A Framework for Knowledge Integration in Conversational Information Retrieval

Praveen Acharya[1,*], Noriko Kando[2] and Gareth J.F. Jones[1]

[1]*School of Computing, Dublin City University, Dublin, Ireland*
[2]*National Institute of Informatics, Tokyo, Japan*

Abstract

In both traditional and conversational information retrieval, users can choose to engage in exploratory searches with varying levels of knowledge about the subject of their information needs. They formulate search queries to express these needs, which are then used by retrieval systems to find relevant information. Users with substantial prior domain knowledge about the topic of the information need can create sufficiently rich queries which include appropriate domain-specific vocabulary, leading to retrieval of relevant search results, while those with limited domain knowledge struggle to formulate effective queries. The latter must refine their queries over multiple search passes as they learn more. This iterative search process imposes a high cognitive load and limits the effectiveness of traditional search systems. Conversely, conversational information retrieval (CIR) offers a multi-turn, iterative process where the user and the system can work collaboratively to help the user satisfy their information needs with reduced cognitive effort. With each interaction, information is progressively accumulated aiding users in better understanding the topic and improving their knowledge. By representing the user's knowledge and its continuous refinement, a CIR system can better comprehend and respond to the information need and support the users in satisfying their information needs, resulting in more effective search outcomes. However, existing CIR systems lack a framework for representing the user's knowledge during the current search dialogue. Leveraging a user's prior knowledge and information gathered during each interaction can potentially enhance CIR system performance by guiding subsequent system actions. To address this, we propose a framework for capturing and utilizing knowledge in CIR. This framework aims to improve the performance and adaptability of conversational search systems, making them more effective and responsive to users' evolving information needs.

Keywords

Knowledge Integration, Conversational Search, User Knowledge, Framework for CIR

1. Introduction

When using a search system with informational intent [1] to acquire information about a topic, users exhibit varying levels of familiarity with the topic. This knowledge disparity influences how different users formulate their search queries, resulting in different degrees of precision and search effectiveness. For example, an expert (knowledgeable) user can specify their information need with sufficient detail (well-defined query, including correct use of domain-specific vocabulary) for the search system to retrieve relevant documents. In contrast, a non-expert (ill-informed) user will have difficulty specifying their information need, leading to under-specified (vague) queries and poorly retrieved documents. The range of query specificity, influenced by the user's knowledge, is illustrated in *Figure* 1. Users with limited knowledge about the topic often struggle to accurately articulate their information needs, a challenge referred to as the *non-specifiability of need problem* [2]. Consequently, their queries might not precisely convey their requirements, making it difficult to retrieve relevant documents and often resulting in unsatisfactory search results. Therefore, a user's knowledge of the search topic greatly influences the search process, and a search system that can adapt based on this knowledge is highly desirable.

In exploratory information search scenarios, fulfilling a user's information need is typically a multi-turn and iterative process. Unlike straightforward searches, exploratory searches involve dynamic
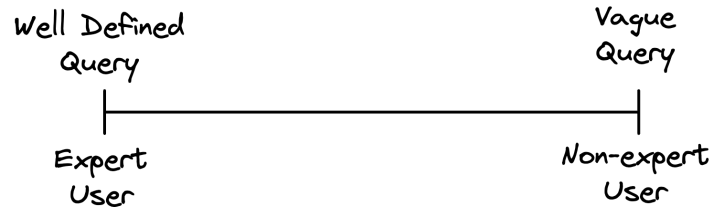
Figure 1: Query Specificity and User Knowledge range.

interactions where each query-response cycle provides valuable information that incrementally builds the user's understanding of the topic. This iterative process allows users to refine their queries based on the feedback received from previous searches. As users gather more information, their knowledge about the topic deepens, enabling them to formulate more precise queries. This ongoing cycle of query refinement and knowledge acquisition is often crucial to the user in satisfying complex information needs. Therefore, a search system designed specifically to support exploratory information retrieval should facilitate this iterative process, offering mechanisms to guide users through successive stages of query formulation and refinement. Ultimately, such a system would help users navigate through vast information spaces, progressively honing in on the information they seek.

