University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 9, Issue 1, 2025
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/

Analyzing User Information Seeking Strategies in Interactive Information Retrieval Systems: A Mixed-Methods Evaluation Framework

A. R. Nangraj, M.S. Chandio, Yarsir Arfat Malkani, Quratulain Nizamani
Institute of Mathematics and Computer Science, University of Sindh, Jamshoro
rehman.nangraj@usindh.edu.pk, m.s. mschandio@hotmail.com, yasir.malkani@usindh.edu.pk, qn@usindh.edu.pk

Abstract: Understanding how users adapt their ISS across task types is critical for evaluating and designing effective IIR systems. While traditional evaluations focus on performance metrics like relevance or precision, they often fail to capture user behavior's cognitive and strategic dimensions. This study proposes a mixed-methods evaluation framework to analyze ISS across factual, exploratory, and comparative tasks. A user study involving 30 participants used a custom academic search interface. Data were collected through interaction logs, satisfaction surveys, and post-task reflections. The findings show that strategy use is highly task-dependent. Factual tasks encouraged direct lookup behaviors, whereas exploratory and comparative tasks involved iterative query refinement, multi-document synthesis, and opportunistic exploration. Satisfaction correlated positively with strategy diversity; expert users demonstrated greater adaptability in complex tasks. Five dominant strategy types were identified and linked to task complexity and user expertise using quantitative log data and qualitative user narratives. This study advances IIR evaluation by offering a replicable, strategy-sensitive framework beyond traditional click-based metrics. The framework provides actionable insights for designing adaptive, user-centered search systems that support a range of strategic behaviors across varying information needs.

Keywords: Interactive Information Retrieval, Information Seeking Strategies, Mixed-Methods Evaluation, User Behavior, Task-Based Search.

I. Introduction

In the era of information abundance, effective search is no longer defined solely by the ability to retrieve relevant documents but by the user's capacity to navigate, evaluate, and strategically interact with vast information environments. Interactive Information Retrieval (IIR) is a central paradigm for understanding and designing search systems that support such complex, dynamic information behaviors (Ingwersen & Järvelin, 2005; Kelly, 2009). IIRs move the discourse beyond the static query-response models to dynamic, iterative interaction between users and systems repeated over their search processes as user goals, strategies, and cognitive states change over time in the search session.

In the IIR's previous work, it is argued that users apply a whole range of Information Strategies (ISS) that are relevant not only to the goals and the prior knowledge of the user but also to the task, the interface design, and the context, in which they are working (Belkin, 1995; R. W. White & Roth, 2009). Traditional models like (Bates, 1989) berrypicking theory and Kuhlthau, (1993) information search process illuminate the nonlinearity, affective, and recursive nature of human information seeking. Recent empirical work has also shown how users adopt a variety of strategies (orienteering, parallel exploration, and opportunistic query reformulation) to make sense of information in academic, professional, and everyday search contexts (Athukorala et al., 2016).

However, these people's efforts and the evaluation frameworks for IIR are still insufficient in terms of how they account for and analyze human users' strategies in real-world tasks. Performance metrics commonly studied include precision, dwell time, or query length, but these often do not shed light on the cognitive or strategic underpinnings of search behavior (Kelly et al., 2015). Furthermore, studies that look to describe ISS frequently use qualitative methods, think-aloud protocols, or interview data but fail to place this data within scalable, repeatable frameworks of analysis (Aloteibi, 2020; Maxwell & Azzopardi, 2016). There is consequently an increasingly pressing need for mixed methods of evaluation that allow connecting behavior log analysis with an interpretive, contextually sensitive understanding of user strategy.

In this study we address that gap by suggesting an evaluation methodological framework based on mixed methods in order to assess user information-seeking strategies in IIR systems. Using behavioral interaction logs, task-level user feedback, and a priori observations, the framework facilitates the systematic investigation of how users change strategies under varying task types and interface conditions. Moreover, the framework was implemented in a user study with diverse IIR tasks (from factual look-up to exploratory synthesis) that were recorded using real-time interaction data and post-task reflective measures. Supplemented by system-centric and stand-alone behavioral indicators, the hope is to provide a comprehensive view of search as a strategic, situated practice.

