From Data to Dialogue: Leveraging the Structure of Knowledge Graphs for Conversational Exploratory Search

Phillip Schneider[1], Nils Rehtanz[1], Kristiina Jokinen[2] and Florian Matthes[1]

[1]Technical University of Munich, Department of Computer Science, Germany

[2]AI Research Center, National Institute of Advanced Industrial Science and Technology, Japan

{phillip.schneider, nils.rehtanz, matthes}@tum.de

kristiina.jokinen@aist.go.jp

Abstract

Exploratory search is an open-ended information retrieval process that aims at discovering knowledge about a topic or domain rather than searching for a specific answer or piece of information. Conversational interfaces are particularly suitable for supporting exploratory search, allowing users to refine queries and examine search results through interactive dialogues. In addition to conversational search interfaces, knowledge graphs are also useful in supporting information exploration due to their rich semantic representation of data items. In this study, we demonstrate the synergistic effects of combining knowledge graphs and conversational interfaces for exploratory search, bridging the gap between structured and unstructured information retrieval. To this end, we propose a knowledge-driven dialogue system for exploring news articles by asking natural language questions and using the graph structure to navigate between related topics. Based on a user study with 54 participants, we empirically evaluate the effectiveness of the graph-based exploratory search and discuss design implications for developing such systems.

1 Introduction

Dialogue systems, or conversational interfaces, have been a long-established research area in natural language processing (NLP). Over several decades, their capabilities have continuously improved, and due to recent advances in neural network-based methods, they have progressed even further. Today, we see widespread adoption of dialogue systems across a plethora of domains, including customer service, education, healthcare, home automation, and finance, to name a few (Maedche et al., 2019). There is currently significant public interest in auto-regressive large language models, such as PaLM or GPT-3 and its successors Chat-GPT and GPT-4 (Brown et al., 2020; Chowdhery et al., 2022; OpenAI, 2022). These generative models are fine-tuned with instructions and can produce
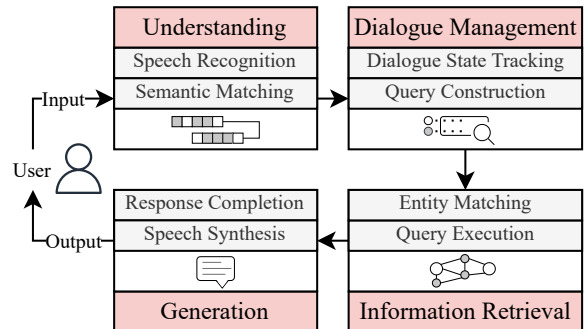

Figure 1: Conceptual architecture of the exploratory search system.

human-like responses for a variety of prompts, but they are susceptible to generating text that lacks factual grounding in credible information sources. Nevertheless, this trend is another driving force behind the paradigm shift in moving human-machine interactions toward the medium of language.

Although most dialogue systems can be categorized as either task-based, performing specific functions like controlling smart devices, or non-task-based, engaging in social chit-chat, another use case that is gaining research attention is the application of conversational interfaces for information retrieval. While traditional search systems are limited to single-turn keyword query-response or question-answer interactions, conversational search systems support multi-turn information-seeking dialogues (Radlinski and Craswell, 2017). Hence, conversational interfaces are particularly suitable for exploratory search settings, in which information seekers have a vague intention of what they are looking for because they are unfamiliar with a complex and multifaceted topic (White and Roth, 2009). Conversational search systems enable users to refine their queries iteratively, evaluate search results, give feedback, and narrow down the information space. This intuitive, dialogue-based interaction theoretically promises to overcome long-known information retrieval issues like vocabulary mis-

match or the anomalous state of knowledge, which restricts users from forming meaningful queries without prerequisite knowledge in an unfamiliar domain (Belkin, 1980; Furnas et al., 1987).

Knowledge graphs can capture the interconnected nature of domain-specific knowledge and have proven to be a powerful data structure for dialogue systems and numerous NLP tasks (Ji et al., 2022; Schneider et al., 2022). As a rich representation of entities and semantic relationships, scholars have studied them to improve dialogue systems by enabling contextual utterance understanding, facilitating decision-making, or enhancing retrieval-based response generation. There are several reasons why knowledge graphs are suitable for exploratory search. For example, knowledge graphs can help with the disambiguation of complex search queries, which is helpful for topics where the meaning of terms may vary depending on context. In addition, the graph structure allows users to navigate through related concepts and explore different facets of a topic. Because these connections approximately reflect associative thinking patterns of human cognition, they can aid serendipitous discovery and novel insights into topics (Sarrafzadeh et al., 2014).

