

# CAIR'17: First International Workshop on Conversational Approaches to Information Retrieval at SIGIR 2017

Hideo Joho  
University of Tsukuba  
[hideo@slis.tsukuba.ac.jp](mailto:hideo@slis.tsukuba.ac.jp)

Lawrence Cavedon  
RMIT University  
[lawrence.cavedon@rmit.edu.au](mailto:lawrence.cavedon@rmit.edu.au)

Jaime Arguello  
University of North Carolina at Chapel Hill  
[jarguell@email.unc.edu](mailto:jarguell@email.unc.edu)

Milad Shokouhi  
Microsoft  
[milads@microsoft.com](mailto:milads@microsoft.com)

Filip Radlinski  
Google  
[filiprad@google.com](mailto:filiprad@google.com)

## Abstract

The First International Workshop on Conversational Approaches to Information Retrieval (CAIR'17) was held on August 11th, 2017 in Tokyo, in association with SIGIR 2017, the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. CAIR'17 brought together academic and industry researchers to present and discuss new research on conversational approaches to search, particularly over speech. As speech interfaces to devices are becoming more pervasive, and with the advent of Personal Digital Assistants (PDAs), investigating and designing more effective and efficient means of interacting with information sources and search engines is growing in importance. Spoken interfaces to search engines opens opportunities through the possibility of greater interactivity but also requires rethinking, e.g., how results are presented. CAIR'17 addressed a range of such issues via invited keynote presentations, research papers presented orally and via posters, and a panel discussion involving researchers from interactive IR, IR evaluation, and spoken dialogue systems.

## 1 Introduction

The First International Workshop on Conversational Approaches to Information Retrieval (CAIR'17) <sup>1</sup> [2] was held in Tokyo on August 11th, 2017 as a full-day workshop associated with SIGIR 2017. The workshop was organised in response to the growth in speech interfaces,

---

<sup>1</sup><https://sites.google.com/view/cair-ws>

---

particularly on mobile devices, and to Personal Digital Assistants (PDAs) such as Siri, Cortana, Google Assistant, Alexa and others, and the opportunities afforded by this burgeoning mode of interaction. Interactive spoken interfaces provide opportunities for more efficient interaction with services and information; however, understanding and design of effective use of speech interfaces specifically for Information Retrieval (IR) and Presentation has been under-investigated. Current systems can handle single queries and can seamlessly present an answer over speech to factoid-style questions. However, the question of presenting answer lists to complex information requests has been little explored. Similarly, effective interaction patterns supported by speech information retrieval, as well as strategies employed by human users on such problems in associated tasks, such as collaborative search have been little explored.

The Call for Papers specifically welcomed reports on techniques to support complex and multi-turn user-machine dialogues for information access and retrieval, and multi-modal interfaces for interacting with such systems; studies of human-human interaction (e.g., collaborative search) to inform the design of conversational search applications; and approaches to evaluation of conversational approaches. Such issues have been studied within the Spoken Dialogue Systems research community, although not so much in the context of the rich informational requests over unstructured data typically of interest to IR research. Hence, representatives of the dialogue community and the associated special interest group, SIGdial (Special Interest Group on Discourse and Dialogue)<sup>2</sup>, were invited to participate. The workshop attracted over 70 registered participants.

## 2 Program at a glance

The workshop program included two keynotes, eight oral presentations in three sessions, poster presentations, and a panel discussion, along with a hands-on session with commercial speech-based interaction systems. The structure of the workshop was as follows.

- Keynote Talk I: Search Failed? Let's talk (Ron Kaplan, A9.com)
- Hands on with commercial devices
- Oral Session 1: Datasets
- Panel discussion on evaluation of conversational search
- Keynote Talk II: End-to-end learning for task-oriented conversational systems (Jason D. Williams, Microsoft Research)
- Oral Session 2: Models and Interactions
- Oral Session 3: User Cases and Evaluation
- Open Discussion + Closing

## 3 Keynote speakers

Two keynote presenters were invited, both with experience in spoken dialogue research and with current industry research affiliations.

---

<sup>2</sup><http://www.sigdial.org>

---

The first keynote speaker **Ron Kaplan** is currently Vice President and Chief Scientist at A9, Amazon's search capability; he was previously VP and Chief Scientist at Nuance, CTO and Chief Scientist at Powerset (and similarly at Bing after Powerset's acquisition by Microsoft). Over the course of three decades at Xerox PARC, Ron was one of the leading researchers in the field of Natural Language Processing.

The second keynote speaker **Jason Williams** performs research on statistical approaches to spoken dialogue systems. He is a Principal Researcher at Microsoft Research and was previously at AT&T Research. His PhD at Cambridge University involved novel approaches to spoken dialogue management via the use of POMDPs. Jason is currently the president of SIGdial.