Traditional information retrieval (IR) systems lack the capability to engage interactively with users. These systems are entirely user-driven, offering no interaction beyond basic relevance feedback methods. This places a high cognitive load on users, who must formulate progressive queries, examine retrieved results, and determine whether their information needs are satisfied. If not, they must repeat the process, which can be cumbersome and inefficient. Conversational Information Retrieval (CIR) addresses this limitation by incorporating an interactive and collaborative process between the system and the user. CIR systems reduce the user's cognitive load by adapting to user preferences and providing more personalized assistance. This interaction is often modelled as a user model that the system can exploit to better understand and meet the user's information needs as illustrated in *Figure* 2.

An important component of a user model for use in a search system is the representation of their *Knowledge*, potentially alongside that of other users with similar interests. As discussed earlier, knowledge plays a crucial role in the search process, enabling systems to be better able to satisfy a user's information needs. The significance of the role of knowledge exploitation in the search process is heightened in CIR systems because knowledge can aid in understanding users' needs and facilitating effective dialogue with the user and tasks such as query expansion, clarification, and information elicitation. Radlinski and Craswell [3] delved into conversational approaches to information retrieval and identified adaptability in conversation as a key characteristic for effective CIR systems. This adaptability allows search systems to dynamically adjust conversational dialogue based on users' existing knowledge and newly acquired information, continuously refining the process until users' information needs are met.

We argue that since *Knowledge* plays such a critical role in effective CIR, the user's domain knowledge—both prior to and developed during a search conversation should be modelled to support search systems in better understanding the user's information need and guiding the dialogue's progression [4]. Consequently, we seek to answer the following questions related to the role of knowledge in conversational information retrieval:

1. **RQ1:** How can knowledge be captured and utilized in conversational information retrieval?
2. **RQ2:** Does knowledge integration improve the effectiveness of a conversational search system?

To address these questions, we propose a framework for knowledge integration in conversational information retrieval. In this paper, we discuss the problem and several associated research challenges in implementing this framework.
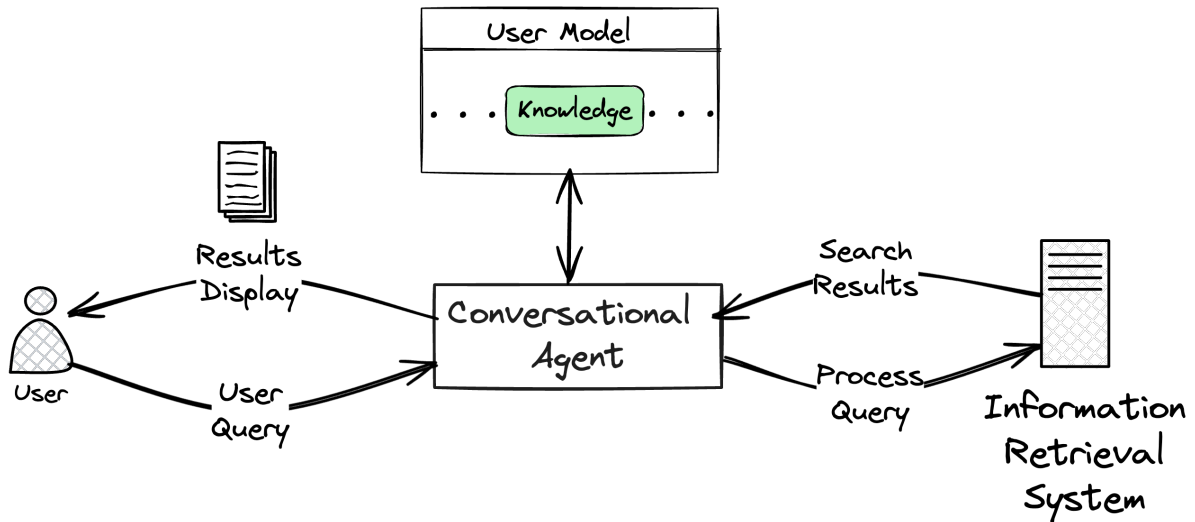
Figure 2: Knowledge Integration in Conversational Information Retrieval (CIR).