The relationship between task type and strategy use is of specific interest in the study. Previous research has demonstrated that the complexity and structure of tasks are key determinants of user behavior (Li & Belkin, 2008;

Wildemuth, 2004). For example, factual tasks tend to evoke rather direct solution strategies such as pinpoint questioning and minimal reformulation. At the same time, exploratory tasks promote more iterative, reflective behaviors (comparison, backtracking). Instead, in comparative tasks, users may need to make sense of information from different perspectives, that is, strategies that facilitate source triangulation and critical judgment (Aula et al., 2010; R. W. White & Roth, 2009). By explicitly associating strategy patterns with task properties, this framework facilitates richer interpretations of not only system performance but also user experience.

IIR has gone a long way in simulating user interaction, and no truly integrated evaluation of fragmented information-seeking strategies that are reduced at times to very small behavioral trace and at times to qualitative information only. To address this gap, we make our contribution with an integrated mixed method approach to the analysis of user strategies that can facilitate the extensive and scalable research of user strategies in various types of tasks. A combination of these quantitative and qualitative dimensions also provides a more holistic view of search behavior, a phenomenon that eventually leads to the creation of IIR systems not only effective but also with an understanding of context and user-friendly.

Another theme adopted in this research is the importance of differences in strategy in the context of users even within a particular task. Users vary in their knowledge about the domain, search skills, cognitive style, and emotional reactions all factors that may impact strategy choice and adaptation. Accounting for this diversity is key when designing adaptable IIR systems to respond to individual needs, and for assessing how systems support different types of information behavior. The framework can address this by categorizing user strategies not only based on the task but also based on the user profiles and the behavioral clusters that are derived from the interaction data.

II. Literature Review

Understanding how users seek information within IIR systems has long been central to information science research. Early models of IR model the efficiency of system output using static queries and relevance judgments, but the focus of study over the past two decades has begun to move towards understanding users as active and adaptive agents whose behaviors change through a search session (Ingwersen & Järvelin, 2005; Kelly, 2009). In this context, ISS has become an important concept in the description and evaluation of user interaction with IR systems, particularly when uncertainty, ambiguity, or an exploratory orientation is involved (Trippas et al., 2025; Zerhoudi & Granitzer, 2024).

Foundational work in ISS emphasizes that human searching is non-linear, iterative, and context-dependent. Model-based on berry picking The berrypicking model of Bates (1989) offered an alternative to the traditional 'the user makes a single query, and the system provides a fixed set of documents' model by showing that instead, users can berrypick, i.e., grab a small quantity of information from one place visit another, do the same, etc. Likewise, Kuhlthau's (1993) Information Search Process (ISP) included the concept that affective states feelings of uncertainty, confidence, and frustration play a critical role in influencing user strategy. These models formed the basis for viewing information-seeking as an active process with strategy selection, adaptation, and reflection (Adhav & Singh, 2024; Kiesel et al., 2024).

Belkin (1995) also elaborated on search as a learning process where "users explore, compare, and synthesize during their quest to gain knowledge." This perspective has since been empirically validated in school and professional settings, with evidence that users use multiple strategies determined by their tasks, prior knowledge, and system affordances (Azzopardi et al., 2024; Ji et al., 2024). These are orienteering, successive refinement, query iteration, parallel scanning, and opportunistic browsing.

One common result in the literature is that the complexity and structure of a task have a great impact on user strategies (Engelmann et al., 2023). Quite factual tasks often result in direct and end-focused strategies, such as low-level strategies like the one that consists of writing precise questions and exploring as few documents as possible. Additionally, exploratory or open tasks, which may involve the exploration of a phenomenon or the comparison of competing ideas, will tend to elicit more general and less linear behaviors. White & Roth (2009) found that participants were more likely to reformulate, view down the result pages, and engage in critical evaluation when carrying out complex tasks.

Task-based IR research further proved that this strategy usage changes even during the search session. Aula et al. (2010) observed that participants would change their strategies based on perceived success or failure, frequently mixing strategies (keyword variation combined with skimming or tab switching). These transitions are frequently subtle and context-sensitive; thus demand for fine-grained behavioral analysis and metric evaluation beyond static click metrics or session summaries.

Moreover, the solid theoretical basis and empirical research on ISS are methodologically dispersed. Some depend on qualitative means to obtain strategic behavior in the form of think-aloud protocols or user interviews (Aqle et al., 2022; Pradhan & Maharana, 2022; Solheim Johansen & Borlund, 2022), while others draw on log-based approaches to study search patterns at scale (Athukorala et al., 2016). Both methods have their advantages but have limits by themselves. Based on the qualitative methods, they are context-rich but yet hard to generalize or replicate. Log-based approaches are feasible at large scale but typically have low-resolution strategy representation based on surface-level signals, such as query length or dwell time, that may poorly represent users' cognitive processes or intent.

