1 Conversational Search

Johanne R. Trippas

Conversational search is rising in popularity. Recent advantages in machine learning and natural language processing increasingly enable the development of conversational systems such as chatbots or speech-enabled personal assistants. The general population is adopting conversational systems to give commands and search, or interact with information. This chapter provides an introduction and overview to searching through conversational interactions. This part of the book includes neighbouring research fields of information retrieval, such as spoken dialogue systems and question answering, illustrating the complexity of these conversational systems while arguing about approaching this problem in a multi-disciplinary manner. This chapter aims to provide a starting point to conversational search while identifying research challenges and opportunities through contrasting aspects of traditional web-search and conversational systems. We illustrate that conversational search is a different mode of interaction and needs renewed attention, expanding beyond the query-response paradigm. We highlight the inherently interactive procedure of searching for information. We also show that conversational search is a different mode of interaction and needs a revised focus to define and understand the interaction components.

1.1 Introduction

Conversational search, the process by which searchers find information through conversations, is a highly popular research topic. Searching for information in which a searcher interactively converses with a system is becoming ubiquitous, especially through the use of mobile devices, smart speakers, and commercial intelligent assistants such as Amazon Alexa, Apple's Siri, or Google Assistant. We can interact with these systems to ask for weather updates, find directions, add calendar items, or seek information on any topic covered on the Internet. Furthermore, in its essence, information seeking, the process by which humans find information, has always been regarded as a highly interactive activity [Oddy 1977]. In fact, from early on, searching has been approached as an interactive user–system activity. However, more recently, we have seen a growth driven by machine learning (ML), natural language processing (NLP), and spoken language understanding to help accomplish a genuinely interactive system going beyond the traditional action-reaction search model [Trippas et al. 2018].

In addition, conversational search as a new interaction paradigm introduces opportunities. For example, we can learn how to advance the traditional browser-based "query box" searching, progress on the accessibility of information, and lower barriers of information access through offering search through every-day conversations.

Given that conversation is the natural mode for information exchange in daily life, a conversational format for search input and output is logical. Furthermore, using conversational interactions could overcome the difficulties inherent of searching through ever-growing corpora solely through queries and minimal feedback interactions to the system. Thus, to conversationally search is not only through providing a query on which the system ranks and provides a search engine results page (SERP) but also includes feedback loops to update the results. Furthermore, the act of conversation will help the user to formulate the information need more clearly. For example, the following exchange is an illustration if we would be interacting with a non-conversational system through a keyword query:

USER: Mandatory to wear mask

SYSTEM: In some countries masks are worn in accordance with local customs or in accordance with advice by national authorities in the context of COVID-19. In these situations, best practices should be followed about how to wear, remove, and dispose of them, and for hand hygiene after removal. It is strongly recommended that you wear a face mask when you have visitors at your home, and that your visitors also wear masks. It is also recommended that when you visit others you also wear a mask. You should always use hand sanitiser or wash your hands before and after changing your face mask...

However, the above results selection or presentation is not interactive, is highly cognitive demanding, and does not actively support the searcher in their information need. Instead, the system retrieves and presents a lengthy text of information without breaking it up and thus forcing the user to decipher all the information. Furthermore, no query refinements or reformulations are offered by the system, leaving the user vulnerable for the need to formulate the perfect query from the start. A more interactive system which includes query refinement support and multi-turn input would therefore be more desirable:

USER: Hey, I have a question, is it mandatory to wear a mask?

SYSTEM: Do you mean in general or in public spaces?

USER: I guess, more in public spaces with all the COVID things.

SYSTEM: According to the ABC News, many countries around the world, as well as some US states, ask people to wear face masks in public places to limit the spread of the COVID-19 virus.

USER: Uhm, and what about when I visit my mum?

SYSTEM: In Australia, the current advice from the Federal Government is that it is recommended that when you visit others you also wear a mask.

USER: Right, so I should wear one as well when I go and visit my brother.

SYSTEM: The quick answer is yes, and you should always use had sanitiser or wash your hands before or after changing your mask.

USER: Cool, thanks for the info.

SYSTEM: Any time.

In the above example, the information is presented in a more human-like conversation, including increased natural language interactions. Intuitively, the conversation is broken down into chunks avoiding information overload. In addition, the system understands the context of the user helping them to refine their query and keeps track of previous interactions without the need for tedious repetitions. Furthermore, the system builds a model of the user and their information need through problem elicitation. All these user–system interactions were feasible by having both parties engaged in a human-like conversation.

Conversational systems have many advantages, some include;

1. alleviating the cognitive burden on the user by breaking down the information,
2. assisting the information need formulation through interactions and conversation,
3. very little to no learning is required since human conversation is a natural way of communicating for humans,
4. highly personalisable through including the context of the user and information units,
5. in the case of a voice-based conversational system, accessible when hands and eyes are occupied, making information more accessible, and
6. having the potential to reduce e-waste because they can easily be embedded in existing systems.

1.1.1 Defining Conversational Search

Defining conversational search is hard. Recently, several researchers have attempted to clarify what conversational search is from a user and system point of view [Azzopardi et al. 2018, Radlinski and Craswell 2017], a cognitive perspective [Trippas 2019], or through user–system interactivity [Trippas et al. 2020a]. These definitions were formed by studying past work on human conversations, theoretical instigation, and empirical studies. Previous work also proposed which properties would be desirable for a conversational IR system and system requirements to support the user naturally and interactively [Anand et al. 2020, Culpepper et al. 2018, Deldjoo et al. 2021, Radlinski and Craswell 2017, Trippas 2019]. Through these works, the action space of conversational systems was exposed, making them useful for evaluating systems which claim to be genuinely conversational. However, no formal definition has been finalised due to the complexity of the system.

For example, Radlinski and Craswell [2017] defines conversational search for a chat as:

> *"A conversational search system is a system for retrieving information that permits a mixed-initiative back and forth between a user and agent, where the agent's actions are chosen in response to a model of current user needs within the current conversation, using both short- and long-term knowledge of the user."* [Radlinski and Craswell 2017, p. 120]

Trippas [2019] defined spoken conversational search by the following system properties as:

> *"A spoken conversational system supports the users' input which can include multiple actions in one utterance and is more semantically complex. Moreover, the conversational system helps users navigate an information space and can overcome standstill-conversations due to communication breakdown by including meta-communication as part of the interactions. Ultimately, the conversational system multi-turn exchanges are mixed-initiative, meaning that systems also can take action or drive the conversation. The system also keeps track of the context of particular questions, ensuring a natural flow to the conversation (i.e., no need to repeat previous statements). Thus the user's information need can be expressed, formalised, or elicited through natural language conversational interactions."* [Trippas 2019, p. 142]

Similarly, Vakulenko [2019] defined conversational search by:

> *"The task of retrieving relevant information using a conversational interface is termed conversational search."* [Vakulenko 2019, p. 15]

A recent Dagstuhl report also tried to define conversational search while bridging common terminology used by different research fields [Anand et al. 2020, Arguello et al. 2020]. The authors state that there are many different possible characterisations of conversational search. Nevertheless, they attempted to structure common terminology through several representations. In particular, they characterise conversational search through the typology of conversational search systems, also referred to as the conversational Ψ (psi), by explaining and extending functionalities of information retrieval (IR), chatbots, and dialogue systems as seen in Figure 1.1.

The typology aims to capture the variety of systems which are needed to create a conversational search system. Highlighting that conversational systems need a variety of researchers to fulfil conversational properties. Arguello et al. [2020] also emphasise that many properties

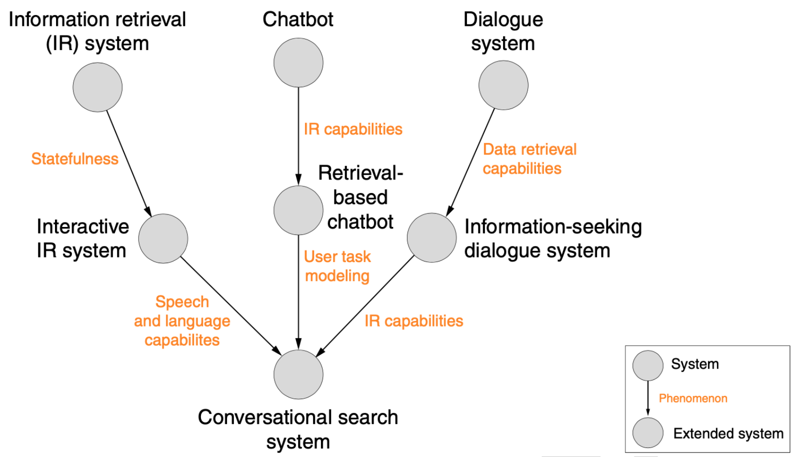

Figure 1.1 The Dagstuhl Typology of Conversational Search [Arguello et al. 2020].

can overlap, thus making it even more challenging to define conversational search. Therefore, they define conversational search as:

> *"A conversational search system is either an interactive information retrieval system with speech and language processing capabilities, a retrieval-based chatbot with user task modelling, or an information-seeking dialogue system with information retrieval capabilities."* [Arguello et al. 2020, p. 52]

1.1.2 Combined Representation of Conversational Search Definitions

The previous section has provided several definitions of conversational search. Many of which are either stated very loosely or providing properties of such systems. Even though the definition is sometimes vague, there are overlaps between the definitions, making them useful to provide a combined view of the predefined system requirements (see Table 1.1).

This representation of conversational system properties highlights the complexity, overlapping, and inter-connectivity of components. We argue that in order to create systems satisfying all these needs, these highly interactive systems have to be created with care and inclusive of the users [Spina et al. 2021]. This means that creating systems with a user-centred design approach (i.e., an iterative design process that focuses on having users at the centre of product design and development) is crucial. Furthermore, issues such as trust and transparency of the systems need to be addressed. For example, when a user has a question such as "Should we abolish the death penalty?", the system should provide sufficient explanations around the pros and cons of the death penalty instead of focusing on a *yes* or *no* answer [Kiesel et al. 2021].

Not only do we need to develop these conversational systems with care towards the user, other issues to work collectively towards the same goal need to be met. For example, a

Table 1.1 Desired system requirements.

		System
1	*Analogy*	Interact through human intelligible dialogue-like conversations, beyond command and control [Trippas 2019]
2	*Language*	Understand and create natural language output and conversational interactions [Arguello et al. 2020, Radlinski and Craswell 2017, Trippas 2019, Vakulenko 2019]
3	*System participation and user revealment*	Interact proactively, through mixed-initiative (implies listening) including clarifying questions and keep track of memory [Radlinski and Craswell 2017, Trippas 2019]
4	*Information request length*	Comprehend longer and more natural information needs [Trippas 2019]
5	*Results presentation mechanism or set retrieval*	Adequately adapt to users' need and context (ranked list is inadequate) including the ability to reason about set retrieval [Radlinski and Craswell 2017, Trippas 2019]
6	*Turn-taking*	Construct and process multi-turn conversations [Radlinski and Craswell 2017, Trippas 2019]
7	*History and memory*	Establish user profile or model over (multiple) sessions [Radlinski and Craswell 2017, Trippas 2019]
8	*Multi-moves*	Enable and support utterances with multiple moves from user and system [Trippas 2019]
9	*Errors*	Anticipate errors and assist with intelligent problem solving (through meta-communication) [Trippas 2019]
10	*Turn-time*	Manage less predictable turn-times [Trippas 2019]
11	*Semantics*	Regulate complex, more discourse sentence constructs [Trippas 2019]
12	*Navigation*	Guide users through multi-dimensional information space [Trippas 2019]
13	*User feedback*	Proactively eliciting feedback on past results [Radlinski and Craswell 2017]

multitude of research fields is needed to create systems that will be genuinely conversational. Different research fields such as NLP, spoken dialogue systems (SDS), artificial intelligence (AI), interactive information retrieval (IIR), or human-computer interaction (HCI) are needed.

