Conversational Bibliographic Search

Markus Nilles[1]

[1]Trier University, Universitätsring 15, D-54296 Trier, Germany

Abstract

Finding experts, publications, and topics is a daily task not only of every scientist and student but also for journalists and people who search for sources when consuming information. To support this process, we aim to develop a conversational search engine with which it is possible to search for experts interactively and to explore interesting publications and topics where existing tools reach their limits. An important aspect of the search is that the search query is formulated in such a way that it leads to the desired result. However, formulating a query by a user or understanding a query by a system are challenging tasks. For example, when a query is formulated too unspecific, the search results might not entirely cover the information need whereby small further pieces of information can help immensely. Current systems do little to accurately understand the user's search intent and offer little support during the search process.

Thus, we designed an interactive search engine which runs in a chat window, so that the query can be specified over several turns until the desired search results are obtained. The search engine initiates the conversation by asking the user what they want to search for. The user answers in natural language or can choose adequate answers suggested by the system. The conversation continues until the user has fulfilled their search need or wants to start the conversation from the beginning in order to perform a new search.

Keywords

Conversational Information Retrieval, Conversational Search, Bibliographic Data

1. Introduction

In almost every area of research, it is necessary to find experts and publications for a topic. Whether to form a new research group, to invite scientists to events, or to recommend reviewers, the challenge is to find suitable experts. However, it is not only scientists who need to find experts, but also people who are no experts such as journalists, e.g., to select suitable interview guests on current news-relevant topics, and laypeople who consume information and seek to have sources with experts supporting these information. Another important research task is to find interesting publications or related work. For a scientist, it is important to know the current state-of-the-art in order to contextualize one's own work and to emphasize what is novel and special about one's own work.

However, identifying suitable experts or publications are difficult tasks not only for computers but also for humans. For example, when a user is looking for experts, they often enter a topic into the search engine, which then checks an index to see which people have published on this topic. A problem arises when a user does not make their query specific enough which can occur on purpose, e.g., when the user makes a navigational search as well as without purpose, e.g., when the user lacks knowledge. As a result, the quality of the search results is not very high and the best results may not be found. A user might also not be aware that the scientists are not well-known in the subject area because the topic of the search query is too broad a topic. As an example, a user is looking for an expert for *Fairness in Information Retrieval*, but only enters the query "*Information Retrieval*" in the search engine. Since experts for *Conversational Information Retrieval* are also experts for *Information Retrieval* to a certain extent, they are also included in the search results, even though they are irrelevant for the user.

Popular search engines for scientific papers, such as Google Scholar[1] or Semantic Scholar[2], search large catalogs of publications and also offer the possibility to browse publications of an author in their profile. Furthermore, statistics such as the number of citations of a publication or the h-index of an author can also be displayed. However, the search options are limited, and the user is only assisted to a small extent in fulfilling their search goal, e.g., by displaying related search terms.

To solve the limitations of insufficient attention to the user's search intent and lack of search support, a conversational search engine is indispensable. A conversational search engine assists users in achieving their search intent through a dialogue using natural language. Thereby it should be possible with the search not only to find experts, but also to explore interesting publications or related topics to the search query. The search is to take place via chat interface and can take several turns.

The conversation is started by the system with an introductory question (such as "*Hello, what are you looking*

35th GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), May 22-24, 2024, Herdecke, Germany.
nillesm@uni-trier.de (M. Nilles)
0000-0002-3449-9319 (M. Nilles)

[1]https://scholar.google.com
[2]https://www.semanticscholar.org

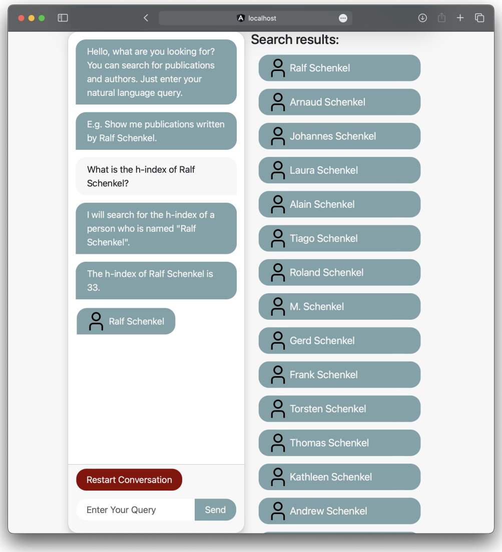

Figure 1: The graphical user interface of Conversational Bibliographic Search with a conversation between the system and a user, in which the user asks an informational question. In the conversation window the system and the user interact in natural language. Results of the retrieval system are displayed in the conversation and additional results are displayed in the result window.

for? You can search for publications and authors. Just enter your natural language query."). After the user entered their natural language query the system infers the user's search intent and uses textual methods to retrieve search results from an index. The search results as well as a natural language answer are displayed to the user. Figure 1 shows the GUI with a sample conversation and search results.