Despite the growing body of research on combining conversational interfaces and knowledge graphs, there is an apparent lack of studies focusing on exploratory search. While a few studies mention the theoretical potential of these technologies for browsing and discovery, the effectiveness of implemented dialogue systems in real-world settings has yet to be studied thoroughly. To address this gap, we developed a knowledge-driven conversational system as a baseline and evaluated it in-depth with a sample of 54 users. We opted for the application domain of news search because it involves browsing numerous topics that change daily, as elaborated upon in Schneider et al. (2023b). At the time of this study, there are no existing commercial or scientific solutions available that support conversational news exploration. Our main contributions are three-fold:

1. We propose exploratory search mechanisms based on the structure of knowledge graphs.

2. We developed a conversational search system for exploring a graph with news articles.

3. We empirically evaluated the developed prototype and publish our source code and collected data in a GitHub repository.[1]

2 Related Work

Integrating knowledge graphs into dialogue systems has been a subject of interest within the research community for many years. Knowledge graphs have the potential to enhance utterance understanding, response generation, and dialogue management. Focusing on the first area, Zhou et al. (2020) employ word- and concept-oriented graphs to better understand short user utterances through contextual enrichment. The authors demonstrate with experiments that their model outperforms existing approaches on recommendation tasks. Similarly, Wang et al. (2020) propose TransDG, a model transferring abilities of question representation and knowledge matching from the task of knowledge-based question answering to improve utterance understanding and response generation. Numerous other studies have investigated using structured information in graphs to ground responses generated by language models (Chen et al., 2019; Zhang et al., 2020; Chaudhuri et al., 2021).

While understanding and generating dialogues play crucial roles, knowledge graph-based dialogue management is especially relevant to conversational search because it is responsible for handling information processing tasks. In a study from Sarkar et al. (2020), a movie recommendation system is proposed that generates suggestions from constructed subgraphs by considering dialogue states and user profiles. Another closely related work is WikiTalk from Wilcock (2012) who introduces a dialogue-based knowledge access system for Wikipedia articles. Although the system does not rely on a knowledge graph with semantically related entities, it uses hyperlinks extracted from articles to transition between topics.

Diverging from previous studies, our research explicitly focuses on the phenomenon of exploratory search, where users are confronted with unfamiliar territories of information (Schneider et al., 2023a). To the best of our knowledge, we are the first to conduct a practical evaluation of how knowledge graphs can be exploited to address the unique challenges posed by exploratory search, providing a first baseline system. Our proposed voice-based conversational agent leverages the semantic graph structure to compress the search space through entity-centric filtering within a targeted scope and

---

[1] github.com/sebischair/kg-conv-exploratory-search

expand the search space by suggesting related entities, familiarizing users with the knowledge domain, and offering opportunities for further exploration beyond their initial query.

3 Knowledge Graph and Conversational Search System Architecture

3.1 Knowledge Graph Construction and Graph-Based Search

To create the graph-based exploratory search system, we gathered a news article corpus from the Tagesschau news program.[2] We automatically scraped and organized all news articles into a knowledge graph structure. Concerning the data model displayed in Figure 2, there are distinct node types for articles, news categories, and entities (e.g., Barack Obama), the latter being instances of entity classes (e.g., human). Each article is part of one news category, such as politics or sports, but can be linked to multiple entity nodes. To derive entities mentioned in a given article text, we performed named entity recognition and linking to the target knowledge base Wikidata.[3] We make use of a wikification model from Brank et al. (2017), which supports multiple languages and provides language-agnostic Wikidata identifiers for the graph, making it feasible to adapt the system to other languages, such as Chinese, Spanish, or English.
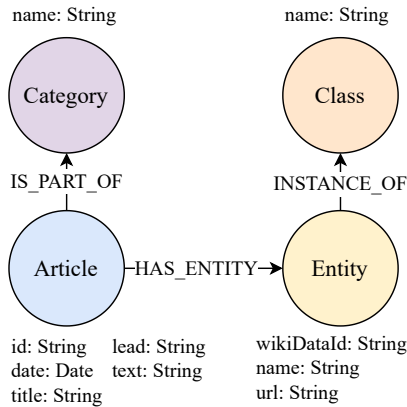
Figure 2: Graph structure of semantic data model.

Over a period of nearly two months, 2118 German news articles were continuously added to the knowledge graph, which comprises a total of 6262 nodes and 19125 relationships. Additional information about the number of nodes and relationships can be found in Figure 5 in Appendix A.

Building upon the constructed news knowledge graph, we can leverage graph-based search mechanisms for the dialogue system. Figure 3 illustrates the possible information-seeking paths as a finite-state machine with eight states (S1-S8). First, a user can initiate the conversation by greeting the agent (S1) or asking for help (S2). Thereafter, the agent introduces three search mechanisms (S3), namely, getting an overview of recent popular articles (S4), searching articles by news category (S5), and entity-based search (S6). We opted to incorporate these three levels of abstraction in order to cover varying degrees of specificity. The agent queries the graph to retrieve articles relevant to the user's search intent and suggests three articles by listing their titles. Next, the user can navigate (S7) by either asking for additional articles or selecting a specific one. The navigation options also include the ability to skip, stop, or repeat articles being read at the moment. Finally, when the agent has finished reading a selected article, it suggests related entities (S8) as a possibility to explore more themes. This enables users to acquire knowledge about existing entities, discover related content, and deepen their understanding of a specific topic.

For each of the eight dialogue states, we defined intents and related entities along with training phrases that were used to fine-tune pre-trained intent detection and entity extraction models from Google. To suggest related entities for a given news article, a count-based suggestion algorithm was used, which is described below in pseudo-code.