In this section, we illustrated the complexity of conversational search. Furthermore, one thing is clear from the definitions as mentioned earlier, conversational search is more than "audiofying" the web.

1.1.3 Chapter Overview

In this chapter, we present the background to relevant previous work in conversational search. In Section 1.2, we first provide information on the history of conversational systems, including speech user interfaces and SDS. In Section 1.3, we briefly review IIR and question answering

(QA) concerning conversational search. We then present past work on modelling information-seeking interactions through dialogue. The section concludes with a summary of theoretical frameworks and properties for conversational search. Next, we outline two fundamental conversational components for interacting with search systems, namely actions related to search and non-search (Sections 1.4–1.5). In Section 1.6, we cover implementation and evaluation topics for conversational systems. Finally, we finish with a conclusion (Section 1.7), some challenges, and summary in Section 1.8.

1.2 Searching Through Conversations

> *"With a conversational interface, people can speak to their smartphones and other smart devices in a natural way to obtain information, access Web services, issue commands, and engage in general chat."* McTear et al. [2016, p. 11]

A conversational system is a broad term for a system with which a user can interact through conversations. Users can interact through "talk" to retrieve information, access services, or issue commands and in which the system responds.

Researchers in AI, speech technology, and HCI have long anticipated and worked towards conversational systems. Many science fiction movies have envisioned how conversational systems would interact and come to fruition, for example, in *Star Wars*, *2001: A Space Odyssey*, or *Her*. Apple continued with the conceptualisation of such system with the *Knowledge Navigator* in 1987. Apple's system could assist users in planning, searching, or communication by including advanced modules such as text-to-speech (TTS), NLP, and natural language generation (NLG). Even though the above systems were imaginary, arguably one of the first working conversational chatbot[1] system was ELIZA [Weizenbaum 1966]. ELIZA was created by Weizenbaum in the mid-1960s and replied to user input imitating interactions between a patient and therapist.

The vision of conversational systems has never ceased and has only been growing stronger. For example, more than twenty years ago in 2001, Berners-Lee et al. [2001] anticipated that the hypertext link infrastructure could be utilised to help with the development of conversational systems. Nevertheless, until the mass introduction and adoption of Siri and Google Now in 2011 and 2012, respectively, the vision of conversational systems received extensive consideration.

All recent developments and advancements in AI, NLP, spoken dialogue management, and speech recognition have driven the renewed and growing interest in conversational systems [Gao et al. 2017, Trippas et al. 2020b, Xiong et al. 2018]. Additionally, smart devices such as smartphones or smartwatches are progressively powerful and being connected to the internet. This means that we are now increasingly capable of accessing robust computing services that support our wants and needs to complex systems. These technological as-

[1] Note, the previous systems included spoken conversational systems

pects, which can support the adoption of conversational search, have been developed rapidly. Nonetheless, truly conversational systems to help users search over the ever-expanding (user-generated) unstructured data, need further advances to become genuinely conversational.

1.2.1 Speech User Interfaces

Human speech through natural language is the most universally used form of communication, as well as the most intricate one. Even though speech is considered as a natural way to interact, speaking to a machine or computer is still mostly seen as "peculiar" [Klemmer et al. 2000, Turunen et al. 2012]. However, with the recent advancements of spoken interactive systems such as Apple's Siri or Google Assistant, both as built-in devices in mobile phones or headphones, talking to a computer is becoming more broadly accepted. The use of speech systems can be helpful in particular circumstances, such as when one's eyes or hands are busy [Cohen and Oviatt 1995], and allows for information to be accessed without needing a keyboard or typing [McTear et al. 2016, Yankelovich et al. 1995]. Furthermore, typing takes longer for many people than speaking [Čech and Condon 1998].

Additionally, these speech systems can be used by people who may otherwise be unable to access information via text, such as people with a visual impairment, people with dyslexia, or low literacy skills [Sahib et al. 2012, Trippas et al. 2021, Turunen et al. 2012]. Even though many people think of current systems such as Apple's Siri, Google Assistant, or Amazon Alexa as speech user interface, previous work on making the web accessible was done through a regular phone. Indeed, many people around the world can access voice-driven applications hosted in an everyday telecommunications network. People can call a specific phone number with their landline or even payphone to access the spoken web without needing the latest technology or internet access, making information even more accessible.

Useful speech user interfaces should require no to little learning of the interface and simulate a human–human conversation [Gardner-Bonneau and Blanchard 2007]. Therefore, the interface should include and mimic typical voice communication efforts such as turn-taking, barge-in (i.e., interrupting), pauses to receive responses, and corrections or repairs if the conversation fails. Furthermore, other techniques such as including discourse markers and sounds as feedback in speech user interfaces may optimise the utility [Arons 1993, Chuklin et al. 2019, Yankelovich et al. 1995].

However, there are also disadvantages and drawbacks to using speech user interfaces. Speech user interfaces are not only hard to build and maintain; they can be more expensive and time-consuming to test and create. Furthermore, it is hard to present graphs and images over a speech channel to users [Trippas et al. 2018]. In a search environment, this will mean that visual representation, such as highlighting query terms, is restricted [Chuklin et al. 2019, Hearst 2009]. Other difficulties speech user interfaces encounter is that the user may talk before the system is ready, not know what to say, read meanings into pauses while the system is still processing, and may find it harder to consume speech input than producing output [Trippas 2019]. Even though there are clear challenges to creating a good speech user interface, overcoming them may have a greater impact on society for equal information access.

1.2.2 Spoken Dialogue Systems

A SDS (often referred to as voice user interface in industry [McTear et al. 2016]) is an example of a speech user interface. In other words, the SDS acts as an interface connecting the user and computer by exchanging knowledge on a turn-by-turn basis [Gibbon et al. 1997]. These SDSs are a platform in which people can interact through spoken natural language with computer systems such as an online flight database. The tasks supported by a SDS are very narrow and the system aims to elicit structured information such as flight destination, departure, airline preference, and return flight. Eliciting this information can be done through open-ended prompts such as “How may I help you?” or through directed dialogue or system-initiative by asking specific information about the user’s flight. A third alternative to retrieve information from the user is often referred to as mixed-initiative, where both user and system can steer the conversation.

A SDS is not just an automatic speech recognition system bolted on to a TTS system. Instead, it involves many technical components in order to “talk” with a human. Indeed, a SDS is a complex system which includes speech and language components, such as speech recognition, natural language understanding, dialogue management, NLG, and TTS. All these different components have to work together in real-time with each other and the users, making SDSs complex systems with different in and output formats (i.e., audio, text, semantics) as seen in Figure 1.2.

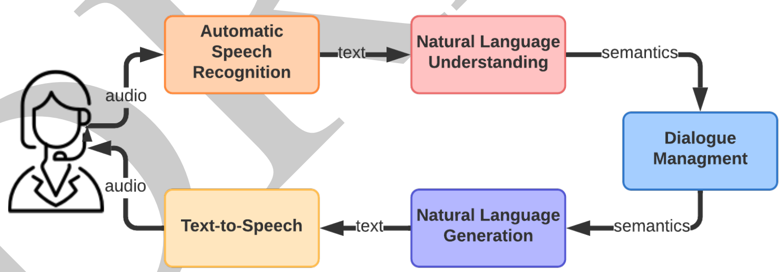

Figure 1.2 A typical SDS pipeline and its components.

In recent years, interest in SDSs has grown, as speech technology [Xiong et al. 2018] and ML for spoken systems [Yang et al. 2018a] have developed. A range of SDSs have advanced, from QA to semi-conversational systems [McTear et al. 2016]. The early research in SDS had been devoted to task-oriented tasks which have defined search boundaries, such as travel planning or route planning, and can be developed with slot filling approaches [Walker et al. 2001]. However, more recently, research efforts have been focused on the task of search for which no clear path or “slots” exist. Thus, for the task of search, rigid plan-based dialogue approaches many not be appropriate [Higashinaka et al. 2014]. Efforts of understanding how to deal with a variety of user utterances and how replies can be structured and generated appropriately are needed [Sugiyama et al. 2013, Trippas et al. 2020a].

1.3 Information Seeking Models, Theories, and Properties

This section aims to introduce several concepts which are used to build upon in conversational search. We first provide a brief introduction to relevant concepts of IIR and QA (Sections 1.3.1–1.3.2). We then introduce additional research around modelling information-seeking interactions through the use of dialogues in Section 1.3.3. Finally, we further specify theoretical frameworks introduced in Section 1.1 and how their properties relate to conversational search (Section 1.3.4).

1.3.1 Interactive Information Retrieval

Traditional IR focuses on developing, evaluating, or indexing information. This form of retrieving information is mainly viewed from a computational perspective and mostly does not involve real people. In contrast, IIR concentrates on the *interactivity* between a user and a system [Borlund 2013]. Focusing on how people interact and search for information while satisfying their information need is the core aim of IIR. Thus, "interactive" suggests the human engagement in information seeking in contrast to classic IR which abstract humans out of the equation.

There are many different ways a user can interact with a search system. The most-used and well-accessed way to information are through typing a query in a browser-based search box on a desktop. However, with the increasing usage of mobile devices, other modes of information access received much attention [Ong et al. 2018]. More recently, with the renewed interest for other modalities of accessing information such as voice-based, searching through conversations has become a popular research topic. Furthermore, the word "conversation" already embeds the interactivity since it is something one cannot do alone and is inherently interactive.

Figure 1.3 is a possible abstraction of an entire search session of a traditional web-based search interaction versus a hypothetical session of conversational search. The web-based search interaction model was proposed by Maxwell and Azzopardi [2016]. Their model considers six actions: *(i)* the application of the query (re)formulation strategies; *(ii)* snippet scanning and assessment; *(iii)* snippet clicking; *(iv)* document reading; *(v)* document assessment; and *(vi)* session stopping. We now transfer these interactions to a different IIR context, the conversational setting, in which a user may not be consuming documents but instead consume "information units" [Allan et al. 2012]. These units are a way to combine information which may be scattered over multiple documents and is in contrast to providing the user with a relevant web-page. In addition, we propose that the user will exhaust an "information space" (i.e., a set of knowledge or information units and their relationship) if they want to learn more or satisfy their information need. This behaviour is in contrast to reading documents. The conversational model includes similar actions but tailored to a conversational setting: *(i)* the application of the information need (re)formulation strategies; *(ii)* information unit valuation; *(iii)* information unit continuation; *(iv)* information unit assessment; *(v)* information space consummation and fulfilment; and *(vi)* session stopping. As illustrated in the example search processes, the concepts of "information units" and "information space" are an alternative perspective on the classical "ranked documents". In addition, this new interaction paradigm also influences the expected chain of interactions. In summary, this new form of search is intrinsically concerned with interactivity and its user, making conversational search an ultimate IIR topic.