Recent work has tried to fill this gap by using mixed methods that combine behavioral evidence, user feedback , and qualitative insights. Kelly et al. (2015) integrated log data and post-task surveys to study user satisfaction and decision strategies. Similarly, Chen et al. (2023) applied clickstream data to explicit process data that reports gazes and strategy shifts in an Exploratory Search study. These findings highlight the utility of mixed methodologies for capturing the multifaceted nature of search strategy, especially in IIR when user intents are complex and changing.

The design of the interface is also important in influencing strategy use. Features like faceted search, snippet preview, and query suggestion, document clustering might be facilitating or limiting user exploration depending on how these features are implemented (Hearst, 2009; Liu et al., 2014). Its user studies did demonstrate that users adapt to the system tools and feedback. For example, Diriye et al. (2010) found that the exploratory search interfaces encouraged users to do more browsing and cross-document comparisons compared with traditional keyword-based systems. This flexibility complicates, even more, the categorization of the strategies because a single user could use different behaviors when interacting with the other HCI tasks.

However, current IIR assessments did not effectively consider the dynamic relation between user, task, and system. When measured in terms of click-through rate or relevance judgments alone, evaluation methods provide an incomplete picture of the extent to which the system supports successful strategy use. There is growing agreement that system evaluation frameworks should include strategy-sensitive evaluation metrics (e.g., reformulation frequency, exploration depth, or strategy-switching indicators) (Shah, 2014; Wadhwa & Zamani, 2021).

In spite of these developments, a consensus and effective approach to evaluating ISS in the IIR context by accounting for both behavioral and qualitative perspectives is still lacking. Most of the research is still siloed into either retrospectively describing strategies or algorithmically inferring strategies without any user validation. Few provide replicable ways to categorize strategies across a wide range of tasks and user types or to connect such strategies to things that can be measured, such as satisfaction, engagement, or task success.

This study aims to fill this gap by suggesting a mixed-method evaluation framework that combines user tasks, behavioral interaction logging, task-based feedback, and qualitative annotation. By learning from earlier models of ISS, we organize this analysis of strategy through observable behavior (e.g., patterns of reformulation, depth of navigation), contextual information (e.g., task type, user profile), and reflective user input uncategorized user Behavior analysis formulating and conducting a complex search is not an arbitrary process but is motivated by a variety of factors. Through our simulations, this study provides a practical toolkit for analyzing how actual users look for information in IIR systems and how to design systems and perform an evaluation to enable a wide range of users with diverse and adaptive information-seeking strategies.

III. Methodology

This study uses a mixed-method evaluation framework to examine how users exhibit diverse ISS while performing various types of tasks in the IIR setting. The study design integrates log-based indicators of behavior and textual feedback to address both the visible and assumed aspects of user behavior. Through these combined lenses, the study aims to develop a richer, context-sensitive view of search strategies and to evaluate to what extent the current IIR systems accommodate varying user needs.

Study Design and Participants

The empirical part of the study is an experiment involving a custom-developed academic search interface. Thirty participants were recruited from a university community consisting of postgraduates and early career researchers in various fields. The participants were chosen to represent a range of domain knowledge and search experience, two factors that are known to affect strategy choice (White, 2016). All participants had experience with web-based search systems and indicated an acceptance of academic search tools in general.

In the course of their participation, each participant carried out the three search tasks, one from each of three search types: fact-finding, exploratory search, and comparative search. The design was of within-subjects with counter-balanced order of course to reduce learning and fatigue. Participants were asked to do their best in performing each task in any way they felt was helpful. Session length was restricted to 15 minutes per task to represent the reality of search constraints although providing sufficient time for strategy use.

Task Typology

The tasks were intended to represent different levels of cognitive complexity and information requirements. Factual questions called for the recall of specific, factual information. Open tasks led to more in-depth inquiry and open-ended learning. Comparison-related items are required to analyze the alternatives or draw distinctions between two competing concepts. The typology is rooted in established frameworks in the fields of IIR and task-based IR that categorize tasks based on their structure, goal clarity, and information uncertainty (Li & Belkin, 2008). Employing task diversity is necessary to see how users change tactics given varying cognitive loads.