The system response consists of multiple parts. It explains how the system understands the user query so that the user can verify that the system had interpreted their query correctly. When the user asks an informational query, the system provides the response in natural language. During navigational search, the user benefits from clarification questions and follow-up query suggestions from the system to clarify or reformulate the question in order to find interesting publications and authors. For example, if a user searches for publications about *NLP*, the system could ask whether the user wants to search for publications about *natural language processing* or *neurolinguistic programming*.

Besides the presentation of the system architecture we implemented and evaluated three components: User Intent Classification & Slot Filling, Search Module and Conversational Module. The User Intent Classification & Slot Filling component has two tasks. The goal of User Intent Classification is to determine the purpose the user wants to accomplish by using the search engine, e.g., finding experts, publications, or topics. Slot Filling extracts the needs from the user utterance. For a predefined set of slots, e.g., author name or publication title, it determines the values for (slot, value)-pairs. The Search Module uses the determined information of the User Intent Classification & Slot Filling component to query an index and retrieve the data, e.g., persons or publications. The Conversational Module generates the natural language answer for the retrieved data, asks clarification questions and suggests follow-up queries. To evaluate the components we also created a dataset consisting of user utterances in the context of bibliographic search.

In the future, we not only want to evaluate the individual components of this conversational information retrieval system for bibliographic data, but also want to work out the advantages and disadvantages of the conversational information retrieval system for bibliographic data, in comparison to already existing systems that do not support the user in their search process via natural language conversations. Which leads us to the research question: *How beneficial is a conversational information retrieval system for the search of bibliographic data?*

2. Related Work

Chat systems such as ChatGPT[3] or Microsoft Bing's new chat mode[4] make conversations between humans and computers more and more natural, and there are virtually no limits to what computers and humans can talk about. With Bing's new chat mode, Microsoft wants to support the user in web search, so that they can submit his query in natural language. When systems support a user in searching for information through natural language interaction, they are called conversational information retrieval systems or conversational search systems. McTear [1] explains what has led to the current advances in conversational interfaces and why they are an interesting topic today.

Zamani et al. [2] use the definition of conversational search systems by Radlinski and Craswell [3] and provide an overview of definitions, applications, interactions, interfaces, design, implementation, and evaluation of conversational information systems, which include conversational search as well as conversational question answering and conversational recommendation. Goa et al. [4] summarize recent advances in conversational information retrieval with a focus on neural approaches and Zhang et al. [5] discuss recent advances and challenges

[3] https://chat.openai.com/chat
[4] https://www.bing.com/search?q=Bing+AI&showconv=1&FORM=hpcodx

of task-oriented dialogue systems in their survey.

There are few systems that represent an entire conversational information retrieval system. These systems include the above mentioned ChatGPT and Bing's chat mode, as well as XiaoIce (Zhou et al. [6]), a chatbot also developed by Microsoft, which can support users in searching and retrieving information. Similarly, few works like the one by Kaushik et al. [7] exist that study the graphical user interface of conversational search systems. They combined a chat window with an extended standard search interface.

While there exist further works that deal with individual components of a conversational information retrieval system, e.g., user intent classification and slot filling (Louvan and Magnini [8]) or response generation (Lajewska and Balog [9]), there is no system that focuses explicitly on the implementation of a conversational information retrieval system for bibliographic metadata. In contrast to the data of the previous mentioned conversational information retrieval systems, the data of bibliographic metadata is extensive but domain-specific and search queries have a vocabulary corresponding to the domain.

Current search engines for bibliographic metadata, such as dblp[5] [10], ResearchGate[6], Semantic Scholar or Google Scholar allow only keyword-based searches. Kreutz et al. [11] presented SchenQL, a query language for bibliographic metadata that allows users to make their queries more easily and precisely. With the conversational search system the user should be able to search for bibliographic metadata with the support of the system without prior knowledge.

