

The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI'22)

Penha, Gustavo; Vakulenko, Svitlana; Dusek, Ondrej; Clark, Leigh; Pal, Vaishali; Adlakha, Vaibhav

DOI

[10.1145/3477495.3531700](https://doi.org/10.1145/3477495.3531700)

Publication date

2022

Document Version

Final published version

Published in

SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval

Citation (APA)

Penha, G., Vakulenko, S., Dusek, O., Clark, L., Pal, V., & Adlakha, V. (2022). The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI'22). In *SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3466-3469). (SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval). ACM. <https://doi.org/10.1145/3477495.3531700>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI'22)

Gustavo Penha
Delft University of Technology
Delft, Netherlands
g.penha-1@tudelft.nl

Svitlana Vakulenko
Amazon Alexa AI
Barcelona, Spain
svvakul@amazon.com

Ondřej Dušek
Charles University
Prague, Czech Republic
odusek@ufal.mff.cuni.cz

Leigh Clark
Swansea University
Swansea, United Kingdom
l.m.h.clark@swansea.ac.uk

Vaishali Pal
University of Amsterdam
Amsterdam, Netherlands
v.pal@uva.nl

Vaibhav Adlakha
Mila, McGill University
Montréal, Canada
vaibhav.adlakha@mila.quebec

ABSTRACT

The goal of the seventh edition of SCAI (<https://scai.info>) is to bring together and further grow a community of researchers and practitioners interested in conversational systems for information access. The previous iterations of the workshop already demonstrated the breadth and multidisciplinary inherent in the design and development of conversational search agents. The proposed shift from traditional web search to search interfaces enabled via human-like dialogue leads to a number of challenges, and although such challenges have received more attention in the recent years, there are many pending research questions that should be addressed by the information retrieval community and can largely benefit from a collaboration with other research fields, such as natural language processing, machine learning, human-computer interaction and dialogue systems. This workshop is intended as a platform enabling a continuous discussion of the major research challenges that surround the design of search-oriented conversational systems. This year, participants have the opportunity to meet in person and have more in-depth interactive discussions with a full-day onsite workshop.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval; Presentation of retrieval results; Evaluation of retrieval results.**

KEYWORDS

conversational search, conversational information access, information-seeking dialogue

ACM Reference Format:

Gustavo Penha, Svitlana Vakulenko, Ondřej Dušek, Leigh Clark, Vaishali Pal, and Vaibhav Adlakha. 2022. The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI'22). In *SIGIR '22: ACM Symposium on Neural Gaze Detection, July 11–15, 2018, SIGIR, Madrid*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3477495.3531700>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '22, July 11–15, 2022, SIGIR, Madrid

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8732-3/22/07.

<https://doi.org/10.1145/3477495.3531700>

1 OVERVIEW

The emergence of devices such as smartphones that enable conversational interactions with personal assistants, coupled with successes in machine learning technologies has led to a new branch of multi-disciplinary research generally known as conversational search. Unlike other dialogue related research areas with focus on chit-chat conversations, conversational search is concerned with information-seeking processes and comes from the search perspective. The shift from web search to conversational search leads to many challenges that remain unanswered. A search-oriented conversational system would be able for example to elicit the information need of the user with clarification questions, find relevant information from different and reliable sources and deliver it to the user through natural language interaction.

Conversational search was recognised as one of the most important topics during the Strategic Workshop in Information Retrieval (SWIRL) workshop 2018 [7], where 60 IR researchers reunited to discuss and identify long-range issues in the field of Information Retrieval. It is currently an active research field as evidenced by a number of related workshops and publications in the top-tier retrieval journals and conferences. For example, the list of accepted papers from SIGIR 2021¹ contains a total of 21 papers with “dialogue” and “conversation” in their titles.

The recent special issue of TOIS 2021 was dedicated to conversational search and three of those papers were already presented in the last year’s SCAI workshop. Most recently, two independent manuscripts were published with the goal to provide a comprehensive overview of the current state of research on conversational search as well as the neural approaches proposed for the individual subtasks [10, 16]. All of this indicates a rapid growth of conversational search as a field and a boosting interest in the relevant topics across different research communities, including Information Retrieval (IR), Machine Learning (ML), Artificial Intelligence (AI), Natural Language Processing (NLP) and Human Computer Interaction (HCI). In particular, an independent conference focused on Conversational User Interfaces (CUI) was established as a spin-off of the ACM CHI Conference on Human Factors in Computing Systems community, highlighting the importance of this topic also for HCI researchers.

¹<https://sigir.org/sigir2021/accepted-papers/>

SCAI'22 will bring together researchers across a wide range of disciplines, opening up the frontiers of IR to absorb and impact the ideas that are currently discussed upon the research agendas of NLP, ML and HCI research. This initiative enables a rich discussion that will stimulate further development of the field of search-oriented conversational AI. Thereby, it also follows closely the tradition established at the Dagstuhl seminar 19461 of November 2019 [3], where 44 researchers across different fields first met to discuss the research agenda for conversational search. Amongst the topics discussed in this seminar were defining, evaluating and modeling conversational search, as well as different application scenarios for conversational search, e.g. e-learning technologies, scholarly assistants etc. Three organisers (Svitlana, Ondřej and Leigh) met at this seminar and decided to pursue this interdisciplinary collaboration on a regular basis, continuing the tradition of the already established SCAI workshop. While interdisciplinary in nature, the core research interest of SCAI has always remained centered on the same goal: **satisfying an information need by means of a conversational interface**.