Algorithm 1 Entity Suggestion Algorithm

Require: news article A, knowledge graph G
1: $linked\_entities \leftarrow$ get_article_entities(A, G)
2: $entity\_freq \leftarrow$ init_empty_dict()
3: **for** e in $linked\_entities$ **do**
4:     $entity\_freq[e] \leftarrow$ get_article_count(e, G)
5: **end for**
6: $sorted\_entities \leftarrow$ sort_by_freq($entity\_freq$)
7: $result\_entities \leftarrow$ get_top_n($sorted\_entities$, n)
8: **return** list of n suggested entities $result\_entities$

3.2 Architecture of Conversational Search System

The system architecture, depicted in Figure 1, consists of four core components for natural language understanding, dialogue management, information retrieval, and natural language generation. It is implemented as a web application using open-source Python libraries and Google Cloud models. Since the interaction happens through a voice in-
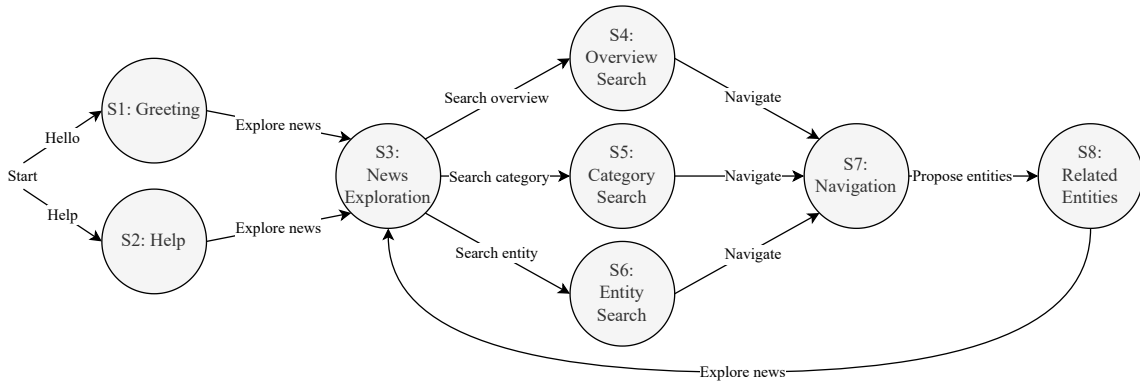
Figure 3: Finite-state model of dialogue states for conversational exploratory search system.

terface, we use pre-trained speech-to-text and text-to-speech models from Google. The front-end is a web application with a single button for activating the microphone. When a user initiates the conversation by clicking the button, the recorded audio is transcribed with the speech-to-text model and sent to the natural language understanding component, which is built with the conversational agent framework Google Dialogflow.[4] This framework is responsible for producing a semantic representation of the user input by classifying intents, recognizing entities, and keeping track of the dialogue state.

The dialogue management, which uses the dialogue state and context to determine the system's actions, is implemented as a Python-based conversation fulfillment service with Flask. After matching the semantic representation of the user input to query templates, the dialogue manager uses the extracted entities and links them to entities present in the knowledge graph to construct graph queries. The knowledge graph itself is stored as a property graph in a Neo4j database.[5] The queries are executed, and the results are combined with a text template in a probabilistic manner to complete a response. Lastly, the text-to-speech model generates an audio file from the response text that is automatically played back to the user. Despite the system's various technical components depicted in Figure 6, the voice interface hides the complexity of the underlying graph-based information retrieval, providing a highly accessible search experience.

4 System Evaluation Results

The goal of the evaluation was to determine the effectiveness of the proposed graph-driven dialogue

system for exploratory search. To achieve this, we conducted an extensive human evaluation focused on the example scenario of a news search. Collecting empirical observations helped to reveal insights into technical limitations and how the conversational search system was utilized by the study participants in a practical setting.

4.1 Study Sample

A diverse sample of participants was acquired by recruiting test users through university courses, social media platforms, and snowball sampling to expand the pool of test users further. The voice-based dialogue system was made available through a minimalist web interface with one button to activate the microphone. Participants received no concrete instructions on how to use the agent or the existing news articles since we wanted to assess the conversational agent's self-explanatory capabilities. They were only told to greet the agent, seek information about news, and afterward report their experience in an online questionnaire. Additionally, we logged all dialogue interactions between the conversational agent and the participants.

The gender distribution among the study's participants was fairly balanced, with 43% being female and 57% male. The participants' ages ranged from 14 to 86, encompassing different age groups. The mean age was 36, and young adults between 20 and 30 accounted for the biggest group. We also inquired about the participants' familiarity with virtual assistants. Of those surveyed, 76% of respondents had previous experience, but those above 50 had considerably less experience. Even though almost all participants consumed daily news, and many were familiar with virtual assistants, only 2 out of 54 had used virtual assistants to ask about the news. Instead, most respondents preferred other

---

channels, such as the Internet, radio, television, or physical newspapers.

Consequently, the evaluation study included a diverse sample of participants across different age groups and genders, where most never had used conversational agents for news search. This renders our participant pool an ideal foundation for evaluating how accessible and effective the proposed conversational exploratory search system is.