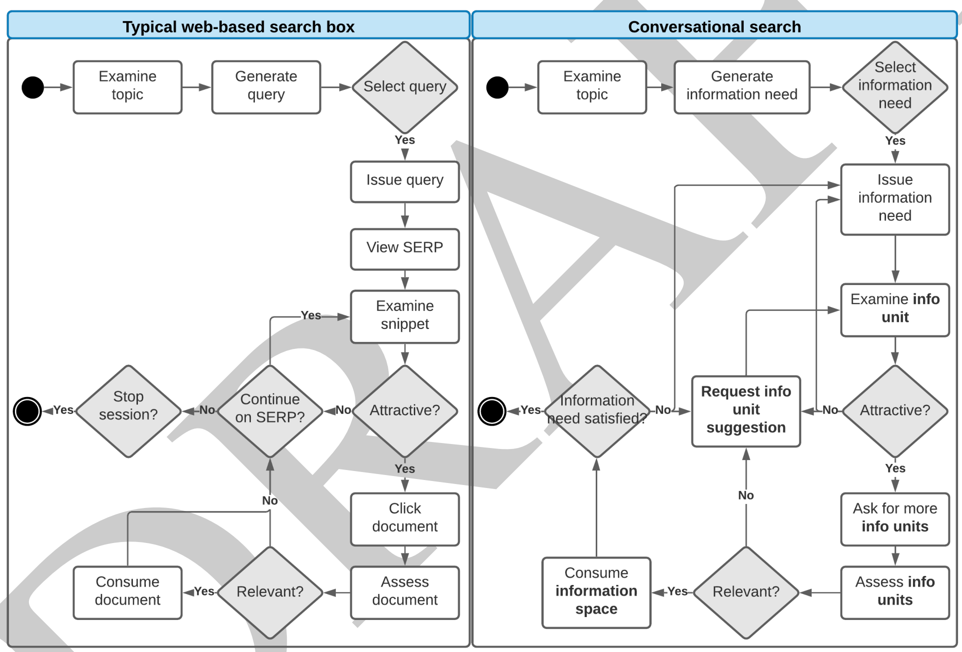

Figure 1.3 Example search process of a typical browser-based search box interactions versus hypothetical conversational interactions.

1.3.2 Question Answering

Question answering (QA) in a text setting has been a longstanding research topic within the IR and NLP community. QA combines the knowledge of these two research fields to retrieve small text snippets containing the exact answer to a query instead of the document lists traditionally returned by text retrieval systems [Voorhees 1999]. The short answers are often retrieved and presented to the user as short text passages, phrases, sentences, or knowledge graph entities [Lu et al. 2019]. These different answering mechanisms allow systems to be flexible while answer both factoid and non-factoid queries.

With the developments around conversational systems, QA work received increased attention in the context of conversational search in IR [Christmann et al. 2019, Kaiser et al. 2020, Qu et al. 2019b]. Indeed, conversational QA can be seen as a subsection of conversational search. Recently, Qu et al. [2019b] defined conversational QA as a "simplified conversational search". That is, conversational QA is not proactively asking clarifications like it is in conversational search. Furthermore, conversational QA can be seen as closely related to machine comprehension [Kenter and de Rijke 2017, Yang et al. 2018b]. The main difference being that conversational QA organises the questions into conversations in comparison to single-turn QA in machine comprehension [Qu et al. 2019b]. This means that leveraging the interaction history and use this history model is crucial to create robust and effective conversational QA systems. For example, the history can be used to help map the state and changes of the information need to inform current or future responses [Qu et al. 2018]. Recent work from Kaiser et al. [2020] also mention the importance of dialogue context as a way to improve conversational QA. That is, the implicit context of previous utterances can be referred to by the user in later interactions. Again, many aspects of QA could be adaptable or expandable to enhance conversational search.

1.3.3 Modelling Information Seeking Through Dialogue

Even though limited research has been devoted to the new search paradigm of conversational search, some early work can provide insights into the aspects of query formulations, results presentation, and filtering. For example, Oddy [1977] introduced man-machine IR through dialogues in the 1970s. He developed a virtual character-based reference retrieval program to help users relevant documents, called THOMAS. THOMAS was designed to act as a human enquirer (or intermediary), allowing the user to select documents without explicitly formulate their information need. Instead, THOMAS was designed to iteratively decrease the possible document pool or search scope by eliciting information through questions without having the user explicitly state queries. THOMAS only presented the possible relevant documents at the end of the interactions and thus acting as a filtering process. Some initial evaluations with the prototype suggested that retrieving relevant documents without the need of formulating queries was feasible.

In the 1980s, nearly ten years after THOMAS, Croft and Thompson [1987] designed the Intelligent Intermediary for Information Retrieval (I^3R) where the system is modelled on an expert intermediary. Again, the idea of only allowing users to search through queries was extended. That is, the I^3R system was able to utilise domain knowledge acquisition, explanation, browsing, retrieval, and evaluation to support the user. Furthermore, the system was capable of engaging in some dialogue with the user to confirm or request further information.

While the above studies were mainly interested in document retrieval, other studies have focused on the discourse aspects of communicating with an information system to elicit expert knowledge. For example, more than three decades ago, Belkin et al. [1987] investigated the interactions between information seekers (clients) and librarians (experts) in which the librarian acted as an intermediary. A coding schema was proposed to annotate these conversations the help design better expert systems. This particular coding schema highlighted that contextual information from dialogues could be deduced, including the description, states, modes of problems at hand, user models, search strategies, and search interactions. In later work, Belkin et al. [1995] considered conversational information-seeking through identifying "strategies" and "cases". Based on these strategies, they proposed scripts which could be used by a system to interact with the user depending on the type of retrieval task. Thus, depending on the kind of information needs, different case-based reasoning selections could be used to decide the next steps and propose the user with choices.

Another body of work which also focused on discourse aspects of conversations included the usage of the function of an utterance. Sitter and Stein [1992] created the COnversational Roles (COR) model. These researchers proposed ways to incorporate searching for information through dialogue but this time with the use of Dialogue Acts (DA) [Sitter and Stein 1992, Stein and Maier 1995]. DA are functions outlined as a schema which represents the generic meaning of an utterance. The COR model provided a general structure for information seeking dialogue. The authors envisaged that this plan-based model could be used to guide users through stages of information seeking. The model is illustrated as a transition network and is primarily focused on tracking conversational commitments, see Figure 1.4. In this model, the actors are noted as A (information seeker) and B (information provider). The circles and squares symbolise the states as part of the dialogue. Arrows represent the progress between the states. For example, in step ① the seeker makes the first move with the possible outcomes outlined in example ②. This atomic move is annotated with DA as *request(A,B)*.

Other features of the COR model incorporate the adaptability in *mixed-initiative* and *meta-communication*. These two features allow users to determine which subsequent actions will occur or ask questions at any given time. Support for mixed-initiative dialogues is desirable in conversational interactions because this allows for more spontaneous and natural interaction. Nevertheless, mixed-initiative dialogues are more difficult and complicated for the system to handle [McTear et al. 2016]. The model also allows for meta-communication by permitting the conversation to go through one of the loops at any point in time. However, only one move in an utterance is plausible according to the COR model, making the model unaccommodating for the versatility of voice input and output. Furthermore, recent work by Trippas et al. [2017] and Vakulenko et al. [2019], suggested that the COR model fails to reflect the structure of an information-seeking dialogue appropriately. In response, Vakulenko et al. [2019] created a new DA-based model called the QRFA model (which stands for Query, Request, Feedback, and Answer loops). These communication loops are modelled of User–Agent feedback aiming

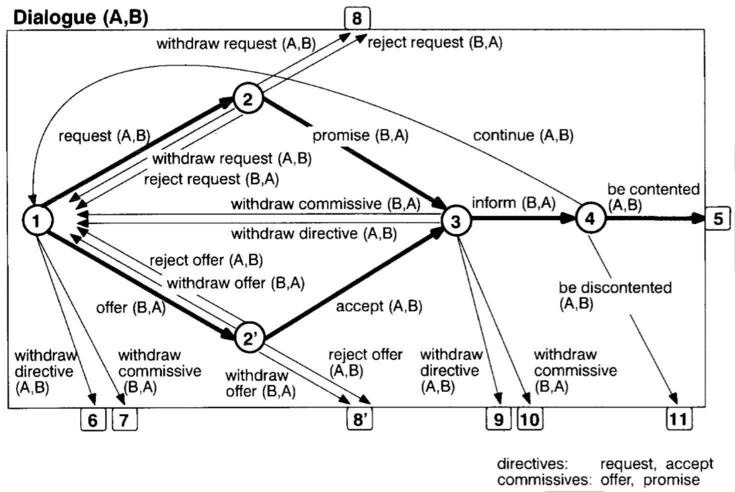

Figure 1.4 COR model by Sitter and Stein [1992].

to explain conversational flows occurring in real interactions. However, the distillation of these DA-based models can be oversimplified and fail to reveal further interactions unique to conversational search [Trippas et al. 2018]. Furthermore, multi-faceted information-seeking commonly include more complex interactions such as several moves in one utterance or users unexpectedly employ unanticipated interaction paths [Belkin et al. 1995, Trippas et al. 2018].

Indeed, the COR [Sitter and Stein 1992], QRFA [Vakulenko et al. 2019] models and scripts [Belkin et al. 1995] enable the forecasting of which kind of interaction will be required following from a previous one. These predictions are a form of discourse cycles and can become “predictable” defaults adaptable to maximise performance by necessitating minimal encoding of the system. Hence, if we could predict, interpret, and explain the user input, without losing much-needed features about the conversation, we may be able to provide suitable responses for the user produced by the system. This section introduced several setting in which conversational search can be used. That is, as a document referencing tool or general search. Other contexts for conversational search include enterprises [Cavedon et al. 2020], cooking [Frummet et al. 2019], or aerospace [Arnold et al. 2020, Gosper et al. 2021]

1.3.4 Theoretical Frameworks and Properties for Conversational Search

In the last few years, several attempts have been made creating theoretical frameworks for conversational search. These proposed frameworks and models are discussed in this section and help define permissible steps or actions which can be taken during an information-seeking conversation by either a user or system. As such, the models reveal which kind of statements can be expected by the user or system. Furthermore, the frameworks have tried to expand on properties conversational systems should have.

For example, in 2017, Radlinski and Craswell [2017] suggested a theory and model for information interaction in a chat setting. They studied previous work on human conversations on which they built a set of properties to understand whether a system can be classified as truly conversational. They provide the needed five properties as RRIMS:

1. User **R**evealment: The system helps the user express (potentially discover) their true information need, and possibly also long-term preferences.
2. System **R**evealment: The system reveals to the user its capabilities and corpus, building the user's expectations of what it can and cannot do.
3. Mixed **I**nitiative: The system and user both can take initiative as appropriate.
4. **M**emory: The user can reference past statements, which implicitly also remain true unless contradicted.
5. **S**et Retrieval: The system can reason about the utility of sets of complementary items.

Radlinski and Craswell [2017] illustrate that different information-seeking tasks commonly use the above properties and thus should be enabled in future conversational systems. In particular, they argue that those five main properties are necessary for many day-to-day information-seeking tasks. Even though the implementation of the proposed model is future work, the authors suggest that no previous systems yet satisfy all their proposed properties of a conversational search system.

A second relevant conceptual framework, based on Radlinski and Craswell [2017]'s theoretical work and Trippas et al. [2017]'s preliminary analysis, was proposed by Azzopardi et al. [2018] in 2018. Their framework aimed to provide a starting point and possible overview of the actions and intents possible between the user and system, including key decision points within a conversation. According to their framework, a conversational system (or agent) should support the user in finding, exploring, and understanding which information units are available (i.e., providing an overview of the information space [Trippas 2015]) to help satisfy the user's information need. This initial idea was made on the assumption that a system wants to help a user, minimise conversational effort, and maximise the range of relevant options provided to the user. They provide the interactions space from both user and agent overview. While validating the framework empirically was future work, the authors also emphasise that other actions, intents, and decisions can be included.