Another important part in the development a conversational information retrieval system is to understand the search behaviour of users. Kuhlthau [12] created a seven step model of users information search process. In another study, Kuhlthau [13] observed students in their search process while they were in high school and again four years later when they were in college to examine changes in their search behavior.

3. System Architecture

The preliminary architecture of the conversational search system for bibliographic data is shown in Figure 2. It consists of four main components: User Intent Classification & Slot Filling, Search Module, Conversational Module, and Conversation History Module.

The task of User Intent Classification & Slot Filling is to determine the search intent of the user and to extract the information given by the user to fulfill their intent. A search intent can be that the user is searching for persons, publications or topics. The User Intent Classification &

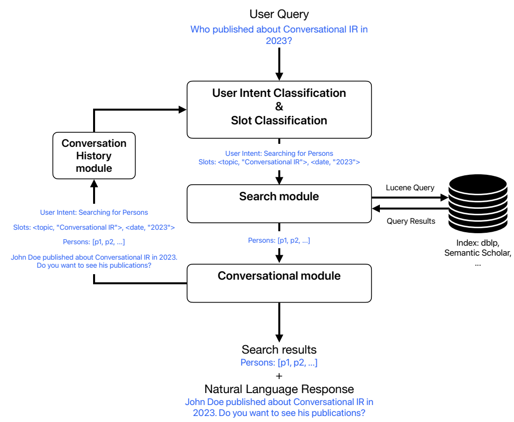

Figure 2: Architecture of the conversational search system for bibliographic metadata.

Slot Filling module consists of a BERT [14] model which takes the user utterance and the conversation history of the session as the input, and outputs an user intent label for the whole utterance and a slot label for each token of the utterance. For the slot labeling the BIO-tagging format is used. A token can be classified as the **B**eginning of a slot value, as **I**nside of a slot value, or as **O**utside of a slot value.

With the determined user intent and (slot, value)-pairs, the Search Module searches an index which contains the bibliographic data. The slots are mapped to different fields in which a query processor searches for the corresponding values.

The Conversation History Module stores the user queries as well as the system responses to update the dialogue state after each turn. The dialogue state is considered by the User Intent Classification & Slot Filling component to predict the user intent and slot values of the next user utterance. The user utterances can also be used to analyze the user's search tasks and to improve the system.

4. Preliminary Implementation

In this section we give an overview of the progress we made so far and describe the implemented components of the system architecture (Section 3) in more detail. To evaluate the components, we built a dataset of user utterances (Section 4.1). The dataset is used to train and evaluate the User Intent Classification & Slot Filling component (Section 4.2). Initially, the conversational information retrieval system supports 29 predefined user intents (17 for searching publications and 12 for searching persons) and 33 predefined slots (16 from publications and 17 from

[5] https://dblp.uni-trier.de
[6] https://www.researchgate.net

persons).

With the obtained information from the User Intent Classification & Slot Filling component an index containing the bibliographic data (Section 4.3) is queried by the Search Module (Section 4.4). The Conversational Module (Section 4.5) uses the results from the User Intent Classification & Slot Filling component and the results from the Search Module to generate a natural language answer to the user.

4.1. User Utterances

Because there is no dataset of user utterances in the domain of search for bibliographic data with labelled slots for the tokens of the user utterances, we created a dataset containing 620 user utterances.

For the predefined set of user intents and slots, we randomly combined a user intent with a slot. We created one (user intent, slot)-pair for each user intent and one (user intent, slot)-pair for each slot. We formulated for each pair a user utterance, a user of a conversational information retrieval system might ask a conversational search system for bibliographic data. E.g., we formulated for the pair (`publication`, `publication.topic`) the utterance *"Who has published a paper about Conversational Information Retrieval?*.

Then we used ChatGPT to rephrase these utterances and checked the returned reformulations. We made sure that the reformulated utterances still have the same meaning and contain the same slots as the given utterances. We added nine reformulations and the original utterance to the dataset. In total the dataset contains for each utterance ten different formulations.

The slot values of each formulation are filled with random values from the bibliographic database, e.g., the `publication.topic` slot is filled by keywords contained in the database. To allow the system to recognize questions that are not related to any of the predefined intents, we added questions from the Quora-Question-Pairs dataset[7]. We used the created dataset to train and evaluate the User Intent Classification & Slot Filling component.