4.2 Conversation Log Analysis

An important part of our empirical evaluation involved logging and analyzing the dialogues between the conversational agent and test users. Over a span of two weeks, 54 participants acted as test users of the exploratory news search system. Because a few users started more than one conversation, we collected 65 dialogues with a total sum of 2172 dialogue turns, where one turn refers to a single utterance by either the agent or user. On average, each dialogue session comprised 33.4 turns. The mean duration of a dialogue session was around 9 minutes and 11 seconds, with a standard deviation of 7.8 minutes. The longest recorded conversation lasted for roughly 41 minutes, and the shortest for 22 seconds.

Concerning the three different search intents, entity-based search was by far the most commonly detected intent, with a share of 10.2% of all dialogue turns, followed by searching for a general overview and searching for news categories with 2.2% and 0.9%, respectively. Navigating intents, such as selecting or skipping articles, were also frequently used, making up 15.8% of the turns. Whenever the agent could not determine the intent of a user's utterance, it replied with a fallback clarification response. To give an example, some users asked for information about the weather or about what a specific person did on a particular day, although we did not include intents for these types of requests. In other cases, issues arose due to nonsensical audio transcripts generated by the speech-to-text model. Overall, the fallback response appeared 192 times in the logs, accounting for a share of 8.8%.

Since entity-based search proved to be the most popular search mechanism in the conversation logs, we examined which specific entities users were looking for. Table 1 lists the number of the top three recognized and unrecognized entity classes, as well as their accuracy scores, indicating how often the

conversational agent could match them correctly to the graph entities. Of the 222 entity-related search requests in the logs, the conversational agent successfully recognized 130 entities (59%) and linked them to entities in the news knowledge graph. Entities of the class country appeared most often and were recognized with a very high accuracy of 94%. Similarly, classes like person or city also had high accuracy scores of at least 70%. Considering the unrecognized entities, we see that entities of the classes sports, person, and nutrition were among the most common, although they constituted a minority compared to the long tail of other classes. There were numerous topics that users searched for that were not covered by the Tagesschau news portfolio. These included news about celebrities, computer games, local events, and weather reports. A detailed discussion of reasons regarding issues with unrecognized entities is part of Section 5.

Entity Types	Count	Accuracy
Recognized Entities	130	0.59
Country	53	0.94
Person	19	0.70
City	14	0.78
Other classes	44	-
Unrecognized entities	92	0.41
Sports	10	0.17
Person	8	0.70
Nutrition	7	0.29
Other classes	67	-

Table 1: Count and accuracy of recognized and unrecognized entity classes.

Our underlying hypothesis regarding the combination of knowledge graphs and conversational agents is that the connections within the graph structure can be leveraged for exploration purposes. In theory, human thinking patterns revolve around mental associations between concepts (Morewedge and Kahneman, 2010), which knowledge graphs can model as directed edges between two entities. Through our comprehensive system evaluation, we gathered numerous dialogues that serve as evidence of participants intuitively using the suggested linked entities to explore related articles on a given subject. To illustrate this point, Figure 4 shows one exemplary conversation transcript translated from German to English. The corresponding knowledge graph entities are shown next to the conversation logs.

The dialogue begins with the user's news request about France. In response, the system uses entity-based search to retrieve three articles linked to the