More recently, Trippas et al. [2020a] proposed the the spoken conversational search annotation schema, *SCoSAS*. This model is different from the models above in this section due to the empirical evidence on which it is built. Indeed, the SCoSAS was carefully developed over several years including; the built of the first fully labelled dataset for conversational search, *SCSdata*; the definition of the first multi-level annotation schema for conversational search, SCoSAS; the creation of a new conversational search model particularly for a speech-only communication channel; and the validation of the proposed SCoSAS model. In particular, Trippas et al. [2020a] aimed to better understand how search results could be presented over an audio-only channel while not overwhelming the user with information, nor leaving them uncertain as to whether they covered the information space. By proposing the SCoSAS, they enabled the investigation of interactivity in spoken conversational search. With this model, they also confirmed that interactivity and proactivity need to be incorporated to overcome the complexity posed by the audio-only channel. The SCoSAS model highlighted the complexity of conversational search and accentuated the need for non-search interactions (i.e., discourse functions) to be integrated into future conversational systems. Furthermore, by outlining the possible interactions, they have opened up the facets or units researchers should investigate.

As seen in Figure 1.5, the SCoSAS model by Trippas et al. [2020a] proposed that utterances could be categorised as either *Task Level* (i.e., conversations on the topical search task), or *Discourse Level* represented conversations about the mechanism of the dialogue (i.e., the function but not the task). Each of the levels (Task and Discourse) consist of eight sub-themes explaining the function or action of that utterance for that level. The Task Level consists of the following four sub-themes: *Information Request, Results Presentation, Search Assistance*, and *Search Progression*; Discourse Level consists of *Discourse Management, Grounding, Navigation*, and *Visibility of System Status* functions. We will use the Discourse Level functions to discuss and frame non-search actions in Section 1.5.

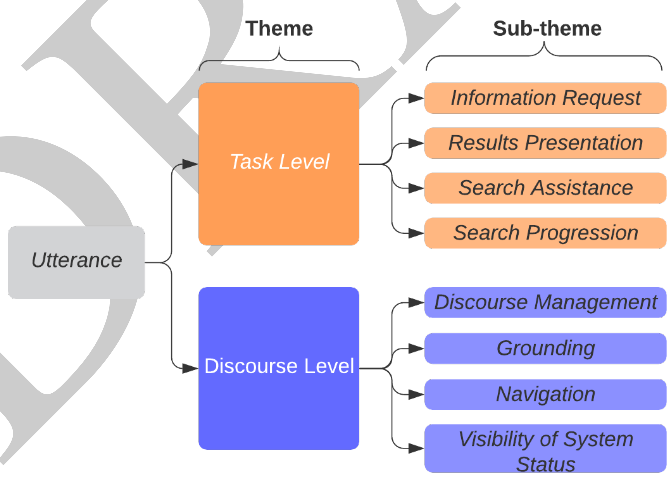

Figure 1.5 SCoSAS model by Trippas et al. [2020a].

In summary, Section 1.3 provided an overview of existing models of search in which dialogue is central and specifically tailored conversational search models. In the next section, we will use these models to explain different actions related to both search and non-search interactions.

1.4 Fundamental Search Actions In Conversational Search

Several actions in conversational search are independent of the interaction mode, be it typed through online chat, spoken through voice, or a combination of both. We present the possible interactions and functional goals for conversational search through the lens of information search behaviours [Hearst 2009, Marchionini 1997]. The *primary* (or *atomic*) actions which take place between a system and a user are *(1)* expression of the information need or query formulation (Section 1.4.1), *(2)* examination of the results or selection recommendations through results presentation and answer organisation (Section 1.4.2), and *(3)* reformulation of the problem (Section 1.4.3). Section 1.4 aims to provide an overview of research related to those fundamental searching actions in conventional browser-based search box setting while contrasting them with conversational search.

1.4.1 Query Formulation

How people express their information need or gap in their knowledge has always been a central part of IIR. These expressions or queries are often seen as the primary way for users to interact with search engines. Traditionally, search queries submitted to a browser-based search box are search intent statements which are explicit and short. These queries are then used by the system to establish which information is relevant to that query and show these relevant results to the user. However, query behaviour changes when query boxes are altered [Agapie et al. 2013]. For example, it has shown that query length increases if a query box has a halo or glow effect which changes as a query becomes longer [Hiemstra et al. 2017].

Query behaviour also changes when the interaction mode changes modality (text versus spoken). Several researchers have studied search queries in which they compared the features between spoken and written queries in both log analyses and lab-based experiments. For example, Crestani and Du [2006] and Guy [2018] showed that spoken queries are often longer, have a more varied language, and are more verbose. More recent work also suggests that the query word order and therefore the manipulation of information intent differs between spoken and written queries [Smirnova 2020]. Furthermore, Guy [2018] also described that voice queries have unique characteristics. For example, voice queries are closer to natural language (they include wh-word queries –what, why, who, and where), the topics are different, and user behaviour (time of use and clicks) differs. A recent observational lab study by Trippas et al. [2018] suggest that queries in a voice-only environment vary more than a traditional browser-based search box setting. They observed a varied range as to how information requests were formed, from more traditional short queries to entirely natural language expressions which were detailed and delicately formulated. These voice queries could also be constructed with

Table 1.2 Examples of different spoken query styles on the task of finding an online tool to help create and format LaTeX tables.

Characteristics	**Query example**
Query-like	*"make latex tables"*
Natural language type query	*"can you please give me a online website to generate and format latex tables"*
One utterance with multiple moves	*"is there a way to automatically generate latex tables or does overleaf have one and can it drag and drop an example"*

several needs in one expression, which is in contrast to a traditional search box query in which only one need is expressed per query.

Nevertheless, there are some conflicting views on the length of text versus voice queries. Even though several different studies reported that voice queries are longer than the average typed text query [Crestani and Du 2006, Guy 2016, Yi and Maghoul 2011] (often referred to as verbose queries [Gupta and Bendersky 2015]), these results are in contrast to research by Schalkwyk et al. [2010]. Instead, Schalkwyk et al. [2010] reported that voice queries were shorter than typed text queries. Either way, all the above research strengthens the argument that there are behavioural variations between different modalities and interfaces. This indicates the importance of understanding how users express their information need.

More importantly and perhaps the more extensive difference between a traditional search box approach and a conversational search is that a conversational system aims to support multi-turn information need revealment or development. Thus going beyond the typical action-reaction paradigm of search engines. In order to support these multi-turn actions, the system requires memory for both in-session and over multiple turns [Trippas et al. 2018]. For example, a user could express their information need as this need unfolds in real-time in collaboration with the system. Thus, exploring and exhibiting the knowledge gap and allowing for a multi-faceted information need [Radlinski and Craswell 2017, Trippas et al. 2018]. This multi-turn information need support also indicates that the system actively anticipates information need changes, and participates in the refinement or elicitation of the problem statement. In addition, such proactive systems aide with the offloading of the user's cognitive resources to formulate *precise* or *oneshot* queries (see Figure 1.6). Indeed, the conversational interactions support tight collaboration between the user and system, enabling for even non-formalised or ill-formed queries to be expressed [Taylor 1962, Trippas 2019]. Searching for information is an inherently interactive procedure. Hence that the support and usage of the interactivity as an advantage with the system is ever increasingly important for effective system design and user experience [Belkin et al. 1995].

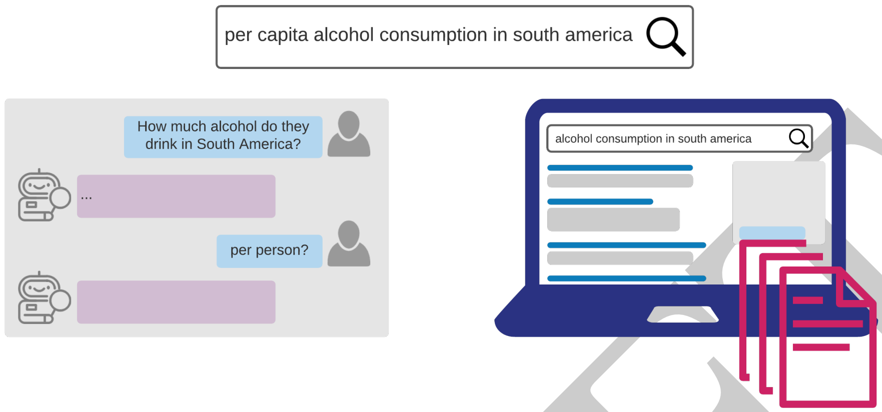

Figure 1.6 Example of search through conversational search in which information needs are refined iteratively versus action-reaction paradigm of a traditional SERP with a oneshot query. (Information need: "*Per capita alcohol consumption in South America*")

As described earlier, the query formulation stage needs to support the user in many different forms of expressing their need. In the formulation stage, the user should also be allowed to revise, refine, or expand on their initial query [Azzopardi et al. 2018] as well as querying inside the given results, spelling out their query, or embellish their query [Trippas et al. 2018, 2020a]. The system can assist in this stage by enquiring for further information about the information need, offering a query refinement option, or rephrase [Trippas et al. 2020a]. While this elicitation process may help the system to understand what the user is searching for, users can benefit from this process too. In particular, this proactive elicitation process can help users to develop their mental model of their search need and information space. Nevertheless, it will be challenging to effectively elicit the user's need without having them to repeat their information need unnecessarily.

1.4.2 Results Presentation and Answer Organisation

While many researchers have investigated the presentation of search results in a traditional search box setting, only recently, studies have investigated the presentation of results for conversational search [Chuklin et al. 2018, Jiang and Ahuja 2020, Ren et al. 2020, Trippas et al. 2015b, Vtyurina 2019]. Furthermore, despite the need for presenting search results in different ways for mobile devices or smartwatches, the traditional SERPs are still dominating the search results presentation style. These SERPs are a representation or surrogate of the underlying document which the user can assess for the relevance to their query. This means that SERPs are still displayed as "ten blue links" which is far away from genuine conversational interactions. Thus, no matter the underlying document, search results are always presented similarly to the SERP list. Nevertheless, with the movement of conversational systems, the results can be presented dynamically and in a more adaptable way. This is a research area for incremental improvement in how conversational search will be used.

Despite the well-known investigation of the summary length in a traditional "ten blue links" setting [Cutrell and Guan 2007, Kaisser et al. 2008, Maxwell et al. 2017], hardly any research has been conducted in understanding results presented in a conversational setting. Early work on the presentation of search results over an audio-only channel was investigated by Trippas et al. [2015a]. They studied user preferences for summary length with a novel crowdsourcing setup and investigated the impact of summary length by comparing preferences between audio-only and text results. While they observed that users preferred longer, more informative summaries for text presentation, this was not the case for audio summaries. In contrast, the results indicated that different presentation styles were preferred depending on the query style. More specifically, shorter audio summaries were favoured for factoid or single-facet queries while for ambiguous or multi-facet queries, user preferences were not as clear. More recently, Vtyurina et al. [2020] further investigated results summaries by comparing audio versus text presentation and suggested that there is a difference in user-preference depending on the presentation medium of the search results. Their work suggests that this user-preference difference may be due to the perceived workload of text versus audio, with listening to search results being more cognitively demanding. Furthermore, Vtyurina et al. [2020] also suggests that even by just reading out the search results, users can still identify relevant results in the audio presentation. Nevertheless, they highly recommend improving the audio results to overcome the cognitive challenges of an audio-only communication channel.