4.2. User Intent Classification & Slot Filling

The User Intent Classification & Slot Filling module has two tasks. The first task is to determine the user intent from the user utterances and the second task is to extract the slot values from the user utterances. We used a joint intent classification and slot filling model based on BERT [14, 15].

[7]https://www.kaggle.com/datasets/quora/question-pairs-dataset

To determine the user intent and the slot values, the BERT model adds a special token at the beginning of the user utterance. The output for the special token is the user intent label of the user utterance. For each token of the user utterance the slot label is returned. A token can either be classified as the **B**eginning of a slot value, as **I**nside of a slot value, or as **O**utside of a slot value. The BERT model achieves an accuracy of 0.955 for predicting the user intent. The f1-score for predicting slot values is 0.968.

4.3. Bibliographic Database

As a database, we use the dblp and extended it with data from standard data sources such as Semantic Scholar. The dblp is a bibliography containing information of computer science journals and proceedings. We enriched the dblp data with information retrieved from the Semantic Scholar API[8]. The Semantic Scholar API provides information about publications and persons which is not available in the dblp, e.g., the abstract, a summary or academic categories of the publication. This information will be displayed to the user. We use the dblp xml file[9] to build a Lucene[10] index containing the dblp data and the retrieved information from Semantic Scholar. The index is queried by the Search Module.

4.4. Search Module

The Search Module queries a Lucene index containing the bibliographic data. The results of the User Intent Classification & Slot Filling determine how the query will be build. The query consists of multiple subqueries. For each slot a subquery will be added to the query if the User Intent Classification & Slot Filling module detected values for this slot. The subquery of each slot will then search in one or multiple fields of the index. The results will be displayed to the user.

4.5. Conversational Module

The Conversational Module generates the natural language answer of the system. It uses templates to generate the answer. The determined user intent and slots by the User Intent Classification & Slot Filling module and the retrieved results by the Search Module are inserted in the templates. Currently, we use templates to summarize the results of the User Intent Classification & Slot Filling components, and to formulate the natural language response to the user request.

[8]https://www.semanticscholar.org/product/api
[9]https://dblp.org/xml/
[10]https://lucene.apache.org/core/

5. Discussion & Future Work

We discussed the idea of combining the search of bibliographic metadata by means of a conversational retrieval system. We presented an architecture of such a system and so far, we implemented three components of the system: User Intent Classification & Slot Filling, Search Module and Conversational Module. Next, we will implement the remaining component, the Conversation History Module. The Conversation History Module tracks the dialogue state of the conversation between the system and the user. It incorporates previous turns of the conversation into the User Intent Classification & Slot Filling component. With the current dialogue state from the conversation history and a new user utterance the system is able to detect changes in the user intent and the (slot, value)-pairs. To train the system for multi-turn conversations, we will create a multi-turn conversation dataset for bibliographic search through studying how a conversation between a user and conversational search system might evolve during a search session. Each turn of the dataset's conversations will be annotated with the current user intent and (slot, value)-pairs.

Besides the evaluation of the individual components of our proposed conversational information retrieval system Conversational Bibliographic Search, we will evaluate the system as an entirety in user studies. To examine the usefulness of a conversational retrieval system for bibliographic data and to answer the research question, we plan to evaluate our system by comparing it to existing bibliographic search engines in terms of effectiveness, efficiency, and user satisfaction. We will also compare our system with approaches that use Large Language Models (LLMs). Because of the high computation and storage cost we are currently not planning to train a single LLM for bibliographic search. Another disadvantage of using a single LLM instead of our approach could be the problem of LLM hallucination.

Instead, we want to explore the extent to which LLMs can be used in each component. For example, we could use LLMs to generate the natural language response. LLMs could also be trained to ask clarification questions and to suggest follow-up queries to the user.

In user studies, we want to identify how a conversational information retrieval system can help users to fulfill their information need. We also want to gain insights into the information tasks of different user groups, e.g., students vs. more advanced researchers. After identifying user tasks, we examine for which task users benefit the most from an information retrieval system supporting conversational search and infer from these results for which user group a conversational information retrieval system would be most useful.

