering for personalized web search with BERT-based interest modeling. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2023; pp. 2127–2131.

[7] Liu, X.; Croft, W.B. Statistical language models for information retrieval. *Found. Trends Inf. Retr.*, 2005, **1**, 1–103.

[8] Ricci, F.; Rokach, L.; Shapira, B. *Introduction to Recommender Systems Handbook*; Springer: Boston, MA, USA, 2011.

[9] Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 2005, **17**, 734–749.

[10] Lops, P.; De Gemmis, M.; Semeraro, G. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*; Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Eds.; Springer: Boston, MA, USA, 2011; pp. 73–105.

[11] Burke, R. Hybrid web recommender systems. In *The Adaptive Web*; Brusilovsky, P., Kobsa, A., Nejdl, W., Eds.; Springer: Berlin, Heidelberg, Germany, 2007; pp. 377–408.

[12] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Minneapolis, MN, USA, 2019; pp. 4171–4186.

[13] Reimers, N.; Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, China, 2019; pp. 3982–3992.

[14] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

[15] Grootendorst, M. *KeyBERT: Minimal keyword extraction with BERT embeddings*. Available online: https://github.com/MaartenGr/KeyBERT (accessed on 30 August 2025).