For example, Chuklin et al. [2018, 2019] investigated results presentation by "highlighting" or "bolding" key answer sections through the modification of the audio itself. That is, they manipulated the search results by modifying prosody features such as pauses, speech rate, and pitch. They showed that by emphasising the answer through the audio modifications increases the informativeness of the summary. Figure 1.7 is an illustration of modifying the identified answer in the audio file in which the answer is manipulated in different ways. This research adds to the view that as well as what a conversational system displays, it is equally essential to consider how the system displays information [Thomas et al. 2018, 2020, Trippas 2019].

Besides from changing the length of a summary or highlighting key answers, different mechanisms or strategies have been proposed for answer organisation. For example, in an observational study by Trippas et al. [2018], the authors indicate that the traditional idea of a SERP and linked documents to that SERP is changing. Instead, they suggest that those clear pre-existing boundaries of documents and SERPs are fading. Their results suggests that not all users needed to know whether they were read out a summary or directly from the original document. Furthermore, this fading of document boundaries may be beneficial and multi-document summarisations could be tested as an alternative to SERPs. More proactive ways of summarising through different information unit manipulation strategies (i.e., personalisation,

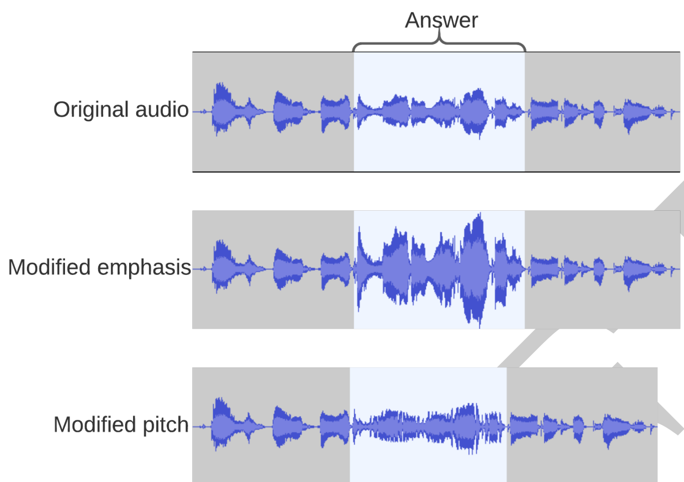

Figure 1.7 Visualisation of "highlighting" answers in audio results.

interpretation of graphic information, or comparing results against each other) are also identified as potential mechanisms [Trippas et al. 2020a]. Vtyurina et al. [2020] also suggest to avoid abbreviations, truncated sentences, and repetitions of words in result presentation over audio, mitigating overloading the user cognitively. Furthermore, other structural information, such as the duration of the audio, maybe beneficial meta-information for users [Vtyurina et al. 2020].

Similarly to Trippas et al. [2018, 2020a]'s suggestions, Azzopardi et al. [2018]'s model includes the opportunity of using summarisation, comparing, and manipulating subsets as a way to present results. Thus, different ways for the system to reveal the possibilities of presenting the underlying corpus [Radlinski and Craswell 2017]. Furthermore, a recent study with goal-oriented flight booking tasks has suggested that different revealment strategies may impact user performance re-iterating to the need for further research in results presentation [Dubiel et al. 2020].

Conversational QA will be helpful to overcome some challenges in results presentation and answer organisation. For example, when search results are presented to the user and the user selects a result, perhaps transferring the user to that document itself by going through the content may not be helpful. Instead, the system could direct the user to the answer. Hence, not only does the system need to find the relevant document, but the relevant passage also has to be played [Ajmera et al. 2011]. Furthermore, the use of "information units" instead of a "ranked list of documents" may support the new paradigm of conversational search correctly [Allan et al. 2012].

The research mentioned in the this section on information presentation strengthens the suggestion that more sophisticated presentation techniques, for example, through conversations, are needed to support the user with their different information needs. Thus, just translating text search results into audio is not desirable. Even though users may be able to identify the most relevant result, this simplistic translation is not optimised for audio-only presentations of search results, and it hurts user experience. Thus, similarly as in the query formulation stage, we observe differences between the modality of the interaction, audio versus visual. These differences warrant further investigation into presenting results through conversations.

1.4.3 Query Reformulation and Refinements

Query reformulation and refinement is a key search strategy to modify a previous search query in order to optimise the retrieved information. Users can reformulate their query themselves or use suggestions from the system. For example, users can refine, disambiguate, or embellish their query through accepting query auto-complete suggestions located in the query box during their query formulation. Furthermore, they can accept clarifying suggestions from the system on the SERP through query facets or related search examples as seen in Figure 1.8.

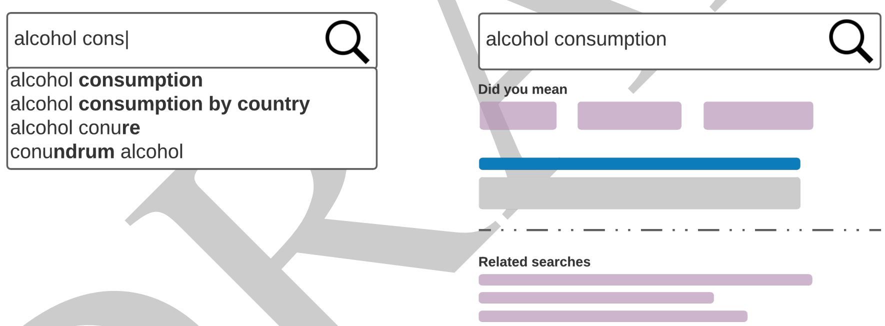

Figure 1.8 Different techniques to help the user formulate their query. Query suggestions while the user is typing before submitting a query or consequent refinement or clarifying suggests on the SERP.

These query reformulation or refinement techniques provide guidance for users to resubmit a query. Users can take advantage of spelling correction or related query suggestion informing them about the available information space. Many researchers have suggested that these query suggestions can be seen as a first step to making information systems truly interactive. For example, it has been suggested that these clarifications or suggestions could overcome reformulation problems by proactively providing them [Aliannejadi et al. 2019, Rosset et al. 2020, Zamani et al. 2020a].

Clarifying intents has recently become a popular research topic and can be seen as a method of making web search interactive [Aliannejadi et al. 2019, Braslavski et al. 2017, Kiesel et al. 2018, Zamani et al. 2020a,b]. In a traditional web-based search box, the user is mainly in control over the content they are consuming. For example, if a ranked list is retrieved, including suggestions, the user can opt for only inspecting the results while ignoring these suggestions altogether. Even though these suggestions may not be consumed, displaying them is reasonably straight forward. In contrast, in a conversational setting and especially in a voice-only environment, the "displaying" of query suggestions is not straightforward [Kiesel et al. 2018]. Instead, the screen real-estate is smaller, making every move from the system more essential. Nevertheless, in an audio-only setting, reading out all these suggestions may overwhelm the user. However, a study by Kiesel et al. [2018] suggested that voice assistants should always ask for clarification when ambiguous queries are detected and that users in general do not mind the interruption and thus not harming the interactions.

Further investigation of this query reformulation stage suggests additional opportunities to guide the user through reformulation queries proactively. For example, through search assistance and progression, a system could suggest switching search engines (academic search versus regular search), proactively ask for information judgements to adapt the information space [Trippas et al. 2018], or elicit more information when the system is not confident, before providing a response [Qu et al. 2019a, Trippas et al. 2018]. Conversational systems can also track and predict the user's intent, to proactively suggest search directions [Qu et al. 2018, 2019a]. Indeed, the interactive nature of this stage invites for further proactivity from the system, and thus, minimising the cognitive burden on the user.

As suggested in the results presentation and answer organisation section (Section 1.4.2), boundaries between the document and SERP presentation is fading. While those boundaries may be changing, other boundaries such as query formulation and reformulation are suggested to be declining too [Liao et al. 2020]. These ever-fading boundaries add to the needed interactivity and transparency of the system to predict when a query is related to the current search session, or the user has moved on to a new topic.

In summary, similar atomic functions which are seen in traditional browser-based web search are observed in conversational search as well. Nevertheless, further investigation on more advanced support for query formulation is needed. Similarly, refined results presentation techniques without just translating results into speech for spoken conversational search (i.e., audiofying) are needed too. Thus, supporting the user through query reformulations while leveraging interactivity are crucial for good user experience in conversational search. Finally, all the introduced research in Section 1.4 reinforce the need to investigate, design, and evaluate conversational interactions centred around how users would interact in a human–human conversation.

1.5 Fundamental Non-search Actions In Conversational Search

In the previous section, different aspects of conversational search interactions for query formulation, results presentation, and query reformulation were discussed. In this section, we focus on non-search interactions which can take place between the user and system as specified in the SCoSAS model (see Figure 1.5). More precisely, we disseminate non-task specific interactions of search which have been studied in a spoken conversational search setting, that is *(1)* discourse management (Section 1.5.1), *(2)* navigation (Section 1.5.2), *(3)* grounding (Section 1.5.3), and *(4)* visibility of system status (Section 1.5.4). These four aspects have been empirically identified in spoken conversational search. Intuitively, future search systems, even if they use non-speech functions, could integrate aspects of audio-only specific suggestions.

1.5.1 Discourse Management

In a conversational setting between two humans, we encounter discourse interactions (i.e., the communication of a series of linked utterances). Some of these utterances are part of the search interactions while others are part of the conversational coherence and cohesion between the speakers. This section discusses the latter; this includes utterances to check whether a message has been sent and understood by a communication partner [Schiffrin 1985]. As seen in the next example, Speaker 2 is verifying what they had heard, enabling Speaker 1 to understand what Speaker 2 had understood:

Speaker 1: Could you take the potatoes out of the fridge and put them into the oven?

Speaker 2: Sorry, the tomatoes? What do you want me to do...

Speaker 1: No, the potatoes, can you put them in the oven?

These interactions are often seen as an interruption in the communication flow and thus are linked to the repair of communication breakdowns [Gilabert et al. 2009]. However, it has also been suggested that these repairs can have a positive side-effect and thus influencing the information elicitation. For example, the communication repairs can contribute to negations about the subject and help the information exchange [Pica 1994].

In traditional web search, where information is exchanged in a text form (i.e., through written queries and SERPs) these repair interactions are absolute. That is because the user themselves can skim and revisit what they submitted as query and results retrieved by the system. However, when we change the interaction mode from a visual to an audio-only setting, these repair actions to manage the discourse become crucial for the information exchange [Trippas 2019]. That is, Trippas et al. [2020a] observed different functions of the discourse management utterances, for example, often an information request was echoed, asked to repeat a previous utterance, or confirmed a command.

Furthermore, these discourse utterances were also observed to be independent of the role, that is either an information seeker (i.e., user) or system can utilise them. We, therefore, suggest that linguists can play an important role to help validate these conversational systems in collaboration with IR researchers and thus promoting interdisciplinary research.

1.5.2 Navigation

Navigation is a fundamental element of being human. Walking to the local grocer, cycling to university, or searching for friends in the park, all these movements suggest navigation. According to the Cambridge Dictionary[2], navigation is the science of finding a way from one place to another linked to transport. Furthermore, with the rise of electronic documents, it is also defined as the act of moving around within a website or between websites [Maglio and Matlock 1999]. Research in information visualisation has suggested that presenting information in 3D can help shift the user's cognition load to the human perceptual system [Robertson et al. 1991].