2. Background and Related Works

Modern information retrieval systems are utilized for a myriad of tasks, each contingent upon the user's specific objectives. Users approach these systems with varying intents. These intents are commonly classified into three categories: informational, navigational, or transactional [1]. Among these intents, a notable focus has emerged on leveraging retrieval systems for learning purposes, a concept termed *Searching-as-learning (SAL)* [5, 6, 7]. This paradigm reflects a shift towards utilizing retrieval systems not solely for accessing information but also as tools for learning. In the context of SAL, users engage with retrieval systems with the explicit aim of acquiring knowledge about a specific topic. Through iterative interactions with the retrieved documents, users progressively accumulate information to satisfy their learning goals. This process entails not only accessing relevant documents but also assimilating and comprehending the information contained within them.

A study focusing on the differences in searching patterns between domain experts and non-experts regarding a shared subject is reported in White et al. [8]. Their findings highlight distinctions in query formulation strategies across various levels of expertise, emphasizing the importance of understanding how different users approach search tasks. Furthermore, the level of domain knowledge influences how the user formulates their queries [9, 10, 11], with non-experts using more keywords than experts and experts producing more new keywords than non-experts [12, 13]. Moreover, Zhang et al. [14] found that insights from data collected during the search process can provide valuable indications of a user's domain knowledge suggesting that analyzing user interactions with retrieval systems can yield valuable information about their knowledge. A study by Hagen et al. [15] showed that users can learn query terms while engaging in searching and reading activities. This suggests that during the search process, users gradually refine their query formulation and enhance their understanding of the topic through iterative interactions with search results highlighting the importance of incorporating knowledge mechanisms into the retrieval systems. A search system can assist users in achieving their learning objectives more quickly by estimating how much they learn. For example, it can do this by retrieving documents that match not only their specific query but also their existing knowledge about the topic.

Several frameworks have been developed to integrate user knowledge into retrieval systems [16, 17, 18]. Câmara et al. [19] introduced a framework focused on representing user knowledge during search sessions. This framework estimates a user's knowledge about a specific topic by maintaining an internal representation that continuously updates throughout the session. They achieve this by employing a combination of keyword-based methods and Large Language Models (LLMs) based methods. Their
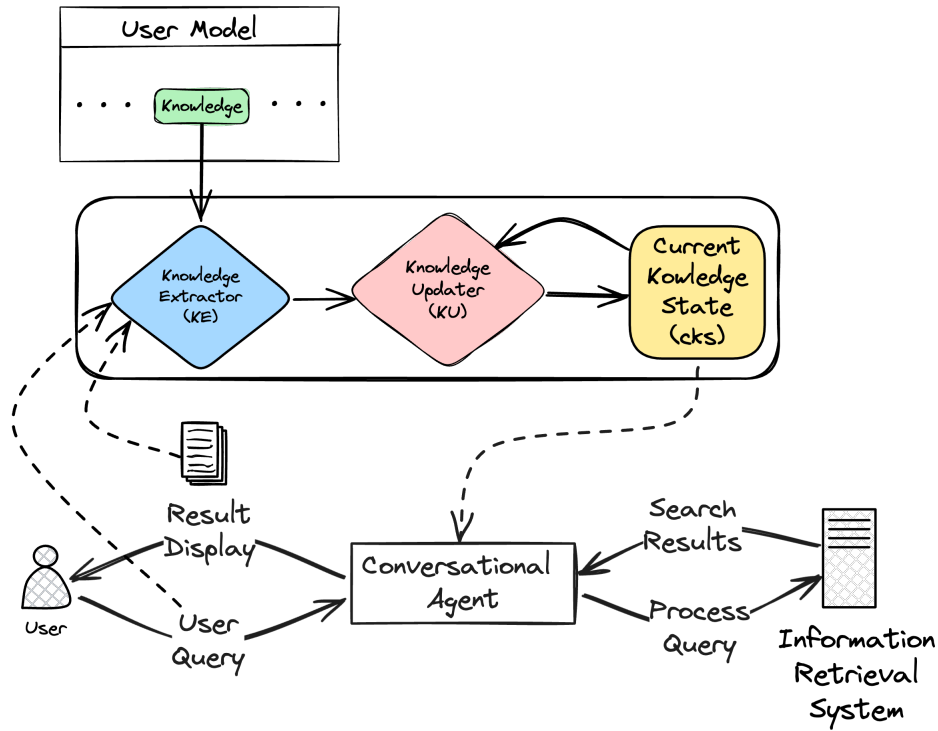
Figure 3: Proposed Framework Architecture