Data Collection

The behavior log was recorded through a behavior-logging system embedded inside the search interface. The system recorded timestamped events, which include the submission of the query, the number of times the query was reformulated, a click on the document, the dwell time (amount of time spent viewing a document), the presence of tab activity on the tags, and the session duration. This enabled the reconstruction of user interaction sequences at the fine-grained level necessary for discriminating and categorizing search strategy.

Qualitative data was simultaneously collected by using a pre-and post-task questionnaire, as well as brief semi-structured interviews. Pre-task questionnaires were used to record what participants knew about the topic and whether they were familiar with it. Surveys administered following task completion gathered both Likert-scale ratings of task difficulty, satisfaction, and perceived success, and open-ended questions on search behavior and reasoning. Participants were invited to detail their strategies in their own words, explaining how and why they selected specific queries, documents, or exploratory paths.

One group of participants (n = 10) additionally participated in think-aloud sessions, where they were encouraged to speak their thoughts on the task. These sessions were transcribed and coded as complements to log data to gain further understanding of user intent, perception, and judgment.

Mixed-Methods Strategy Coding

The analysis was coding and sorting out strategies from users' interaction logs and self-report behavior. Also, adopted an abductive approach to strategy categorization whereby theory shapes but is not allowed to dictate analysis (Creswell & Plano, 2017).

First, raw interaction logs were parsed into sequences of query episodes, which were defined as runs of actions initiated by a new query or reformulation. Computed for each episode behavioral evidence based on the reformulation rate and the amount of step-by-step progress (result depth explored, time spent per document, frequency of switching from one query or tab to another). Episode-level strategy labels were assigned in terms of direct lookup, successive refinement, orienteering, berry picking, and synthesis on the basis of these cues through an extension of taxonomy (R. White, 2016).

Secondly, the thematic coding of user feedback was used to confirm and refine the strategy names. Statements like "I reformulated my query because I didn't find what I was looking for" or "I have three different tabs open to check sources" mapped to behavioral characteristics such as query reformulation and parallel exploration. Inter-coder reliability was scored on a subset of the data (Cohen's κ = 0.84), demonstrating strong agreement.

The integration of these sources allowed this strategy categories to be based on not only observable activities but also user purpose and understanding, which is a primary benefit of using a mixed-methods evaluation (Creswell & Plano, 2017).

Analysis and Metrics

Description and inferential statistics were employed in analyzing the quantitative findings. Primary behavioral measures were the average number of reformulations per task, average query length, and the nature of the reformulation (e.g., changes in topic or specificity). Depth of exploration was measured by the number of documents encountered, average time spent with documents, and scroll depth within sessions. Task performance was measured with the time for each task and the overall number of actions (e.g., queries, clicks, tab switches). Subjective measures (i.e., user satisfaction and perceived task success) were collected with a self-assessment of the two ratings on a 5-point Likert scale after each task.

To compare task types and user profiles, repeated-measures ANOVA was used, and the Bonferroni adjustment was applied for multiple comparisons. Correlations between behavioral indicators (i.e., number of reformulations or dwell time) and user-reported satisfaction were tested using Pearson's correlation coefficients. Interview and open-ended survey information were blended with similar behavioral log events using a convergence coding matrix that mapped log data events with theme and strategy codes. This afforded a view of patterns that were universal and that differed at the individual strategy level, among users on tasks and within tasks, a view of searching in context.

IV. RESULTS AND ANALYSIS

The study results present the findings in four main categories: query reformulation behavior, depth of exploration and patterns of interaction, satisfaction and perceived task performance, and an overarching typology of search strategies.

Query Behavior and Reformulation Patterns

The degree of user reformulation differed markedly between task conditions as the result shows that users reformulated a query most often in the exploratory task (M = 3.8) followed by comparative (M = 3.5) and a factual task

(M = 2.1). A repeated measures ANOVA revealed that there was a task type effect for reformulation frequency (F(2, 58) = 6.79, p <. 01), and further post hoc analyses revealed that factual tasks differed significantly from exploratory ones (p =. 004).