Furthermore, SCAI has a continuous tradition of organising shared tasks relevant to conversational search scenarios, such as clarifying question generation (ClariQ [2]) in 2020 and conversational QA (SCAI-QReCC [14]) in 2021. This year, we continue to further develop the infrastructure and methodologies required for evaluation of conversational QA. The task of conversational QA builds on the conversational passage retrieval task (see, e.g., TREC CAsT setup [8]) and extends it to also produce a concise answer based on the retrieved passages, which was shown to be a challenging task for evaluation in the previous task we organised last year, which relates to the challenge of evaluating natural language generation performance in general [14]. This year we invest further into this research direction and extend our evaluation with an interactive human evaluation phase, similar to previous ConvAI dialogue evaluation challenges [9] and evaluation of interactive IR systems.

In the following section, we provide more details on the topics for this edition and elaborate more on the purpose of the workshop. In Section 3, we describe relations to the previous SCAI editions, followed by the workshop format in Section 4. The details of the shared task and the organizers' bios are provided in Sections 5 and 6, respectively.

2 THEME AND PURPOSE

Our workshop focuses on different aspects of a conversational search system including a broad range of topics related to conversational search, such as:

- theoretical models of conversational search with implications for design and evaluation methodology (e.g., successful interaction strategies);
- information retrieval and conditional text generation (e.g., results presentation to support information-seeking conversations);
- conversational recommendation (e.g., preference elicitation);
- mixed initiative (e.g., clarifying and suggested questions);
- online metrics, offline benchmarks and user simulation (e.g., dialogue satisfaction and search task success);

- fairness, transparency and trust (e.g., bias in search results);
- domain-specific applications (e.g., e-learning, cooking and shopping assistants);
- personalizing conversations (e.g., user modeling).

The purpose of the workshop is to connect researchers working on different aspects related to conversational search, stimulate them to discuss the current challenges and identify promising directions for future work.

3 PREVIOUS EDITIONS

SCAI is one of the first and long-standing workshops dedicated to conversational search. The first SCAI workshop was a full-day workshop at ICTIR 2017 [5], attracting 96 registrants, 8 invited talks, 6 contributed papers, and a discussion panel. The 2nd edition [6] was held at EMNLP 2018 (Brussels, Belgium) and received 22 submissions, of which 6 were accepted for oral presentation (27%) and 7 accepted as posters (59% total acceptance rate). The workshop attracted around 400 participants. The 3rd edition was a special off-cycle half-day event that only featured invited speakers from start-up and industries focusing on conversational technology. It was co-organized along with The Web Conference 2019 (San Francisco, USA). The 4th edition was held at IJCAI 2019 (Macao, China), received 7 valid submissions, 4 were accepted (57%). The 5th edition was held at EMNLP 2020, received 8 submissions and accepted 5 (62%). It also hosted the ClariQ shared task with 8 participating teams [2]. The 6th edition was held in 2021 as a fully online independent event. The format was shifted towards all invited paper presentations. We held 4 sessions with 15 oral paper presentations focused on evaluation, personalisation and question answering. Every session was concluded with a short panel discussion between all the presenters and the session chair. A demo session with 14 posters was organised on the GatherTown platform. SCAI'21 also featured a shared task on conversational question answering, SCAI-QReCC [14], and the three best approaches were presented in the final 5th session of the day. A more detailed report on the last year's edition appeared in the latest issue of the SIGIR Forum (December 2021).

Previous workshop venues were picked to increase participants' geographic and topical diversity, which served its purpose. The previous editions of the SCAI workshop also received a generous support from sponsors, such as Google, Microsoft, Facebook, Amazon, Bloomberg, and HuggingFace. This funding helped us to secure student travel grants and increase diversity. It was also used to collect new datasets that play an important role for advancing the research and help the community. This year, we aim to continue our vision of promoting diversity and serving as a discussion platform for the recent advances and open challenges in conversational search.

4 FORMAT

The workshop is a full day onsite event, also allowing for hybrid participation and broadcast on Youtube in real time. The schedule includes five sessions: two morning and two afternoon paper sessions, with a poster session over lunch. The regular sessions feature 3–4 invited presentations with a short panel discussion among all the presenters in the end of each session.

Table 1: Program committee.

Name	Affiliation
Mikhail Burtsev	MIPT & DeepPavlov
Aleksandr Chuklin	Google Research
Jeff Dalton	University of Glasgow
Claudia Hauff	Delft University of Technology
Evangelos Kanoulas	University of Amsterdam
Julia Kiseleva	Microsoft Research
Verena Rieser	Heriot-Watt University & ALANA AI
Xiaoyu Shen	Amazon Alexa AI
Stefan Ultes	Mercedes-Benz Research

Similarly to the previous year, invited speakers will present their most recent work. We selected highly relevant presentations that are cohesive within and between the sessions, which allows us to have a productive discussion by comparing and contrasting the presented papers. This approach also allowed us to attract the most recently published work, including yet unpublished but high-quality manuscripts. The workshop selected presenters based on the following criteria: recent papers (last 2 years) related to the field of search-oriented conversational AI (as highlighted in Section 2) that were accepted in top conferences and journals.