study demonstrates that this internal representation effectively correlates with users' actual knowledge levels. Expanding on this work, the framework was extended to incorporate named entities, leveraging the relationships between these entities to better represent and measure user knowledge during a search session [20]. This enhancement suggested that utilizing named entities complements earlier approaches, offering a more nuanced estimate of users' knowledge. Additionally, Nasser et al. [21] utilized knowledge graphs to represent both the user's knowledge and the knowledge goal, demonstrating that the graph-based approach can capture complementary aspects of knowledge.

Despite the importance of the role of knowledge in the search process highlighted in the aforementioned work, the formal utilization of knowledge in CIR remains surprisingly unexplored. Leveraging a user's prior knowledge and their knowledge of the search topic accumulated during a conversational dialogue would appear to have the potential to significantly enhance CIR system performance by guiding subsequent actions resulting in more efficient and effective search outcomes. The remainder of this paper discusses a proposed framework for the exploitation of knowledge in CIR with a focus on the theoretical aspects of the proposal and reviews some relevant operational approaches used in previous studies, leaving the full implementation details of the framework for future research.

3. Framework

Our proposed framework for integrating knowledge into CIR is built around three primary components. In the following sections, we discuss each component in detail, explaining their roles and functionalities within the framework.

3.1. Components

3.1.1. Knowledge Extractor (KE)

The knowledge extractor is responsible for identifying and extracting pertinent information from various sources, including queries, documents, and user knowledge history. It analyzes these sources to

extract what is deemed relevant knowledge *e.g.* topical knowledge, task knowledge etc...

$$K_{\text{source}} = KE(\text{Source}) \tag{1}$$

where *source* refers to any of the following: Query (Q), Document (D), or User Knowledge (U_k).
The Knowledge Extractor plays a crucial role in various parts of the proposed framework. The user's interaction with a CIR system begins with the issuance of a query. The **KE** extracts knowledge from the user query (K_{query}) as well as from the existing user knowledge history stored in a user model (K_{U_k}). This initial extraction of knowledge constitutes the initial current knowledge state ($cks_{initial}$) in the CIR process.

$$cks_{\text{initial}} = \{K_{\text{query}}, K_{U_k}\} \tag{2}$$

Subsequently, the **KE** is used again after the CIR system retrieves relevant documents. The decision of which documents to extract knowledge from depends on whether the user has interacted with these documents.

$$K_{\text{Document}} = KE(\text{Documents}) \tag{3}$$

3.1.2. Current Knowledge State (cks)

The Current Knowledge State reflects the current state of knowledge in the conversation at any given moment. It primarily interacts with the **KU**, which is responsible for incorporating new information and is updated at each step of the conversation. As the dialogue progresses, the **cks** is continuously updated to accurately represent the user's information needs and the context of the conversation. This ensures that the conversational system can take various actions based on strategies tailored to the user's current knowledge state, facilitating a more natural and fluid progression of the dialogue. Consequently, the system can effectively adapt to the user's evolving needs throughout the interaction.