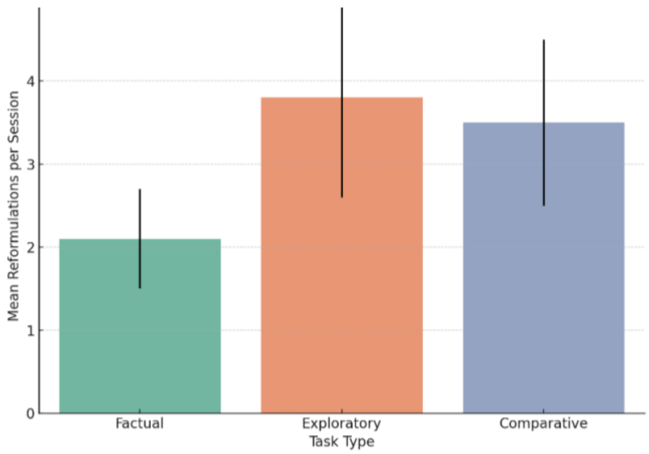

Figure 1: Average Query Reformulations By Task Type

Exploratory and comparative tasks involved iteration of query development to a larger extent than did factual tasks, indicating an increased need for cognitive and information processing. For the exploratory and comparative tasks, in particular, participants explained that they needed to "try a few different angles" or "reword the question" when their initial queries returned results that were too broad or ambiguous. Reformulations frequently ranged from topic expansion (e.g., adding context-specific keywords) to re-focusing on the basis of previous results. This finding is consistent with the conclusion that more complex tasks produce more dynamic searching.

Exploration Depth and Interaction Patterns

Exploratory actions (document inspection, scroll depth, and tab use) also varied across tasks. As shown in Figure 2, available attention (rank fused clicked) was highest for comparative tasks (M = 58 sec), followed by exploratory tasks (M = 51 sec) and factual tasks (M = 36 sec). A 2A was computed for a task type × user expertise interaction, which was significantly affecting dwell time (p <. 05), which means that cognitive demand and user profile are both significant factors influencing engagement depth.

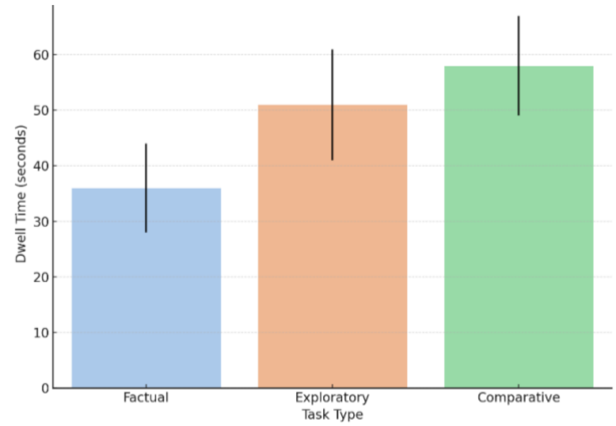

Figure 2: Average Dwell Time Per Document By Task Type

Users interacted more with documents in complex, open-ended tasks, indicating more evaluation and synthesis. Browsing tasks resulted in more expansive scrolling and deeper processing of longer documents. Comparison tasks were highly salient for multi-tab behavior; users commonly had multiple sources open at once to contrast perspectives. "I found myself having three articles open at once and switching continually to compare the points each one was making," one participant wrote.

This conduct aligns with the antagonistic feature of comparative synthesis where the triangulation of evidence and ideas between the sources is congruent. The trends highlight the significance of IIR systems in facilitating tab-handling, parallel browsing and document annotation tools to help facilitate synthesis.

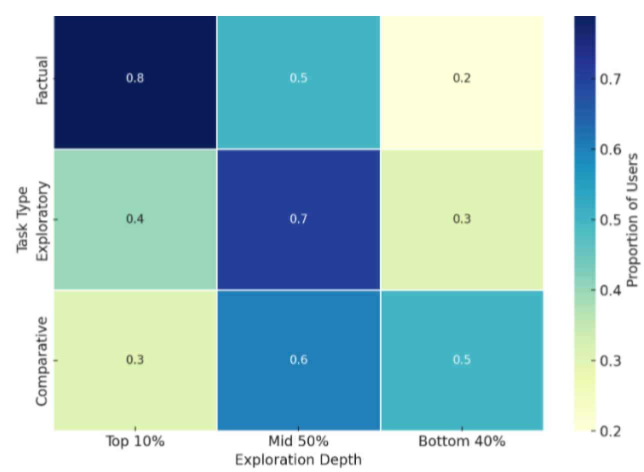

Figure 3: Session Behavior Heatmap

The exploration profundity of document distribution of the various tasks. With tasks based on facts, it was found that a high proportion of users (80 percent) were in the top 10 percent in terms of exploration depth, and thus it was found that even target-directed queries often necessitated expansive scanning to confirm a fact or compare formulations. On the other hand (the exploration of the information scent), the proportion of exploratory tasks showed a higher number (70%) in the middle-range exploration group, which indicates moderate sustained browsing behavior that is characteristic of sense-making and learning activities.