References

[1] M. F. McTear, The rise of the conversational interface: A new kid on the block?, in: J. F. Quesada, F. Martín-Mateos, T. López-Soto (Eds.), Future and Emerging Trends in Language Technology. Machine Learning and Big Data - Second International Workshop, FETLT 2016, Seville, Spain, November 30 - December 2, 2016, Revised Selected Papers, volume 10341 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 38–49. URL: https://doi.org/10.1007/978-3-319-69365-1_3. doi:10.1007/978-3-319-69365-1_3.

[2] H. Zamani, J. R. Trippas, J. Dalton, F. Radlinski, Conversational information seeking, CoRR abs/2201.08808 (2022). URL: https://arxiv.org/abs/2201.08808. arXiv:2201.08808.

[3] F. Radlinski, N. Craswell, A theoretical framework for conversational search, in: R. Nordlie, N. Pharo, L. Freund, B. Larsen, D. Russel (Eds.), Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR 2017, Oslo, Norway, March 7-11, 2017, ACM, 2017, pp. 117–126. URL: https://doi.org/10.1145/3020165.3020183. doi:10.1145/3020165.3020183.

[4] J. Gao, C. Xiong, P. Bennett, N. Craswell, Neural approaches to conversational information retrieval, CoRR abs/2201.05176 (2022). URL: https://arxiv.org/abs/2201.05176. arXiv:2201.05176.

[5] Z. Zhang, R. Takanobu, Q. Zhu, M. Huang, X. Zhu, Recent advances and challenges in task-oriented dialog systems, Science China Technological Sciences 63 (2020) 2011–2027.

[6] L. Zhou, J. Gao, D. Li, H. Shum, The design and implementation of xiaoice, an empathetic social chatbot, Comput. Linguistics 46 (2020) 53–93. URL: https://doi.org/10.1162/coli_a_00368. doi:10.1162/coli_a_00368.

[7] A. Kaushik, V. B. Ramachandra, G. J. F. Jones, An interface for agent supported conversational search, in: H. L. O'Brien, L. Freund, I. Arapakis, O. Hoeber, I. Lopatovska (Eds.), CHIIR '20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020, ACM, 2020, pp. 452–456. URL: https://doi.org/10.1145/3343413.3377942. doi:10.1145/3343413.3377942.

[8] S. Louvan, B. Magnini, Recent neural methods on slot filling and intent classification for task-oriented dialogue systems: A survey, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020, International Committee on Computational Linguistics, 2020, pp. 480–496. URL: https://doi.org/10.18653/v1/2020.coling-main.42. doi:10.

18653/V1/2020.COLING-MAIN.42.

[9] W. Lajewska, K. Balog, Towards filling the gap in conversational search: From passage retrieval to conversational response generation, in: I. Frommholz, F. Hopfgartner, M. Lee, M. Oakes, M. Lalmas, M. Zhang, R. L. T. Santos (Eds.), Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023, ACM, 2023, pp. 5326–5330. URL: https://doi.org/10.1145/3583780.3615132. doi:10.1145/3583780.3615132.

[10] M. Ley, DBLP - some lessons learned, Proc. VLDB Endow. 2 (2009) 1493–1500. URL: http://www.vldb.org/pvldb/vol2/vldb09-98.pdf. doi:10.14778/1687553.1687577.

[11] C. K. Kreutz, M. Wolz, J. Knack, B. Weyers, R. Schenkel, Schenql: in-depth analysis of a query language for bibliographic metadata, Int. J. Digit. Libr. 23 (2022) 113–132. URL: https://doi.org/10.1007/s00799-021-00317-8. doi:10.1007/s00799-021-00317-8.

[12] C. C. Kuhlthau, Inside the search process: Information seeking from the user's perspective, J. Am. Soc. Inf. Sci. 42 (1991) 361–371.

[13] C. C. Kuhlthau, Perceptions of the information search process in libraries: a study of changes from high school through college, Inf. Process. Manag. 24 (1988) 419–427. URL: https://doi.org/10.1016/0306-4573(88)90045-3. doi:10.1016/0306-4573(88)90045-3.

[14] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), 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, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 4171–4186. URL: https://doi.org/10.18653/v1/n19-1423. doi:10.18653/V1/N19-1423.

[15] Q. Chen, Z. Zhuo, W. Wang, BERT for joint intent classification and slot filling, CoRR abs/1902.10909 (2019). URL: http://arxiv.org/abs/1902.10909. arXiv:1902.10909.