## 4 Panel

While a number of different topics were considered for the panel discussion, the organisers decided that Evaluation presented both a critical and pressing need. The panel brought together researchers that provided strong expertise on topics related to Conversational IR, including Interactive IR, Evaluation over Complex Queries, and Spoken Dialogue Systems Evaluation. The ensuing discussion covered ground over techniques, frameworks, and the need for resources to support evaluation.

Chair

- Mark Sanderson (RMIT University)

Panelists

- Nick Belkin (Rutgers University)
- James Allan (University of Massachusetts at Amherst)
- Jason D. Williams (Microsoft Research)
- Emine Yilmaz (University College London)

## 5 Main themes

The rest of this report aims to highlight the main themes that emerged from the keynote presentations, oral presentations, panel and open discussions. It should be noted that many ideas and suggestions given in this report came from the active audience, who are named explicitly here.

### 5.1 Concept and novelty of conversational search (CS)

Given that this was the first edition of the CAIR workshop, a theme frequently visited during the workshop was defining the concept of conversational search itself. There was some level of consensus that simply replacing text-based inputs with voice-based inputs would not make a search conversational. Although it is premature to provide a clear definition of CS, the following is seed ideas and elements discussed during the workshop.

An interesting contrast was made with conventional search engines, where the returned search results can be presented augmented with rich factoid information. Effective access to

---

such a rich SERP via speech-based channels would be impossible since speech-based access is necessarily sequential, and thus does not allow the whole picture to be shown at once. In addition, a user's expectation of voice-based search engines could be influenced by the experience of current SERP interfaces.

On the other hand, presentations involved examples of strong cases where conversational interaction between users and systems can solve common but complicated search tasks. For example, Kaplan presented a product search scenario where a user has a task of buying a garbage disposal unit, with a number of choices that are difficult for a non-expert to select between (such as wattage and noise levels). Conversational search systems could help elicit such a user's needs through explanations of the choices being made. Another case is where a user does not know the name of a product, and can only offer a description of his or her need. In particular, Kaplan discussed desirable properties of conversations: incremental, educational, explanatory, and expeditious.

The panel session on evaluation also raised key questions regarding the concept and novelty of CS. Why focus on speech-based approaches to IR interaction? What are the differences from user-librarian interaction scenarios? Finally, literature mentioned as a useful source includes Nordlie [8] and Radlinski and Craswell [9].

## 5.2 CS behaviour and interaction

Given that conversational search is different from conventional search, how does it affect people's searching behaviour or interaction between users and search systems? This was a second emerging theme in the workshop and several papers provided insights as follows.

Arguello et al. [1] analyzed the differences between queries (or information requests) produced using voice (versus text) and intended for a human intermediary (versus a search engine). Additionally, the authors focused on search tasks associated with a domain knowledge constraint – the user requires information that is appropriate for either a domain novice or expert. Results found that both factors (referred to as the *medium* and *target* of the information request) had an effect on several outcomes, including: (1) participants' perceptions about their own request, (2) the level of natural language structure of the request, and (3) extent to which participants explicitly stated a preference for novice- or expert-appropriate information. Results from this study suggest that higher expectations about a system's ability to understand and respond to requests may influence users to say more about their information need. However, such additional information may not necessarily produce better search results using current retrieval techniques.

Mehrotra et al. [7] presented a log analysis of the Microsoft Cortana intelligent assistant for Windows 10. The analysis suggests important differences between how users interact with an intelligent assistant such as Cortana and how users interact with a web search engine in the desktop domain. First, results suggest a need for a shorter inactivity threshold for identifying session boundaries – the evidence suggests that the 30-minute inactivity threshold typically used for web search may be too long. Second, as might be expected, voice queries were slightly longer than textual queries. Third, most queries were associated with desktop search (e.g., search for local files), followed by general web search, search for short answers (e.g., weather, how-to, and language translation), and system commands (e.g., reminders, text messages, alarms, music controls, etc.). Finally, short sessions were more often associated with desktop search, while longer sessions were more often associated with system commands.

Lui et al. [6] presented a position paper on specific use cases where dialogue-based in-

---

teraction is appropriate and the different challenges associated with conversational IR. In terms of appropriate use cases, the authors cited query disambiguation, as well as eliciting information about facet-value pairs that may be relevant to the user’s task, such as asking about geographical location. In terms of key challenges, the authors cited the following: (1) the need for anaphora resolution (e.g., being able to resolve pronouns used to refer to an entity mentioned in a previous query), (2) automatically ignoring terms that are not central to the query topic (e.g., “find information about...”), (3) deciding between different types of system responses such as eliciting information or presenting search results, (4) deciding when to respond with a specific answer versus search results, (5) maintaining an internal representation of the conversation, and (6) using common sense knowledge to infer important missing terms that should be added to a query before issuing a search.