Thus, the aim of navigational commands or utterances in conversational search is to enable the searcher to manoeuvre around an online information space. Supporting this movement between information units enables the searcher in their information exploration or finding. Even though navigational interactions in web search and information visualisation have received much attention, only limited conversational search research hints towards navigation and information spaces [Azzopardi et al. 2018, Trippas et al. 2020a, Vtyurina et al. 2020].

It has been suggested through empirical observations that a searcher utilises a different kind of navigational actions for spoken conversational search. For example, searchers can access specific sources, navigate or traverse between documents, single out particular documents or topic, and read more from a document or the next document [Trippas et al. 2020a]. Imagine a user wants to know about the *health benefits of milk*. The system could collect all possible information such as benefits, disadvantages, nutritional information, or pasteurisation processes of cow's milk as a first topic, see Figure 1.9. The system could also collect all this information for other topics such as almond milk (Topic 2) as part of the information space. Other less convenient links within the information space of milk health befits could be the impact of milk on climate change from greenhouse gasses (Topic 3). The system now needs to support the user help create a mental model of the information space and support the understanding of complex links and structural information unit relationships. Indicating relationships could potentially be done through non-speech sounds or discourse markers similarly as discussed in Section 1.2.1.

Further work through interviews by Vtyurina et al. [2020] has also suggested that a conversational system should support searchers' conceptualisation of an information space. For example, searchers often create a mental map of which paths or landmarks they have come across. Thus a system should support the wayfinding of information units and understand the referencing of these landmarks [Trippas 2019, Vtyurina et al. 2020].

The advantages of allowing and supporting the user in a conversational area to roam and navigate are that they have the freedom to explore information spaces. However, from a system perspective, it is difficult to distil a multi-dimensional field of information to a narrow

[2] https://dictionary.cambridge.org/dictionary/english/navigation

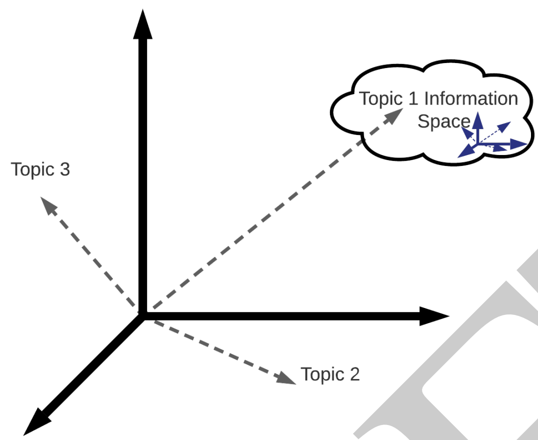

Figure 1.9 Visualisation of a possible information space in audio results.

audio-only channel and support the user's movement through the multi-dimensional field. Understanding how people conceptualise information space is vital to enable users to find their information in an accessible manner. Thus, further research could investigate how searchers navigate the electronic world of information units in an audio-only setting, how to support searchers to refind information, and how to visualise this multi-dimensional navigational interaction.

1.5.3 Grounding

Clark and Brennan [1991] defined grounding in communication as "sharing and synchronising mutual beliefs and assumptions" and is seen as a fundamental action in communication between people. In a recent conversational search study, it has been suggested that grounding is used by conversational participants to identify whether they have understood each other [Trippas et al. 2020a]. For example, by making sure the

- listener understood the speaker;
- speaker knows the listener understood them;
- listener knows the speaker knows the listener understood them.

That is, grounding in conversational search can involve communication actions such as paraphrasing the information a seeker had understood as a way to synchronise the information seeking process. Furthermore, it was observed that these grounding functions allowed the information-seeker to share their beliefs and values around information units, utilising them as feedback. These grounding interactions were thus supporting the system to update the information knowledge model of the searcher.

Grounding can also help as a collaborative activity to update the belief that someone (i.e., the user or system) has and their mental model of the information space or their interlocutor. As a consequence, this grounding process aids the continuous updating of these respective models. Nevertheless, this feedback loop of grounding is under-examined for conversational search, and further investigations are needed to understand how systems can employ and support these grounding functions.

1.5.4 Visibility of Information-seeking Partner Status

One of the ten heuristics of user interface design by Nielsen [2005] states that "The system should always keep users informed about what is going on, through appropriate feedback within reasonable time." through visibility of system status. Thus, interactions between the system and user are about communication and transparency of each other's state, providing some predictability and control.

It has been suggested that feedback on the information-seeking partner and thus sharing on what was taking place on a users' side in conversational search will need to be implemented in future systems [Trippas et al. 2020a]. It has been observed that these feedback loops can be manifested in different ways, for example, by informing whether they are processing information or are overwhelmed by the information. Furthermore, it has also been suggested that a partner's status could be shared about whether certain information had already been encountered. Thus, the visibility of the partner's status helps the actors in the conversation to assess each others' state. Further research is needed to conceptualise how to communicate a partner's state in a conversational search interaction without disrupting the knowledge flow.

1.6 Implementing and Evaluating of Conversational Systems

Despite the popularity of conversational search, the existing systems are limited and have basic functionality [Dalton et al. 2018a]. Furthermore, current systems such as Google Assistant or Amazon Alexa provide their frameworks and platforms, which means that interactions can not be easily compared between systems, and research findings need to be generalised with caution. As a way of simplification, conversational search can be interpreted as a task-oriented dialogue with the information-seeking as the task and can follow the SDS modular architecture as described in Section 1.2.2. This SDS modular architecture is the pipeline, with each module being responsible for part of the task [Walker et al. 1997]. However, unlike very defined tasks in SDS, searching for information is more complex and user–system relationship is different. For example, users often do not know how to express their information needs or precisely what they are looking for, resulting in more complex dialogue than SDS. Simultaneously, the system knows which information is available, however, the user does not. Now, multi-turn interactions are needed to elicit the required information from the user and reveal which information is available to search, thus defining the information space.

All of the above makes developing conversational systems more challenging. One step towards studying conversational search is by creating new frameworks so that other researchers can also A/B test, log interactions, or build test collections [Dalton et al. 2018a]. One of these platforms is Macaw, a high-level architecture for conversational search as an open-source modular framework created by Zamani and Craswell [2020]. This modular design enables the study of conversational approaches, including mixed-initiative interactions, document retrieval, recommendation and question answering. Researchers can change each of the modules or new modules can be added, enabling ongoing research. For example, as seen in Figure 1.10, the user interface specific modules can be changed or different databases could be inserted to store all the interactions.

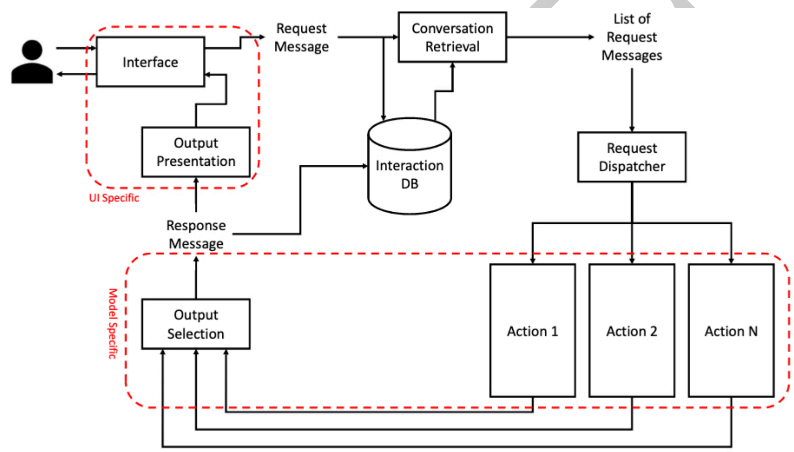

Figure 1.10 Macaw's high-level architecture by Zamani and Craswell [2020].

1.6.1 Evaluation of Conversational Systems

Evaluating IR systems, including conversational systems, in itself is a challenging task [Liu et al. 2021, Penha and Hauff 2020]. In addition, the evaluation of these systems is an ever-evolving research area. The evaluation of conversational systems often follows a similar path as traditional IR systems and is usually divided into offline and online approaches [Zamani et al. 2022].

1.6.1.1 Offline Evaluation

Traditionally, system-oriented evaluation in IR followed the Cranfield paradigm in which evaluation was conducted offline [Cleverdon et al. 1966]. In offline evaluation, the performance is measured on a static collection with pre-recorded responses or ground truth given a query and a set of documents to calculate the ranked list. This means that offline evaluation enables the reproducibility of evaluation, allowing to compare different systems but often seen as a *simplification* of the problem and user interactions. One offline evaluation project which is used for conversational search is the TREC CAsT [Dalton et al. 2018b, 2020]. CaST provides researchers with a large-scale reusable test collection for conversational search, including labelled relevance assessments, conversational queries, and information needs. The task has contributed to problems for conversational search such as ranking, conversational language understanding, and context.

With the move to conversational systems, which is inherently more interactive, there are many different aspects to measure than queries, documents, or ranked lists. For example, how do we incorporate multi-turn conversations beyond turn-based evaluation? How do we measure end-to-end interactions? How do we measure all the different kinds of conversations possible?

1.6.1.2 Online Evaluation

Online evaluation, in contrast to offline evaluation, aims to include real users, studying how they interact with a system and thus helping researchers to understand real-world interactions [Hofmann et al. 2016]. The advantage of online evaluation is that the user–system interactions can be observed while interacting with the system and thus evaluated holistically and robustly. Furthermore, online evaluations are more predictive of real-world system performance and likely show the current limitations. However, because online evaluation requires a working system enabling the observation of user interactions, it means that online evaluation methods usually do not have ground truth annotations [Balog 2021]. Furthermore, since there are limited working conversational systems, studying conversational systems is often restricted to researchers at companies that provide conversational systems (i.e., Google, Amazon, Apple). However, more recently, Amazon has taken the lead in giving non-Amazon researchers access to Amazon's platform to advance conversational systems through the Alexa Prize Challenge [Bowden et al. 2019]. This challenge is to advance conversational systems such as achieving natural, coherent and engaging dialogues. While this challenge is a step in the right direction, it is still limited to social conversations (i.e., chit-chat) and less around search. Similarly, as with the offline evaluation, the online evaluation will need to overcome many issues such as: How do we incorporate the user's feedback in the online evaluation in an ethical manner? How do we scale online evaluation? How can we integrate live user feedback without disrupting their search task?

1.6.2 Data for Conversational Systems

Much data is needed to train and evaluate conversational systems. Not only does a dataset for conversational search consist of queries, documents, and labels, it often requires much more information, such as turns taken and previous context. Furthermore, instead of providing the full documents, the shorter output is given in many existing conversational datasets, mimicking a conversational system output. These more concise outputs can be a short text passage or a reply which provides the direct answer.

Most of the existing datasets are not based on users interacting with existing systems. Instead, several efforts have created datasets from observational collections [Thomas et al. 2017, Trippas and Thomas 2019, Trippas et al. 2018], existing forum conversations [Penha et al. 2019, Qu et al. 2018], or crowdsourcing [Liao et al. 2021]. However, this poses the question of how representative these datasets are, whether they are biased towards particular settings, or whether the topics discussed represent real interactions. For example, two conversational search datasets which were developed with very similar protocols, SCSdata [Trippas et al. 2018] and MISC [Thomas et al. 2017], the authors of these datasets unidentified a range of subtle differences between them. They also encouraged researchers to be cautious of conversational search dataset re-use [Trippas and Thomas 2019].