Also, there was increased polarization of competitive tasks. Although 60 of the users continued to do midrange exploration, 50 percent of the users were also within the bottom 40 percent with the shallower browsing behavior being also evident in certain instances. The first possibility is that this means that a few good sources have been selectively processed or that it is due to cognitive limitations arising out of the need to match complex material.

The heatmap also shows that the task-dependent variation in the extent of the exploration of search results that users do is significantly dependent on search systems that support both breadth and depth engagement with the search system, based on the task context.

Satisfaction and Task Success

Self-rated satisfaction differed between task types. The average satisfaction of factual tasks was the highest (M = 4.4 on a 5-point scale), while ordeal had the lowest (M = 3.7), which was intermediate between the averages of comparative (M = 4.1) and exploratory (M = 3.7). Exploratory tasks were rated as the most difficult task, and satisfaction was positively related to strategy diversity (r =. 62, p <. 01), particularly if users felt they had "learned something useful" in the absence of a (clear) answer.

Inexperienced participants responded less positively to exploratory tasks, feeling insecure and unsure when to stop or how to 'measure' success. These findings are consistent with affective models of information seeking, including Kuhlthau's (1993) model, which highlights cognitive- and affective-based aspects of complex search tasks.

Table 1. Satisfaction Scores by Task Type

Task Type	Mean Satisfaction	Std. Dev.
Factual	4.4	0.5
Exploratory	3.7	0.7
Comparative	4.1	0.6

Satisfaction was greatest for tasks with clear, informational goals (factual) and lowest for tasks that involved open-ended exploratory. However, in-task variation was affected by user expertise, the strategy employed, and perceived learning.

After triangulating user logs, user feedback, and themes from interviews, found five emergent search strategies: direct lookup, iterative refinement, exploratory browsing, comparative synthesis, and opportunistic discovery.

Integrated Strategy Typologies

Table 2. Dominant Strategy Types by Task Category

Task Type	Most Common Strategies	Behavioral Characteristics
Factual	Direct lookup	Short, targeted queries; minimal reformulation; low document depth
Exploratory	Iterative refinement, browsing	Frequent reformulation; deep scrolling; varied navigation paths
Comparative	Comparative synthesis, refinement	Multi-tab usage; side-by-side source evaluation; cross-document reading

In exploratory tasks, users usually start with an initial broad search and iteratively narrow the scope of the search by exploring result lists or examining batches of documents. Such think-aloud data suggests statements like, "I was trying to understand the big picture before I narrowed in on what I found," which are consistent with berrypicking and orienteering behaviors (Bates, 1989; Belkin, 1995).
Comparison between working on tabs and comparing across sources was common in comparative tasks. During these sessions, the individuals would tend to refer to the benefits, drawbacks, or disparities as the characteristic of capitalizing on the strategic synthesis instead of direct, board-game-like access to information. This also took more time to dwell and scrolling to reach the desirable information of a given source.

The opportunistic findings were observed in all types of tasks but were particularly high during the exploratory session. In other instances, users would make middle-task or change of topics in a task due to the discovery of not anticipated but interesting information. This flickering behavior that might have sometimes not been what the user first intended was termed by users as a net positive attribute particularly those with greater knowledge of the domain.
The results suggest the notion that the type of strategies employed is highly influenced by the type of task that will be involved, in that the more open-ended the task type, the more the reformulation rates, exploration, and cognitive challenging activity will be induced. Moreover, the strategy is not only difference as per the type of tasks but also among user profiles, i.e. domain knowledge, confidence and experience.

The combination of the behavioral data and qualitative information allowed discovering fine-grained structure of strategies that validated the theoretical models of ISS and revealing system support that did not involve explorative or comparative behaviors. It is based on these insights that the discussion in the next section is built on the implications that these observations have on the IIR system design and evaluation frameworks.

Discussion

This research deals with the generalization of the ISS of users in various types of tasks in an IIR system, using a mixed method that involves a combination of behavioral and user responses which are related to qualitative data. The result of the study indicates the significant role of task complexity, user expertise and search context in the strategy selection and strategy adaptation.