State (S)	User Utterance (U) and Agent Reponse (A)	Knowledge Graph

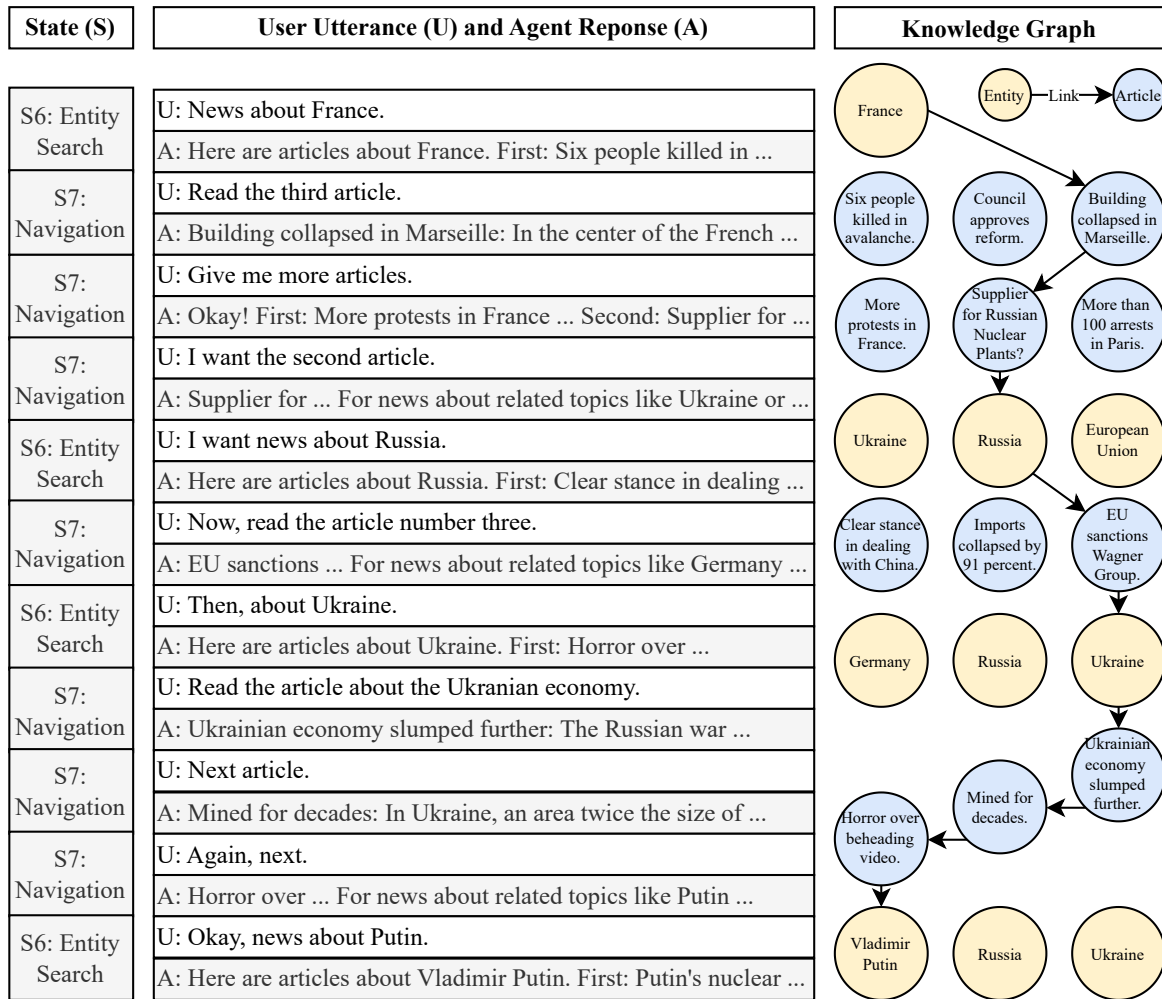| State (S)	User Utterance (U) and Agent Reponse (A)
S6: Entity Search	U: News about France.
	A: Here are articles about France. First: Six people killed in ...
S7: Navigation	U: Read the third article.
	A: Building collapsed in Marseille: In the center of the French ...
S7: Navigation	U: Give me more articles.
	A: Okay! First: More protests in France ... Second: Supplier for ...
S7: Navigation	U: I want the second article.
	A: Supplier for ... For news about related topics like Ukraine or ...
S6: Entity Search	U: I want news about Russia.
	A: Here are articles about Russia. First: Clear stance in dealing ...
S7: Navigation	U: Now, read the article number three.
	A: EU sanctions ... For news about related topics like Germany ...
S6: Entity Search	U: Then, about Ukraine.
	A: Here are articles about Ukraine. First: Horror over ...
S7: Navigation	U: Read the article about the Ukranian economy.
	A: Ukrainian economy slumped further: The Russian war ...
S7: Navigation	U: Next article.
	A: Mined for decades: In Ukraine, an area twice the size of ...
S7: Navigation	U: Again, next.
	A: Horror over ... For news about related topics like Putin ...
S6: Entity Search	U: Okay, news about Putin.
	A: Here are articles about Vladimir Putin. First: Putin's nuclear ...

Figure 4: Visualization of entity-centric search history in constructed news knowledge graph.

entity France. The user chooses the third headline, and the agent reads the article on a collapsed building in Marseille. Then, the user expresses the desire to get more articles, leading the system to provide three additional articles related to France. Among these, the user selects the second article focusing on a French company's plans to supply nuclear fuel rods for Russian power plants. Subsequently, the agent's algorithm suggests the three most relevant entities mentioned in the article: Ukraine, Russia, and European Union. At this point, the user wants news about Russia and opts for an article concerning EU sanctions against the Russian Wagner Group. Following this, the agent once again suggests three entities and the user requests news related to the country of Ukraine. After skipping two articles, the user listens to the third one about a beheading video of a Ukrainian soldier and receives three related entities again. Continuing the exploratory news search, the user requests news about Russian President Vladimir Putin.

4.3 Human Evaluation

After the participants tested the conversational news exploration, they were asked to complete an evaluation questionnaire. The goal was to document their usage experiences and perceived satisfaction with the system. The questionnaire comprised both open-ended and closed-ended questions. For example, it included free-form text fields to share experienced issues and suggestions for improvement. In addition, questions about overall satisfaction, quality of recommendations, the relevance of information content, and other aspects were asked. These were rated on a Likert scale from 1 to 5, with 1 representing "very poor" and 5 "very good". Furthermore, participants were asked questions from the System Usability Scale (SUS), an established method developed to determine the usability of an application (Brooke, 1996). In this, ten questions are asked on a 5-point Likert scale, ranging from 1 ("strongly disagree") to 5 ("strongly

agree"). The SUS score is calculated using a mathematical formula, and its resulting value ranges from 0 to 100, where a value of 68 is considered average. It is important to note that the SUS scores of two given applications are not directly comparable.