In summary, more efforts are needed on evaluating the end-to-end process beyond simply interaction success. Even though many reusable test-sets have been made for conversation tasks, many of these datasets simplify the difficulties of interactions and make assumptions about user behaviour and system capabilities [Zamani et al. 2022]. Indeed, new datasets and protocols to create these datasets to enable better reproducibility and validity are needed. Lastly, the simulation of users for evaluation promises new avenues [Balog 2021, Maxwell and Azzopardi 2016]. Recent advancements in speech recognition technology and natural language understanding are leading to renewed attention on digital personal assistants, such as Apple's Siri, Microsoft Cortana, and the Google Assistant. Each ofthese assistant platforms provides its own framework to create agents [Dalton et al. 2018a].

Recent advancements in speech recognition technology and natural language understanding are leading to renewed attention on digital personal assistants, such as Apple's Siri, Microsoft Cortana, and the Google Assistant. Each ofthese assistant platforms provides its own framework to create agents [Dalton et al. 2018a].

1.7 Conclusion

As demonstrated in this chapter, research on conversations is not new in IIR. However, there is a resurgence of interest in searching through conversations, including in the theoretical aspects of abstracting and defining conversational interactions. The resurgence has previously been referred to as the *conversational search revolution* [Trippas 2019]. We reviewed existing research on conversational models highlighting that in order to build systems for conversational search, it is essential to define expected and allowable actions during the search conversation. Furthermore, we show that conversational search interactions are taken differently than traditional text-based search, suggesting that new approaches are needed for this unique search paradigm. We outlay multiple non-search and discourse interactions which could be used to support users in their search. In summary, since searching for information is inherently an interactive procedure, interactivity is key to overcome imposed difficulties.

1.7.1 Challenges on Developing Conversational Systems

The highly interactive conversational search paradigm is a promising and exciting research frontier. While this new wave of conversational system interest has been moving very quickly in the last few years, many challenges still need to be addressed. How can we create a usable end-to-end conversational system? How can we scale the testing of these systems with real users? Should we incorporate empathy into conversations, how, and in which situations? How do we pro-actively support the user in their information seeking journey? How can we make conversational systems transparent and interpretable for everyone (which will be increasingly complex with neural models)? How do we create conversational systems that adapt to new environments, deal with multi-party seekers, refind and revisit answers, or identify interaction costs?

1.8 Summary

In this chapter, we reviewed prior studies related to conversational search and how this conversational search differs from spoken or web-based search. We began this chapter by explaining where conversational systems are located historically, and the vision people had for these systems. We then reviewed previous research in IIR, highlighting the importance of studying existing models for information through dialogue and new theoretical frameworks and properties of conversational search. Concerning the conversational information seeking process and its fundamental search actions, we outlined how these actions have been studies and which other research opportunities exist. In addition, we examine the actions which take place in non-search interactions with respect to supporting the search process. The implications of genuine conversational search systems and interactions are plentiful, from making searching for information more intuitively to making information more accessible. We identified some challenges for implementing and evaluating conversational systems. Finally, we emphasise the importance of interactivity to overcome imposed difficulties, include investigations of non-search interactions to optimise user experience, and include multi-disciplinary research to advance in conversational search.

Bibliography

E. Agapie, G. Golovchinsky, and P. Qvarfordt. 2013. Leading people to longer queries. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 3019–3022.

J. Ajmera, A. Joshi, S. Mukherjea, N. Rajput, S. Sahay, M. Shrivastava, and K. Srivastava. 2011. Two-stream indexing for spoken web search. In *Proceedings of World Wide Web Conference (WWW)*, pp. 503–512.

M. Aliannejadi, H. Zamani, F. Crestani, and W. B. Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 475–484.

J. Allan, B. Croft, A. Moffat, and M. Sanderson. 2012. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012 the Second Strategic Workshop on Information Retrieval in Lorne. *ACM SIGIR Forum*, 46(1): 2–32.

A. Anand, L. Cavedon, H. Joho, M. Sanderson, and B. Stein. 2020. Conversational Search (Dagstuhl Seminar 19461). *Dagstuhl Reports*, 9(11): 34–83.

J. Arguello, L. Cavedon, J. Edlund, M. Hagen, D. Maxwell, M. Potthast, F. Radlinski, M. Sanderson, L. Soulier, B. Stein, J. Teevan, J. Trippas, and H. Zamani. 2020. Defining conversational search. In A. Anand, L. Cavedon, H. Joho, M. Sanderson, and B. Stein, eds., *Conversational Search (Dagstuhl Seminar 19461)*, pp. 49–55. Dagstuhl.

A. Arnold, G. Dupont, C. Kobus, F. Lancelot, and Y.-H. Liu. 2020. Perceived usefulness of conversational agents predicts search performance in aerospace domain. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, CUI '20.

B. Arons. 1993. SpeechSkimmer: interactively skimming recorded speech. In *Proceedings of User Interface Software and Technology (UIST)*, pp. 187–196. ACM Press.

L. Azzopardi, M. Dubiel, M. Halvey, and J. Dalton. 2018. Conceptualizing agent-human interactions during the conversational search process. In *SIGIR 2nd International Workshop on Conversational Approaches to Information Retrieval (CAIR'18)*. 8 pages.

K. Balog. 2021. Conversational ai from an information retrieval perspective: Remaining challenges and a case for user simulation.

N. J. Belkin, H. M. Brooks, and P. J. Daniels. 1987. Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies*, 27(2): 127–144.

N. J. Belkin, C. Cool, A. Stein, and U. Thiel. 1995. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems with Applications*, 9(3): 379–395.

T. Berners-Lee, J. Hendler, O. Lassila, et al. 2001. The semantic web. *Scientific American*, 284(5): 28–37.

P. Borlund. 2013. Interactive information retrieval: An introduction. *Journal of Information Science Theory and Practice*, 1(3): 12–32.

K. K. Bowden, J. Wu, W. Cui, J. Juraska, V. Harrison, B. Schwarzmann, N. Santer, S. Whittaker, and M. Walker. 2019. Entertaining and opinionated but too controlling: a large-scale user study of an open domain alexa prize system. In *Proceedings of the 1st International Conference on Conversational User Interfaces*, pp. 1–10.

P. Braslavski, D. Savenkov, E. Agichtein, and A. Dubatovka. 2017. What do you mean exactly? analyzing clarification questions in cqa. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, p. 345–348.

L. Cavedon, B. Fröhlich, H. Joho, R. Song, J. Teevan, J. Trippas, and E. Yilmaz. 2020. Scenarios that invite conversational search. In A. Anand, L. Cavedon, H. Joho, M. Sanderson, and B. Stein, eds., *Conversational Search (Dagstuhl Seminar 19461)*, pp. 66–69. Dagstuhl.

C. G. Čech and S. L. Condon. 1998. Message size constraints on discourse planning in synchronous computer-mediated communication. *Behavior Research Methods, Instruments, & Computers*, 30(2): 255–263.

P. Christmann, R. Saha Roy, A. Abujabal, J. Singh, and G. Weikum. 2019. Look before you hop: Conversational question answering over knowledge graphs using judicious context expansion. In *Proceedings of International Conference on Information and Knowledge Management (CIKM)*, CIKM '19, p. 729–738.

A. Chuklin, A. Severyn, J. R. Trippas, E. Alfonseca, H. Silen, and D. Spina. 2018. Prosody modifications for question-answering in voice-only settings. *arXiv preprint arXiv:1806.03957*, pp. 1–5.

A. Chuklin, A. Severyn, J. R. Trippas, E. Alfonseca, H. Silen, and D. Spina. 2019. Using audio transformations to improve comprehension in voice question answering. In *Proceedings of International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF)*, pp. 164–170. Springer.

H. H. Clark and S. E. Brennan. 1991. Grounding in communication. In *Perspectives on Socially Shared Cognition*, pp. 222–233.

C. W. Cleverdon, J. Mills, and E. M. Keen. 1966. Factors determining the performance of indexing systems, (Volume 1: Design). *Cranfield: College of Aeronautics*.

P. R. Cohen and S. L. Oviatt. 1995. The role of voice input for human-machine communication. *Proceedings of the National Academy of Sciences*, 92(22): 9921–9927.

F. Crestani and H. Du. 2006. Written versus spoken queries: A qualitative and quantitative comparative analysis. *Journal of the American Society for Information Science and Technology*, 57(7): 881–890.

W. B. Croft and R. H. Thompson. 1987. I3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science*, 38(6): 389–404.

J. S. Culpepper, F. Diaz, and M. D. Smucker. 2018. Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *ACM SIGIR Forum*, 52(1): 34–90.

E. Cutrell and Z. Guan. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 407–416.

J. Dalton, V. Ajayi, and R. Main. 2018a. Vote goat: Conversational movie recommendation. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, SIGIR '18, pp. 1285–1288. Association for Computing Machinery, New York, NY, USA.

J. Dalton, C. Xiong, and J. Callan, 2018b. The TREC 2019 Conversational Assistance Track (CAsT). http://treccast.ai/. (Last accessed January 24, 2019).

J. Dalton, C. Xiong, and J. Callan. 2020. Trec cast 2019: The conversational assistance track overview. *arXiv preprint arXiv:2003.13624*.

Y. Deldjoo, J. R. Trippas, and H. Zamani. 2021. Towards multi-modal conversational information seeking. In *Proceedings of the ACM Conference on Research and Development in Information Retrieval (SIGIR'21)*, p. 1577–1587.

M. Dubiel, M. Halvey, L. Azzopardi, D. Anderson, and S. Daronnat. 2020. Conversational strategies: impact on search performance in a goal-oriented task. In *The Third International Workshop on Conversational Approaches to Information Retrieval*.

A. Frummet, D. Elsweiler, and B. Ludwig. 2019. Detecting domain-specific information needs in conversational search dialogues. In *Proceedings of the 3rd Workshop on Natural Language for Artificial Intelligence*, volume 2521.

N. Gao, D. W. Oard, and M. Dredze. 2017. Support for interactive identification of mentioned entities in conversational speech. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 953–956.

D. Gardner-Bonneau and H. E. Blanchard. 2007. *Human factors and voice interactive systems*. Springer.

D. Gibbon, R. Moore, and R. Winski. 1997. *Handbook of Standards and Resources for Spoken Language Systems*. de Gruyter.

R. Gilabert, J. Barón, and À. Llanes. 2009. Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47(3-4): 367–395.

S. Gosper, J. R. Trippas, H. Richards, F. Allison, C. Sear, S. Khorasani, and F. Mattioli. 2021. Understanding the utility of digital flight assistants: A preliminary analysis. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, CUI '21.

M. Gupta and M. Bendersky. 2015. Information retrieval with verbose queries. *Foundations and Trends in Information Retrieval*, 9(3-4): 209–354.

I. Guy. 2016. Searching by talking: Analysis of voice queries on mobile web search. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 35–44.

I. Guy. 2018. The characteristics of voice search: Comparing spoken with typed-in mobile web search queries. *ACM Transactions on Information Systems (TOIS)*, 36(3): 30:1–30:28.

M. Hearst. 2009. *Search User Interfaces*. Cambridge University Press.

D. Hiemstra, C. Hauff, and L. Azzopardi. 2017. Exploring the query halo effect in site search: leading people to longer queries. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 981–984.

R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino, and Y. Matsuo. 2014. Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 928–939.

K. Hofmann, L. Li, and F. Radlinski. 2016. Online evaluation for information retrieval. *Foundations and trends in information retrieval*, 10(1): 1–117.

J. Jiang and N. Ahuja. 2020. Response quality in human-chatbot collaborative systems. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*.