Among the significant ones, the one that can be made out is that the type of task has a potent influence on the strategies and their change. Direct facet impoverished lookup behavior (so-called task effectiveness of the percentage of facet impaired searches) was most frequently observed with simple and well-defined tasks, which is likely to be anticipated in reformulation minimalism and superficial document processing (Li and Belkin, 2008; White and Roth, 2009). Exploratory and comparative tasks, in their turn, provoked more active and cognitively challenging strategies like iterative refining, browsing, and synthesis across the sources. These actions can be attributed to such models as berry picking Bates 1989, which is concerned with adaptive, nonlinear searches, and Kuhlthau (1993) affective model, which relates user uncertainty with the maturity of utilizing strategies.

Most crucially, the integration of logs with user feedback informed not only what strategies were employed but also why they were. Participants described changing approaches in response to uncertain answers or when they had acquired a better comprehension of the subject. For instance, a switch of broader to more specific questions or comparing several documents to reconcile conflicting information reflects self-regulation of the learning within a search task. This behavior would be hard to deduce from logs alone, thus the importance of a multi-method approach.

The results point to a number of opportunities in IIR system design and user systems to support more diverse and dynamic user strategies. There are several aspects in which systems should support and acknowledge iteration, like query reformulation and refinement. Dynamic query suggestions, semantic filters , or reformulation advice may help users refine their interests, particularly in an exploratory situation.

Second, multi-document interaction, especially for tasks such as comparison, highlights a requirement for interfaces that facilitate tabbed browsing, document comparison views, and annotation capability. Cross-document synthesis is also a behavior type of users, but present search interfaces usually ignore it. Systems that enable this type of synthesis, e.g., with the use of visual overlays or structured comparison mechanisms, can support users to make complex decisions more effectively.

Third, the existence of opportunistic discovery behaviors implies that IIR systems may need to subordinate some measure of random 'serendipitous' inquiry to the original query without penalizing deviations not inscribed in the original query. For example, showing thematically related material or enabling topic drift within limits would improve the validity of exploratory search tasks (Sanderson, 2010; Zhang et al., 2020).

This work lends additional support to the fact that pure evaluative metrics are not enough to capture user experience and strategy effectiveness in IIR. Metrics such as click-through rate or session length are unable to differentiate between strategic browsing and confusion or to covet affective or cognitive satisfaction (Kelly et al., 2015; Reinanda et al., 2020; Yang et al., 2016). The log-based on and off-task measures (e.g., reformulation counts, dwell time) were useful behavioral metrics but could only be given depth of interpretation by triangulating with user feedback and qualitative coding.

This study suggests strategy-conscious evaluation models which combine behavioral measures with reflective user information. These are satisfaction scores, perceived task success and verbal preferences of search strategies. By coding the types of strategy and relating them to interaction patterns, to identify an approach that is repeatable to offer the exposure of how users interact with IIR systems beyond and above surface-level efficiency.

In addition, this approach complements recent work in human-centered and explainable IR, which demands evaluations that reflect user intentions, understanding, and agency. To be practicable, evaluation procedures must include performance measures, task-specific behavior patterns, and user-defined measures of success to obtain a complete picture of the system's effectiveness.

An interesting observation is the influence of competence and task competence on exploration strategy variability. Although expert users utilized fewer strategies overall, they relied on synthesis and refinement at an earlier point in time and reported higher satisfaction despite the difficulty. They are also prevented from staying in the direct lookup mode or from becoming frustrated in open tasks (opposite to beginners).

This heterogeneity indicates that IIR systems should provide adaptive interfaces that recognize users' intent and experience. Personalization models founded on session history, reformulation statistics , or time on task may be leveraged to provide scaffolding or interface adaptation aimed at the user's strategy style. These would then enhance language usability for novices but not limit the flexibility for advanced users.

V. Conclusion

This study integrated behavioral log data with user feedback and qualitative insights to understand how users adjust their strategies through fact-finding, exploratory, and comparison-based search tasks. The findings strongly indicate that user behavior is very much task-dependent, and the evaluation metric cannot be the same; even traditional evaluation metrics cannot cover the complexity of real search behavior.

The mixed-methods approach conferred distinct advantages over the use of quantitative or qualitative methods alone, facilitating a richer understanding of user behaviors. This integrative viewpoint should be a model for future evaluations, which should be capable of including user cognition, satisfaction, adaptability as well as system performance. These are useful to guide how search systems may evolve in order to handle new and complex information needs by the many different classes of visually challenged users.