Table 2 lists the mean and standard deviation for the evaluation criteria of the questionnaire. Respondents rated their overall satisfaction score at 3.7, indicating a generally positive perception of the conversational search agent. The relevancy of the provided information received a mean value of 3.8 and strongly correlated with the satisfaction level. Ratings for the control over the dialogue and the agent's suggestions were also favorable, with scores of 3.7 and 3.8, respectively. While respondents rated the agent's capability to correctly understand requests a bit lower at 3.5, they appreciated the agent's comprehensible responses, resulting in the highest average rating of 4.4. In contrast, the perceived human-likeness of the agent got a mean value of 2.8, which is the lowest rating. Regarding the conversational search system's usability, we measured a SUS score of 79.4, surpassing the considered average of 68. It is notable that the age of participants had a significant influence on perceived usability, with those below the sample mean age of 36 giving an average SUS score of 86, compared to 68.3 for participants above 36. This aligns with the observation that older participants reported more difficulties using the system, which was reflected in lower general satisfaction scores. Besides the mentioned correlation between information relevance and overall satisfaction, the latter was also positively correlated with both the agent's suggestion and the control over the dialogue. In our analysis, the questionnaire data showed neither implausible correlations nor unexpected outliers, indicating that the users had no problems understanding and answering the questions.

Evaluation Criteria	Mean	SD
Overall satisfaction	3.7	1.2
Information relevancy	3.8	1.1
Control over dialogue	3.7	1.2
Understanding of requests	3.5	1.3
Comprehensible responses	4.4	0.9
Human-likeness of agent	2.8	1.1
Suggestions of agent	3.8	1.0
System Usability Scale (SUS)	79.4	18.5
SUS (age < 36)	86.0	13.3
SUS (age > 36)	68.3	21.0

Table 2: Overview of evaluation criteria from the questionnaire with mean and standard deviation (SD).

5 Discussion

Our proposed conversational search system was built upon the hypothesis that combining conversational interfaces and knowledge graphs can support exploratory information-seeking. In order to examine the synergistic effects between these two technologies, we developed a baseline system for the application domain of exploring news articles. We conducted an extensive user study with a diverse pool of participants to evaluate the system, gathering empirical observations of how the test users interacted with the conversational agent to explore the underlying news knowledge graph.

One of the key insights from the questionnaire results was that participants generally enjoyed using the system. They found it easy to understand the system's functionalities and expressed a strong liking for its intuitive control through voice commands. This was reflected in a high SUS score of 79.4, indicating a favorable user experience. In the questionnaire's free-text responses, participants emphasized the accessibility of the voice interface, noting its benefits for disabled individuals and situations where typing is impractical, such as when driving a car. Besides, the feedback revealed other aspects that contributed to the overall positive reception of the system. In particular, users appreciated the freedom to search among a plethora of topics and select from related suggestions relevant to their interests. These findings support our hypothesis that the underlying knowledge graph structure can facilitate the selection of relevant information and enhance the information-seeking experience.

However, we also identified certain issues encountered by some users regarding the search-centered dialogue. A subset of participants reported facing technical difficulties with their microphones, which coincided with problems with the agent's understanding of requests and inaccurate retrieval of articles. Interestingly, these issues were more commonly mentioned by older users above the age of 50, who tended to give worse usability scores, too. We assume that these users might have had lower technical literacy and were less familiar with using voice assistants. This observation aligns with related studies involving elderly individuals (Pradhan et al., 2020; Kim and Choudhury, 2021). Additionally, respondents expressed a desire to have more news sources than solely the Tagesschau because they were interested in topics about fashion, music, and local events. Furthermore, participants

Issue Group	Description	Example Dialogue Excerpt
Acronyms	Entities were misrecognized as acronyms, but recognized correctly when spelled out in full.	User: News about "BVB" (football club). Agent: Sorry, I cannot find news for this entity.
Disambiguation	While entities were correctly identified, they were sometimes recognized within the wrong context in cases of word ambiguity.	User: Give me news about the "Wolf" (animal). Agent: Here is news about "Wolf" (article of person with surname Wolf).
Foreign language	Entities in a foreign language were often incorrectly transcribed by the speech-to-text model.	User: I want news about "chaditip" (ChatGPT). Agent: Here is news about "Charité" (hospital).
Fuzzy matching	Sometimes, using fuzzy matching resulted in linking to wrong entities in the knowledge graph.	User: News about "Justin Bieber" (person). Agent: Here is news about "Biber" (animal).
Multiple entities	If two or more entities were mentioned simultaneously, only one entity was recognized.	User: News about "China" and "Taiwan" (city). Agent: Here is news about "China" (city).

Table 3: Identified issues regarding entity recognition and linking in exploratory search dialogues.

suggested improvements in the speech generation of the agent. For instance, they proposed incorporating more diverse responses and enhancing the pronunciation of the German text-to-speech models, especially for English terms.

Another important finding pertains to the effectiveness of entity-based search for discovering news. Despite participants initially having no knowledge about available articles, they intuitively requested news for specific entities of interest. Then, the agent's suggestions of related entities played a crucial role in reinforcing the exploration of related topics. We have logged multiple dialogues that demonstrate this navigation between entities and linked news articles. One concrete example is shown in Figure 4. The dialogue transcript provides tangible evidence of how conversational agents can leverage the semantic graph structure of interlinked entities and a finite set of dialogue states for open-ended exploratory search, helping users overcome the hurdle of not knowing precisely what to ask by offering topics for further exploration.