3.1.3. Knowledge Updater (KU)

The Knowledge Updater, on the other hand, is responsible for seamlessly integrating the extracted knowledge into the existing knowledge state (*cks*). This component ensures that the system is updated with new knowledge at each turn in the conversation. By continuously incorporating new information, the **KU** keeps the knowledge state current and comprehensive, enabling the CIR system to maintain an accurate and up-to-date understanding of the user's needs and the context of the conversation. This continuous updating process is crucial for the system's ability to provide relevant and accurate responses and document rankings throughout the interaction.

$$cks_{\text{current}} = KU(cks_{\text{prev}}, K_{\text{Document}}) \tag{4}$$

where, cks_{current} is the updated knowledge state (*cks*) after the user's interaction with documents. The document is represented by K_{Document}, generated by Equation 3, and KU is a function that takes cks_{prev} and K_{Document} as inputs, combining them into an updated representation of the knowledge state.

These components collaborate to integrate and update the relevant knowledge represented by the current knowledge state (**cks**) in the conversation. The **KE** interacts with the system by extracting knowledge from retrieved documents at each interaction in the conversation. This extracted knowledge is subsequently used by the **KU** to modify and update the **cks**. The updated knowledge state helps the conversational system understand the user's current needs and based on this understanding employ various strategies to take appropriate actions. Additionally, the current state of knowledge in the conversation **cks** can potentially be leveraged and used to rank relevant documents according to the user's current knowledge at each step. *Figure* 3 illustrates how the **Knowledge Extracto**r and **Knowledge Updater** interact within the CIR system.

One significant challenge in dealing with knowledge is how to effectively represent it. Various approaches have been suggested for this, including extracting keywords [22, 17], using concept maps

[23], identifying named entities [20], creating knowledge graphs [21] and using LLMs[19]. The choice of representation method has significant implications for how the knowledge state is modelled during the search process. It also affects operational aspects linked to the *KU* component, influencing the efficiency and effectiveness of knowledge retrieval and application. Thus, selecting the appropriate representation strategy is crucial for optimizing knowledge management and utilization.

4. Concluding Remarks

This paper presents a framework for integrating knowledge into Conversational Information Retrieval (CIR) systems, addressing the challenge of making knowledge a central component of information retrieval. By leveraging both user knowledge and knowledge accumulated during conversations, the framework aims to improve CIR systems' ability to understand user's information needs, engage in contextually relevant dialogue, and provide more accurate responses. The framework consists of three main components: the *Knowledge Extractor (KE)*, *Current Knowledge State (cks)* and the *Knowledge Updater (KU)*. The *KE* extracts relevant information from user queries, documents, and prior knowledge, while the *KU* integrates this knowledge into the current knowledge state, ensuring the system remains up-to-date and comprehensive throughout the interaction represented by the *cks*. Future work will focus on implementing and refining these components, optimizing knowledge integration techniques, and conducting real-world evaluations. Advancing these areas will enhance CIR systems' ability to manage and utilize knowledge more effectively to meet users' evolving information needs.

Acknowledgments

References

[1] A. Broder, A taxonomy of web search, in: ACM Sigir forum, volume 36, ACM New York, NY, USA, 2002, pp. 3–10.

[2] N. J. Belkin, Anomalous states of knowledge as a basis for information retrieval, Canadian journal of information science 5 (1980) 133–143.

[3] F. Radlinski, N. Craswell, A theoretical framework for conversational search, CHIIR '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 117–126. URL: https://doi.org/10.1145/3020165.3020183. doi:10.1145/3020165.3020183.

[4] P. Acharya, Towards effective modeling and exploitation of search and user context in conversational information retrieval, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 5161–5164. URL: https://doi.org/10.1145/3583780.3616005. doi:10.1145/3583780.3616005.

[5] K. Collins-Thompson, P. Hansen, C. Hauff, Search as learning (dagstuhl seminar 17092) (2017).