Recommendations

- ✓ IIR systems should provide more than one strategy in a session, allowing users to shift between lookup, exploration, and synthesizing modes, considering their changing demands.
- ✓ Evaluation frameworks should include precision-based metrics as well as indicators of strategy diversity, depth of engagement, and satisfaction.
- ✓ Personalization and scaffolding techniques could help users learn to use the system expertly and improve accessibility without compromising powerful functionality.
- ✓ Researchers and practitioners would benefit from adopting integrated methods of interaction that unify behavioral logs with user stories to understand why people do what they do.

This study provides a replicable analysis framework that throws light on user strategies in IIR and forms the building block towards systems that similarly can adapt in a cognitive and situational way as well as are effective.

References

[1]. Adhav, H., & Singh, V. (2024). *Modeling Topic Evolution to Steer Interactive Information Search* (pp. 586–596). https://doi.org/10.1007/978-3-031-12700-7_60

[2]. Aloteibi, S. (2020). A user-centred approach to information retrieval. *Apollo - University of Cambridge Repository*.

[3]. Aqle, A., Al-Thani, D., & Jaoua, A. (2022). Can search result summaries enhance the web search efficiency and experiences of the visually impaired users? *Universal Access in the Information Society*, *21*(1), 171–192. https://doi.org/10.1007/s10209-020-00777-w

[4]. Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A., & Vreeken, J. (2016). Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, *67*(11), 2635–2651. https://doi.org/10.1002/asi.23617

[5]. Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 35–44. https://doi.org/10.1145/1753326.1753333

[6]. Azzopardi, L., Breuer, T., Engelmann, B., Kreutz, C., MacAvaney, S., Maxwell, D., Parry, A., Roegiest, A., Wang, X., & Zerhoudi, S. (2024). SimIIR 3: A Framework for the Simulation of Interactive and Conversational Information Retrieval. *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, 197–202. https://doi.org/10.1145/3673791.3698427

[7]. Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, *13*(5), 407–424. https://doi.org/10.1108/eb024320

[8]. Belkin, N. (1995). Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications*, *9*(3), 379–395. https://doi.org/10.1016/0957-4174(95)00011-W

[9]. Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

[10]. Creswell, J., & Plano, C. (2017). Designing and conducting mixed methods research. *(3rd Edition) SAGE Publications.*

[11]. Engelmann, B., Breuer, T., Friese, J. I., Schaer, P., & Fuhr, N. (2023). *Context-Driven Interactive Query Simulations Based on Generative Large Language Models*.

[12]. Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of information seeking and retrieval in context* (Vol. 18). Springer-Verlag. https://doi.org/10.1007/1-4020-3851-8

[13]. Ji, K., Hettiachchi, D., Salim, F. D., Scholer, F., & Spina, D. (2024). *Characterizing Information Seeking Processes with Multiple Physiological Signals*. https://doi.org/10.1145/3626772.3657793

[14]. Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, *3*(1–2), 1–224. https://doi.org/10.1561/1500000012

[15]. Kelly, D., Arguello, J., Edwards, A., & Wu, W. (2015). Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework. *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, 101–110. https://doi.org/10.1145/2808194.2809465

[16]. Kiesel, J., Gohsen, M., Mirzakhmedova, N., Hagen, M., & Stein, B. (2024). *Simulating Follow-Up Questions in Conversational Search* (pp. 382–398). https://doi.org/10.1007/978-3-031-56060-6_25

[17]. Kuhlthau, C. (1993). Seeking meaning: A process approach to library and information services. *Libraries Unlimited.*

[18]. Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, *44*(6), 1822–1837. https://doi.org/10.1016/j.ipm.2008.07.005

[19]. Maxwell, D., & Azzopardi, L. (2016). Simulating Interactive Information Retrieval. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1141–1144. https://doi.org/10.1145/2911451.2911469

[20]. Pradhan, D., & Maharana, B. (2022). Faceted Search Interface for Library Portal to Support Scholarly Information Seeking from E-Resource Collection: A Case of NIT Rourkela. *International Conference on Libraries of the Future: Emerging Trends, KIIT Bhubaneswar and MANLIBNET*.

[21]. Reinanda, R., Meij, E., & de Rijke, M. (2020). Knowledge Graphs: An Information Retrieval Perspective. *Foundations and Trends® in Information Retrieval*, *14*(4), 289–444. https://doi.org/10.1561/1500000063

[22]. Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends® in Information Retrieval*, *4*(4), 247–375. https://doi.org/10.1561/1500000009