### 5.3 Description, prediction, and explanation of CS

Another set of work presented during the workshop was around the theme of how to describe the complexity of CS, and how to develop models that can learn from search interactions.

A necessary step towards developing conversational IR systems is to further understand information-seeking conversations between a searcher and an agent. One approach is to observe and analyze how people interact with a *human* (rather than a computer) playing the role of the “conversational search agent”. Several papers presented at the workshop took this general approach.

Thomas et al. [10] introduced the Microsoft Information-Seeking Conversations (MISC) dataset. In this study, 22 participant pairs completed 4 tasks of varying levels of difficulty and complexity. One participant played the role of the “seeker” and the other played the role of the “intermediary” (and had access to a web search engine). The participants coordinated via an audio link and a video link that was focused on their faces. After completing each task, participants answered questionnaires about their emotions, level of effort, and level of engagement. Additionally, the dataset includes automatic annotations of participants’ communication exchanges, voice signals, and facial expressions. The MISC dataset is freely-available for researchers to download and investigate a wide range of questions: What does an intermediary do? What types of intermediary-seeker interactions predict positive outcomes related to emotions, effort, and engagement? What types of seeker behaviors signal (dis)satisfaction? What types of tactics do seekers adopt and are they similar to the types of tactics adopted by users interacting with more traditional, non-conversational search systems?<sup>3</sup>

Trippas et al. [11] presented a set of guidelines for transcribing and conducting qualitative analysis of information-seeking conversations between human “seekers” and “intermediaries”. In this paper, the authors outlined a series of best practices for two important data analysis tasks: (1) transcribing information-seeking conversations recorded over an audio channel, and (2) performing qualitative coding of conversational data. The authors describe the development of detailed codes to characterize the goal of each utterance and broader codes to characterize more general, information-seeking themes observed in the data. The methodology described in the paper is being used to analyze a set of recorded information-seeking conversations, and preliminary results were presented at CHIIR 2017 [13].

An important challenge for a conversational IR system is be able to make inferences and

---

<sup>3</sup><https://www.microsoft.com/en-us/download/details.aspx?id=55594>

---

answer questions using its conversational history with a user. Kenter and de Rijke [4] presented a novel and efficient algorithm referred to as an Attentive Memory Network (AMN). The proposed AMN algorithm is a sequence-to-sequence neural network and was evaluated using 20 machine reading datasets that contain stories (of 2 to 100 sentences long) and questions/answers about those stories. The 20 datasets are associated with different aspects of natural language understanding, such as counting, deduction, induction, and spatial reasoning. The AMN algorithm achieved a similar level of performance as compared to several baselines and required less computation, making it well-suited for conversational IR.

## 5.4 Evaluation of CS

The theme of evaluation served an important role during the workshop to reflect the novelty of CS as discussed earlier. In addition to the panel discussion on evaluation, two research presentations shed light on this topic.

Kiseleva and de Rijke [5] presented a conceptual paper about the challenges of evaluating intelligent personal assistants on mobile devices, especially as they become more “conversational”. The authors describe how the user- and context-specific nature of intelligent assistants makes it difficult to objectively define a “gold standard” output for a given situation. The authors advocate that future work should focus on inferring user satisfaction using a combination of signals derived from visual data (e.g., facial expressions), gesture-based interactions (e.g., touch paths, touch density, and touch velocity), and voice signals.

An important challenge for a conversational IR system is how to best summarize search results for a user over an audio only channel. Trippas et al. [12] presented two case studies on how to assess the quality of an audio summary for a given search result. Specifically, the paper describes two methodologies for gathering quality assessments of audio summaries over a crowdsourcing platform. The authors describe one interface for gathering quality judgements where the assessor sees the underlying text and listens to different summaries, and a second interface where the assessor uses an audio summary to make a relevance judgement. Additionally, the authors propose different ways of measuring an assessor’s mental workload while listening to an audio summary as a means to indirectly evaluate its quality. The authors propose using the well-established NASA TLX questionnaire and also by considering the assessor’s interactions with the audio player.

During the panel discussions, it was suggested to explore existing or ongoing test collections as potential resources for the evaluation of CS (e.g., TREC Tasks Track). The monograph “Methods for Evaluating Interactive Information Retrieval Systems with Users” [3] was also recommended as a useful starting point for the design of CS evaluation.

## 6 Outlook

As can be seen from the summary of the presentations and discussion, future research in conversational search is wide-open and promising. To support ongoing research on the topic of conversational search, we have instituted two further activities.

One is a resource guide<sup>4</sup>; Conversational search is still an early stage, and thus, compiling relevant materials (e.g., papers, workshops, tutorials, dataset) can facilitate the development

---

<sup>4</sup><https://github.com/ConversationalSearch/ResourceGuide/wiki>

---

of the community. Maintaining the guide is expected to be a community-wide effort, and the wiki can be edited by anyone with a GitHub account.