On one side, our analysis revealed a good performance on correctly matching entities from certain classes, such as cities, countries, and persons, which were also the most frequently requested (see Table 1). On the other side, searching for entities was inherently prone to issues due to the virtually infinite number of possible entity-specific queries, whereas the knowledge graph contained only 2874 entities. To gain a better overview of the problems regarding entity matching in exploratory search dialogues, we identified five issue groups as presented in Table 3. Firstly, when entities were mentioned as acronyms, they were usually misrecognized until they were spelled out in full. Secondly, despite entities being correctly identified, there were a couple of cases where they matched within the wrong context, particularly in situations with word ambiguity.

Thirdly, entities expressed in foreign languages often suffer from transcription errors by the speech-to-text model, resulting in recognition inaccuracies. Fourthly, using fuzzy matching occasionally led to the incorrect matching of entities from the article to the knowledge graph. Lastly, when multiple entities were mentioned simultaneously, only one entity was recognized, potentially omitting relevant information. These findings shed light on the conversational exploratory search system's current limitations and point out areas for improvement in entity recognition and linking.

6 Conclusion

In this study, we propose a conversational exploratory search system with a news knowledge graph. It leverages the graph structure to retrieve relevant news articles and suggest related ones. We demonstrate the effectiveness of this baseline system through an extensive human evaluation, providing evidence that participants used the underlying semantic relations to explore articles across various domains. We discuss not only identified challenges with entity recognition and linking but also technical issues regarding speech recognition and generation. These limitations serve as design implications for building more advanced systems and enable meaningful comparisons with our baseline. Our future work will concentrate on techniques to address challenges associated with accurately matching entities from user requests to the knowledge graph in multiple languages. Moreover, we plan to integrate large language models for enhanced response generation while using the graph structure as explicit knowledge grounding. By doing so, it is our aim to generate more natural, contextually appropriate responses and augment the search experience with personalized news summaries.

Acknowledgements

References

Nicholas J Belkin. 1980. Anomalous states of knowledge as a basis for information retrieval. *The Canadian Journal of Information Science*, 5:133–143.

Janez Brank, Gregor Leban, and Marko Grobelnik. 2017. Annotating documents with relevant wikipedia concepts. In *Proc. of Slovenian KDD Conf. on Data Mining and Data Warehouses (SiKDD)*.

John B. Brooke. 1996. Sus: A 'quick and dirty' usability scale. *Usability evaluation in industry*, 189(194):4–7.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Debanjan Chaudhuri, Md Rashad Al Hasan Rony, and Jens Lehmann. 2021. Grounding dialogue systems via knowledge graph aware decoding with pre-trained transformers. In *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*, pages 323–339. Springer.

Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. Towards knowledge-based recommender dialog system. In *Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1803–1813, Hong Kong, China. Association for Computational Linguistics.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais. 1987. The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514.

Sunyoung Kim and Abhishek Choudhury. 2021. Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior*, 124:106914.

Alexander Maedche, Christine Legner, Alexander Benlian, Benedikt Berger, Henner Gimpel, Thomas Hess, Oliver Hinz, Stefan Morana, and Matthias Söllner. 2019. Ai-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering*, 61:535–544.

Carey K. Morewedge and Daniel Kahneman. 2010. Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, 14(10):435–440.

OpenAI. 2022. Chatgpt: Optimizing language models for dialogue. *OpenAI Blog*.

Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Trans. Comput.-Hum. Interact.*, 27(4).

Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, page 117–126, New York, NY, USA. Association for Computing Machinery.

Rajdeep Sarkar, Koustava Goswami, Mihael Arcan, and John P. McCrae. 2020. Suggest me a movie for tonight: Leveraging knowledge graphs for conversational recommendation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4179–4189, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Bahareh Sarrafzadeh, Olga Vechtomova, and Vlado Jokic. 2014. Exploring knowledge graphs for exploratory search. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, page 135–144, New York, NY, USA. Association for Computing Machinery.

Phillip Schneider, Anum Afzal, Juraj Vladika, Daniel Braun, and Florian Matthes. 2023a. Investigating conversational search behavior for domain exploration. In *Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part II*, page 608–616, Berlin, Heidelberg. Springer-Verlag.

Phillip Schneider, Nils Rehtanz, Kristiina Jokinen, and Florian Matthes. 2023b. Voice-based conversational agents and knowledge graphs for improving news search in assisted living. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '23, page 645–651, New York, NY, USA. Association for Computing Machinery.

Phillip Schneider, Tim Schopf, Juraj Vladika, Mikhail Galkin, Elena Simperl, and Florian Matthes. 2022. A decade of knowledge graphs in natural language processing: A survey. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 601–614, Online only. Association for Computational Linguistics.

Jian Wang, Junhao Liu, Wei Bi, Xiaojiang Liu, Kejing He, Ruifeng Xu, and Min Yang. 2020. Improving knowledge-aware dialogue generation via knowledge base question answering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 9169–9176.

Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98.

Graham Wilcock. 2012. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the Workshop on Question Answering for Complex Domains*, pages 57–70, Mumbai, India. The COLING 2012 Organizing Committee.

Houyu Zhang, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. 2020. Grounded conversation generation as guided traverses in commonsense knowledge graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2031–2043, Online. Association for Computational Linguistics.

Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '20, page 1006–1014, New York, NY, USA. Association for Computing Machinery.