M. Kaiser, R. S. Roy, and G. Weikum. 2020. Conversational question answering over passages by leveraging word proximity networks. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*. ACM.

M. Kaisser, M. A. Hearst, and J. B. Lowe. 2008. Improving search results quality by customizing summary lengths. *Proceedings of Association for Computational Linguistics (ACL)*, pp. 701–709.

T. Kenter and M. de Rijke. 2017. Attentive memory networks: Efficient machine reading for conversational search. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*.

J. Kiesel, A. Bahrami, B. Stein, A. Anand, and M. Hagen. 2018. Toward voice query clarification. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 1257–1260.

J. Kiesel, D. Spina, H. Wachsmuth, and B. Stein. 2021. The meant, the said, and the understood: Conversational argument search and cognitive biases. In *Proceedings of the 3rd Conference on Conversational User Interfaces*, CUI '21, pp. 1–5.

S. R. Klemmer, A. K. Sinha, J. Chen, J. A. Landay, N. Aboobaker, and A. Wang. 2000. Suede: a Wizard of Oz prototyping tool for speech user interfaces. In *Proceedings of User Interface Software and Technology (UIST)*, pp. 1–10.

L. Liao, L. H. Long, Z. Zhang, M. Huang, and T.-S. Chua. 2021. Mmconv: An environment for multimodal conversational search across multiple domains. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 675–684.

Q. V. Liao, W. Geyer, M. Muller, and Y. Khazaen. 2020. Conversational interfaces for information search. In *Understanding and Improving Information Search*, pp. 267–287. Springer.

Z. Liu, K. Zhou, and M. L. Wilson. 2021. Meta-evaluation of conversational search evaluation metrics. *arXiv preprint arXiv:2104.13453*.

X. Lu, S. Pramanik, R. Saha Roy, A. Abujabal, Y. Wang, and G. Weikum. 2019. Answering complex questions by joining multi-document evidence with quasi knowledge graphs. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, p. 105–114.

P. P. Maglio and T. Matlock. 1999. The conceptual structure of information space. In A. J. Munro, K. Höök, and D. Benyon, eds., *Social navigation of information space*, pp. 155–173. Springer.

G. Marchionini. 1997. *Information Seeking in Electronic Environments*. Cambridge University Press.

D. Maxwell and L. Azzopardi. 2016. Agents, simulated users and humans: An analysis of performance and behaviour. In *Proceedings of International Conference on Information and Knowledge Management (CIKM)*, pp. 731–740.

D. Maxwell, L. Azzopardi, and Y. Moshfeghi. 2017. A study of snippet length and informativeness: Behaviour, performance and user experience. In *Proceedings of Conference on Research and*

Development in Information Retrieval (SIGIR), pp. 135–144.

M. McTear, Z. Callejas, and D. Griol. 2016. *The Conversational Interface*. Springer.

J. Nielsen, 2005. Ten usability heuristics. https://www.nngroup.com/articles/ten-usability-heuristics/. (Last accessed January 2, 2019).

R. N. Oddy. 1977. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1): 1–14.

K. Ong, K. Järvelin, M. Sanderson, and F. Scholer. 2018. Qwerty: The effects of typing on web search behavior. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 281–284.

G. Penha and C. Hauff. 2020. Challenges in the evaluation of conversational search systems. In *Converse@ KDD*.

G. Penha, A. Balan, and C. Hauff. 2019. Introducing mantis: a novel multi-domain information seeking dialogues dataset. *arXiv preprint arXiv:1912.04639*.

T. Pica. 1994. Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44(3): 493–527.

C. Qu, L. Yang, W. B. Croft, J. R. Trippas, Y. Zhang, and M. Qiu. 2018. Analyzing and characterizing user intent in information-seeking conversations. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 989–992.

C. Qu, L. Yang, W. B. Croft, Y. Zhang, J. R. Trippas, and M. Qiu. 2019a. User intent prediction in information-seeking conversations. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 25–33.

C. Qu, L. Yang, M. Qiu, Y. Zhang, C. Chen, W. B. Croft, and M. Iyyer. 2019b. Attentive history selection for conversational question answering. In *Proceedings of International Conference on Information and Knowledge Management (CIKM)*, CIKM '19, p. 1391–1400.

F. Radlinski and N. Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 117–126.

P. Ren, Z. Chen, Z. Ren, E. Kanoulas, C. Monz, and M. de Rijke. 2020. Conversations with search engines. *arXiv preprint arXiv:2004.14162*.

G. G. Robertson, J. D. Mackinlay, and S. K. Card. 1991. Cone trees: animated 3d visualizations of hierarchical information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 189–194.

C. Rosset, C. Xiong, X. Song, D. Campos, N. Craswell, S. Tiwary, and P. Bennett. 2020. Leading conversational search by suggesting useful questions. In *Proceedings of World Wide Web Conference (WWW)*, p. 1160–1170.

N. G. Sahib, D. Al Thani, A. Tombros, and T. Stockman. 2012. Accessible information seeking. *Proceedings of Digital Futures*, 12: 1–3.

J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope. 2010. "Your word is my command": Google search by voice: a case study. In *Advances in Speech Recognition*, pp. 61–90. Springer.

D. Schiffrin. 1985. Conversational coherence: The role of well. *Language*, 61(3): 640–667.

S. Sitter and A. Stein. 1992. Modeling the illocutionary aspects of information-seeking dialogues. *Information Processing & Management*, 28(2): 165–180.

A. Smirnova. 2020. Word order communicates user intent in search queries. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 1–8.

D. Spina, J. R. Trippas, P. Thomas, H. Joho, K. Byström, L. Clark, N. Craswell, M. Czerwinski, D. Elsweiler, A. Frummet, S. Ghosh, J. Kiesel, I. Lopatovska, D. McDuff, S. Meyer, A. Mourad, P. Owoicho, S. P. Cherumanal, D. Russell, and L. Sitbon. July 2021. Report on the future conversations workshop at chiir 2021. *ACM SIGIR Forum*, 55(1). ISSN 0163-5840.

A. Stein and E. Maier. 1995. Structuring collaborative information-seeking dialogues. *Knowledge-Based Systems*, 8(2-3): 82–93.

H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami. 2013. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pp. 334–338.

R. S. Taylor. 1962. The process of asking questions. *American Documentation*, 13(4): 391–396.

P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. 2017. MISC: A data set of information-seeking conversations. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*. 6 pages.

P. Thomas, M. Czerwinski, D. McDuff, N. Craswell, and G. Mark. 2018. Style and alignment in information-seeking conversation. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 42–51.

P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. 2020. Expressions of style in information seeking conversation with an agent. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 1171–1180.

J. R. Trippas. 2015. Spoken conversational search: Information retrieval over a speech-only communication channel. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, p. 1067.

J. R. Trippas. 2019. *Spoken Conversational Search: Audio-only Interactive Information Retrieval*. PhD thesis, RMIT, Melbourne.

J. R. Trippas and P. Thomas. 2019. Data sets for spoken conversational search. In *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019). CEUR-WS*, pp. 14–18.

J. R. Trippas, D. Spina, M. Sanderson, and L. Cavedon. 2015a. Results presentation methods for a spoken conversational search system. In *CIKM'15 First International Workshop on Novel Web Search Interfaces and Systems (NWSearch'15)*, pp. 13–15.

J. R. Trippas, D. Spina, M. Sanderson, and L. Cavedon. 2015b. Towards understanding the impact of length in web search result summaries over a speech-only communication channel. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 991–994.

J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. 2017. How do people interact in conversational speech-only search tasks: A preliminary analysis. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 325–328.

J. R. Trippas, D. Spina, L. Cavedon, H. Joho, and M. Sanderson. 2018. Informing the design of spoken conversational search. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 32–41.

J. R. Trippas, D. Spina, P. Thomas, M. Sanderson, H. Joho, and L. Cavedon. 2020a. Towards a Model for Spoken Conversational Search. *Information Processing & Management*, 57(2): 102162.

J. R. Trippas, P. Thomas, D. Spina, and H. Joho. 2020b. Third international workshop on conversational approaches to information retrieval (CAIR'20): Full-day workshop at chiir 2020. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, p. 492–494.

J. R. Trippas, D. Spina, M. Sanderson, and L. Cavedon. 2021. Accessing media via an audio-only communication channel: A log analysis. In *Proceedings of the ACM International Conference on Conversational User Interfaces (CUI 2021)*, pp. 1–6.

M. Turunen, J. Hakulinen, N. Rajput, and A. A. Nanavati. 2012. *Evaluation of Mobile and Pervasive Speech Applications*, pp. 219–262. John Wiley & Sons.

S. Vakulenko. 2019. *Knowledge-based Conversational Search*. PhD thesis, TU Wien.

S. Vakulenko, K. Revoredo, C. Di Ciccio, and M. de Rijke. 2019. QRFA: A data-driven model of information-seeking dialogues. In *Proceedings of the European Conference on Information Retrieval (ECIR)*, pp. 541–557.

E. M. Voorhees. 1999. The trec-8 question answering track report. In *Trec*, volume 99, pp. 77–82.

A. Vtyurina. 2019. Towards non-visual web search. In *Proceedings of Conference on Information Interaction and Retrieval (CHIIR)*, pp. 429–432.

A. Vtyurina, C. L. Clarke, E. Law, J. R. Trippas, and H. Bota. 2020. A mixed-method analysis of text and audio search interfaces with varying task complexity. In *Proceedings of International Conference on the Theory of Information Retrieval (ICTIR)*, pp. 61–68.

M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella. 1997. PARADISE: A framework for evaluating spoken dialogue agents. In *Proceedings of European Chapter of the Association for Computational Linguistics (EACL)*, pp. 271–280.

M. A. Walker, R. Passonneau, and J. E. Boland. 2001. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *Proceedings of Association for Computational Linguistics (ACL)*, pp. 515–522.

J. Weizenbaum. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1): 36–45.

W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, and A. Stolcke. 2018. The microsoft 2017 conversational speech recognition system. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5934–5938.

L. Yang, M. Qiu, C. Qu, J. Guo, Y. Zhang, W. B. Croft, J. Huang, and H. Chen. 2018a. Response ranking with deep matching networks and external knowledge in information-seeking conversation systems. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 245–254.

Y. Yang, Y. Gong, and X. Chen. 2018b. Query tracking for e-commerce conversational search: A machine comprehension perspective. In *Proceedings of International Conference on Information and Knowledge Management (CIKM)*, CIKM '18, p. 1755–1758.

N. Yankelovich, G. Levow, and M. Marx. 1995. Designing speechacts: Issues in speech user interfaces. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 369–376.

J. Yi and F. Maghoul. 2011. Mobile search pattern evolution. In *Proceedings of World Wide Web Conference (WWW)*, pp. 165–166.

H. Zamani and N. Craswell. 2020. Macaw: An extensible conversational information seeking platform. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 2193–2196. Association for Computing Machinery.

H. Zamani, S. T. Dumais, N. Craswell, P. N. Bennett, and G. Lueck. 2020a. Generating clarifying questions for information retrieval. In *Proceedings of World Wide Web Conference (WWW)*.

H. Zamani, B. Mitra, E. Chen, G. Lueck, F. Diaz, P. N. Bennet, N. Craswell, and S. T. Dumais. 2020b. Analyzing and learning from user interactions for search clarification. In *Proceedings of Conference on Research and Development in Information Retrieval (SIGIR)*.

H. Zamani, J. R. Trippas, J. Dalton, and F. Radlinski. 2022. Conversational information seeking. *Foundations and Trends in Information Retrieval*.