[6] S. Y. Rieh, K. Collins-Thompson, P. Hansen, H.-J. Lee, Towards searching as a learning process: A review of current perspectives and future directions, Journal of Information Science 42 (2016) 19–34.

[7] J. Gwizdka, P. Hansen, C. Hauff, J. He, N. Kando, Search as learning (sal) workshop 2016, in: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1249–1250. URL: https://doi.org/10.1145/2911451.2917766. doi:10.1145/2911451.2917766.

[8] R. W. White, S. T. Dumais, J. Teevan, Characterizing the influence of domain expertise on web search behavior, in: Proceedings of the second ACM international conference on web search and data mining, 2009, pp. 132–141.

[9] H. L. O'Brien, A. Kampen, A. W. Cole, K. Brennan, The role of domain knowledge in search as learning, in: Proceedings of the 2020 conference on human information interaction and retrieval, 2020, pp. 313–317.

[10] T. Willoughby, S. A. Anderson, E. Wood, J. Mueller, C. Ross, Fast searching for information on the internet to use in a learning context: The impact of domain knowledge, Computers & Education 52 (2009) 640–648.

[11] C. Dosso, L. Tamine, P.-V. Paubel, A. Chevalier, Navigational and thematic exploration–exploitation trade-offs during web search: effects of prior domain knowledge, search contexts and strategies on search outcome, Behaviour & Information Technology 43 (2024) 2232–2258.

[12] C. Dosso, L. Tamine, P.-V. Paubel, A. Chevalier, The impact of expertise on query formulation strategies during complex learning task solving: a study with students in medicine and computer science, in: Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021) Volume V: Methods & Approaches 21, Springer, 2022, pp. 621–627.

[13] M. Sanchiz, A. Chevalier, F. Amadieu, How do older and young adults start searching for information? impact of age, domain knowledge and problem complexity on the different steps of information searching, Computers in Human Behavior 72 (2017) 67–78.

[14] X. Zhang, M. Cole, N. Belkin, Predicting users' domain knowledge from search behaviors, in: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval, 2011, pp. 1225–1226.

[15] M. Hagen, M. Potthast, M. Völske, J. Gomoll, B. Stein, How writers search: Analyzing the search and writing logs of non-fictional essays, in: Proceedings of the 2016 ACM on conference on human information interaction and retrieval, 2016, pp. 193–202.

[16] A. Hoppe, P. Holtz, Y. Kammerer, R. Yu, S. Dietze, R. Ewerth, Current challenges for studying search as learning processes, in: 7th Workshop on Learning & Education with Web Data (LILE2018), in conjunction with ACM Web Science, 2018.

[17] R. Syed, K. Collins-Thompson, Optimizing search results for human learning goals, Information Retrieval Journal 20 (2017) 506–523.

[18] R. Syed, K. Collins-Thompson, Retrieval algorithms optimized for human learning, in: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, 2017, pp. 555–564.

[19] A. Câmara, D. El-Zein, C. da Costa-Pereira, Rulk: A framework for representing user knowledge in search-as-learning (2022).

[20] D. El Zein, A. Câmara, C. Da Costa Pereira, A. Tettamanzi, Rulkne: Representing user knowledge state in search-as-learning with named entities, in: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, 2023, pp. 388–393.

[21] H. Nasser, D. El Zein, C. da Costa Pereira, C. Escazut, A. Tettamanzi, Rulkkg: Estimating user's knowledge gain in search-as-learning using knowledge graphs, in: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval, 2024, pp. 364–369.

[22] D. El Zein, C. da Costa Pereira, A cognitive agent framework in information retrieval: Using user beliefs to customize results, in: PRIMA 2020: Principles and Practice of Multi-Agent Systems: 23rd International Conference, Nagoya, Japan, November 18–20, 2020, Proceedings 23, Springer, 2021, pp. 325–333.

[23] J. D. Novak, Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations, Routledge, 2010.