A Appendix

The appendix provides further insights into the results of our research, including an overview of the news knowledge graph and the architectural components of the conversational exploratory search system, as depicted in Figures 5 and 6, respectively.
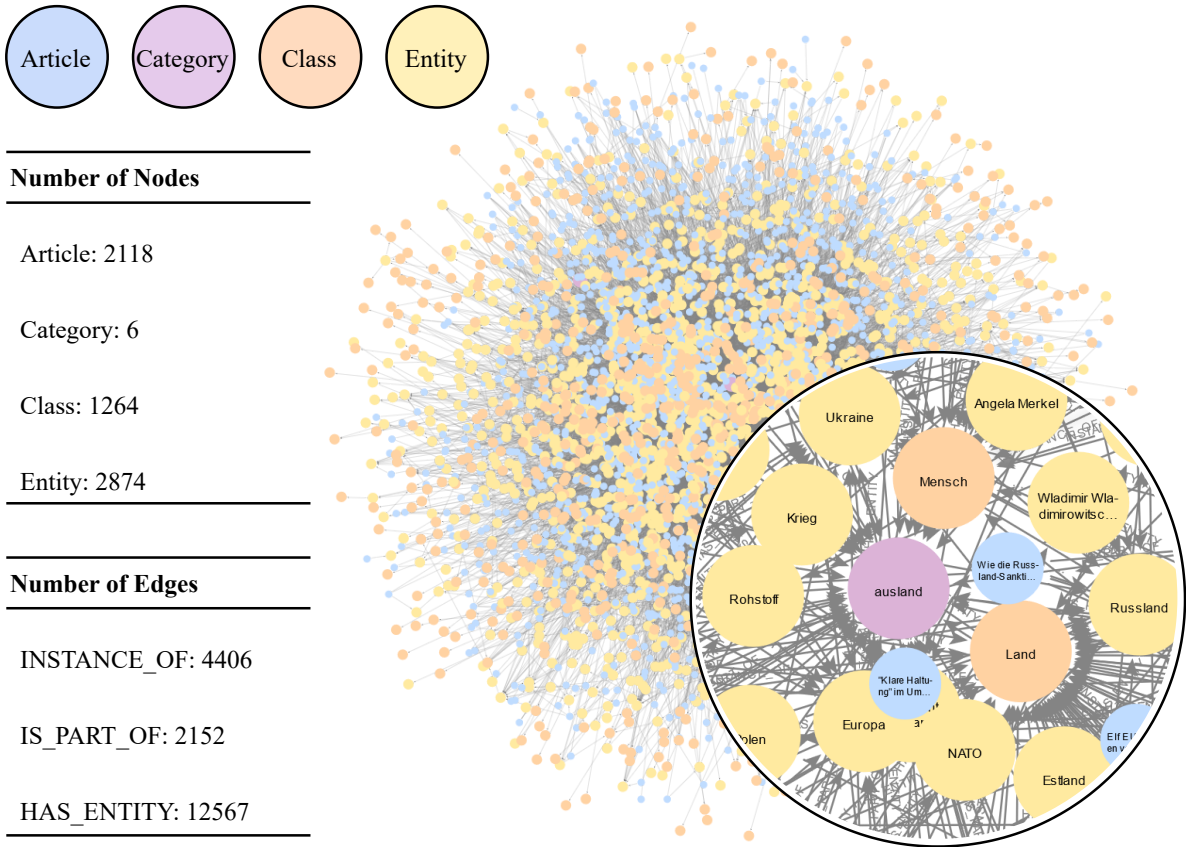

Figure 5: Overview of the different node and edge types in the automatically constructed news knowledge graph.
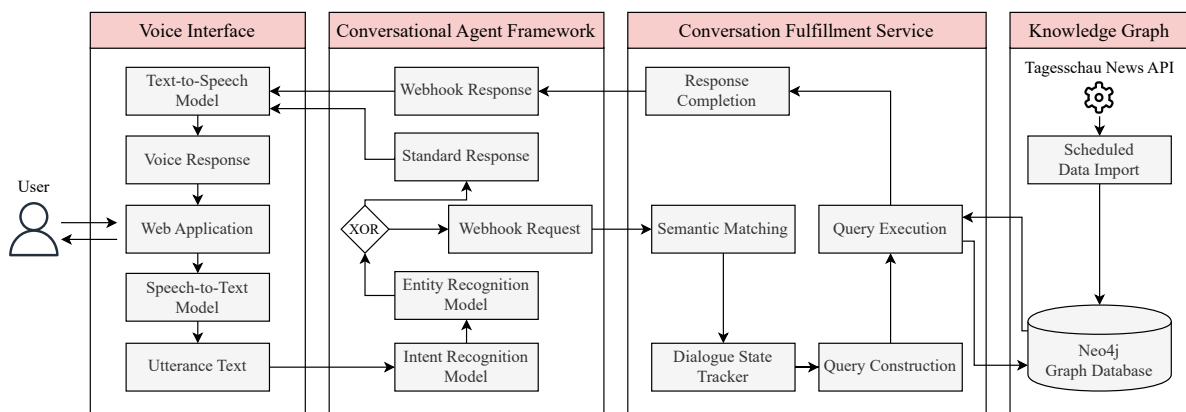

Figure 6: Architectural components of the proposed conversational exploratory search system.