Another is the special topic issue on Conversational Approaches to Information Retrieval at the Journal of the Association for Information Science and Technology<sup>5</sup>. The guest editors of the special topic issue are Jaime Arguello (University of North Carolina at Chapel Hill), Maarten de Rijke (University of Amsterdam), Hideo Joho (University of Tsukuba), and Damiano Spina (RMIT). Paper submissions are due on February 23, 2018.

As mentioned in the Introduction, the workshop also successfully initiated a collaboration with SIGdial community, and we agreed to explore further collaboration by considering a liaison between SIGIR and SIGdial.

We plan to organise a second edition of CAIR workshop in a relevant conference. More details about the CAIR2017 workshop can be found in the ACM blog<sup>6</sup>.

## 7 Conclusions

Voice is a compelling medium for communication with devices and the upcoming years is likely to see an increase in capability and ubiquity of voice-based interaction, driving the need and opportunity for conversational approaches to information search. The popularity of the workshop and the lively nature of the discussions suggests that this is recognised by the SIGIR community. However, the workshop demonstrated there is much work to do to understand the nature of conversational search and, in particular, to develop methodologies and resources for evaluation: developing these is an important challenge for the SIGIR/CAIR research community in the near future.

The CAIR workshop organisers thank SIGdial for offering an opportunity for a student travel grant. We would also like to acknowledge the following people for their contributions to the successful CAIR'17 workshop: Ron Kaplan (Amazon), Jason D. Williams (Microsoft Research), Mark Sanderson (RMIT University), Nick Belkin (Rutgers University), James Allan (University of Massachusetts at Amherst), Emine Yilmaz (University College London), Fernando Diaz (Spotify), Dilek Hakkani-T'ur (Google), Damiano Spina (RMIT University), Fabio Crestani (University of Lugano), Roi Blanco (University of A Coruña), Jeffrey Dalton (Google), Ido Guy (Yahoo Research), Jiepu Jiang (UMass Amherst), Gareth Jones (Dublin City University), Ben Lambert (Spotify), Karthik Raghunathan (MINDMELD), Ruihua Song (Microsoft), Amanda Stent (Yahoo), Paul Thomas (Microsoft), and Johanne Trippas (RMIT University).

## References

- [1] J. Arguello, B. Choi, and R. Capra. Factors Affecting Users' Information Requests. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 4
- [2] H. Joho, L. Cavedon, J. Arguello, M. Shokouhi, and F. Radlinski. First International Workshop on Conversational Approaches to Information Retrieval (CAIR'17). In *Pro-*

---

<sup>5</sup>CFP: [http://onlinelibrary.wiley.com/store/10.1002/\(ISSN\)2330-1643/asset/homepages/JASIST\\_CAIR\\_CFP.pdf](http://onlinelibrary.wiley.com/store/10.1002/(ISSN)2330-1643/asset/homepages/JASIST_CAIR_CFP.pdf)

<sup>6</sup> <https://cacm.acm.org/blogs/blog-cacm/220407-keynotes-at-sigir-2017/fulltext>

---

*ceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 1423–1424, New York, NY, USA, 2017. ACM. 1

- [3] D. Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Found. Trends Inf. Retr.*, 3(1–2):1–224, Jan. 2009. 6
- [4] T. Kenter and M. de Rijke. Attentive Memory Networks: Efficient Machine Reading for Conversational Search. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 6
- [5] J. Kiseleva and M. de Rijke. Evaluating personal assistants on mobile devices. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 6
- [6] Z. Liu, Z.-Y. N. Niu, J.-Y. Nie, H. Wu, and H. Wang. Conversations in IR: Its Role and Utility. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 4
- [7] R. Mehrotra, A. Hassan Awadallah, A. E. Kholy, and I. Zitouni. Hey Cortana! Exploring the use cases of a Desktop based Digital Assistant. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 4
- [8] R. Nordlie. “user revealment”—a comparison of initial queries and ensuing question development in online searching and in human reference interactions. In *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 11–18, New York, NY, USA, 1999. ACM. 4
- [9] F. Radlinski and N. Craswell. A theoretical framework for conversational search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, pages 117–126, New York, NY, USA, 2017. ACM. 4
- [10] P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. MISC: A data set of information-seeking conversations. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 5
- [11] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. A Conversational Search Transcription Protocol and Analysis. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 5
- [12] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. Crowdsourcing User Preferences and Query Judgments for Speech-Only Search. In *SIGIR 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*, 2017. 6
- [13] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. How do people interact in conversational speech-only search tasks: A preliminary analysis. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, pages 325–328. ACM, 2017. 5