5 SHARED TASK

Similar to SCAI'21, we have a shared task on conversational question answering in this edition of SCAI. Our primary objective is to determine the capabilities of current models in real-world conversational question answering settings.

5.1 Datasets

The task is based on two open-domain conversational question-answering datasets – QReCC [4] and TopiOCQA [1]. The two datasets have complementary strengths, which makes this pairing suitable for the shared task. Both the datasets include training and evaluation data for two sub-tasks: (1) passage retrieval – the ability to fetch a passage relevant to the query from a knowledge source, and (2) answer generation – the ability to generate a correct response to the query, given the query and retrieved passage. In addition to this, QReCC also includes data for question rewriting – the task of converting contextualized conversational question into a stand-alone de-contextualized question that satisfies the same information need of the user. On the other hand, TopiOCQA has multiple human annotations per question and every answer is accompanied by a rationale, which is a contiguous span in the ground-truth document.

The two datasets also differ with respect to the information source – QReCC has a larger and more diverse collection of 54M passages based on Commoncrawl, whereas TopiOCQA has 26M passages based on Wikipedia. While the corpus of QReCC is closer to the real-world setting, implementing retrieval models on this huge corpus poses several computational challenges, especially for recent dense retrieval methods [11, 12, 15]. To lay a common ground and to encourage more participation, we will only focus on a subset of QReCC that is based on Wikipedia documents.

5.2 Evaluation

The systems will be evaluated in two stages – automatic evaluation and human evaluation. Similar to previous iteration, the following metrics will be used for various sub-tasks – ROUGE-1 for question rewriting, MRR for passage retrieval and EM and F1 for answer generation. One of the findings of previous edition was that modern QA models are already able to produce fluent answers but we can not always trust those answers to be correct. To estimate faithfulness of the system with respect to knowledge source, we will also be evaluating the generated answers on Knowledge-F1 [13] – a metric which measures overlap of system generated answer with the *knowledge on which the human grounded during dataset collection*. This knowledge is represented as the rationale span in TopiOCQA and the gold passage segment in QReCC.

Even with a wide range of automated metrics, human evaluation still remains crucial for the task due to presence of multiple ground truth answers and their accompanying evidence. We propose to have two-fold human evaluation – (1) dataset specific evaluation, (2) unconstrained interaction with the systems. The first part will be similar to previous edition, where the human annotators will annotate answer plausibility and answer faithfulness for a subset of system generated answers. In the second part, annotators will have an unconstrained interaction with the top five systems shortlisted based on automatic evaluation metrics. The annotators will only be given basic guidelines similar to those given to original annotators of QReCC and TopiOCQA. Similar to the first part, they will label plausibility and faithfulness for every system-generated answer during the interaction. This unconstrained interaction provides a chance for task participants to truly gauge how their system will perform in the real-world. It will also give an estimate of how well the system does when it relies on its own previously generated answers, rather than on gold human answers, which is typical in most conversational question-answering evaluation procedure.

The shared task will be hosted on Codalab.² The docker-based submission workflow of this platform encourages development of reproducible systems, while also allowing the organizers to automatically evaluate systems without revealing the test set.

6 ORGANIZERS

Svitlana Vakulenko (Amazon Alexa AI): Svitlana is an applied scientist in Alexa AI Web Information. She completed her PhD thesis on knowledge-based conversational search at TU Wien and spent two years at the University of Amsterdam as a postdoc. She was the main organiser of SCAI last year, is on the organising team of the DialDoc workshop at ACL 2022 and TREC Conversational Assistance Track (CAST) 2022. Her research work was published in the major IR and NLP venues: SIGIR, WSDM, TOIS, CIKM, ECIR, EMNLP. She also has an extensive experience working in the Semantic Web and Information Systems communities with her publications appearing in the major research venues: ISWC, ESWC, SEMANTICS, ICIS, ECIS.

Ondřej Dušek (Charles University): Ondřej is an Assistant Professor at Charles University in Prague, focusing on NLG and dialogue systems. After his PhD in Prague, he spent 2 years as

²<https://codalab.lisn.upsaclay.fr/>

postdoc at Heriot-Watt University in Edinburgh, where he co-organized the E2E NLG Challenge language generation shared task and co-developed the Amazon Alexa Prize finalist chatbot, Alana. He also was the organiser of last years' SCAI as well as the workshop on NLG evaluation at the INLG 2020 conference. Ondřej has recently secured a starting grant from the European Research Council (ERC) for his NLG research. The output of his project will be a tool able to quickly learn new topics for conversations in different languages.

Leigh Clark (Swansea University): Leigh is a Lecturer in Human-Computer Interaction at the Computational Foundry in Swansea University. His research examines the effects of voice and language design on speech interface interactions and how linguistic theories can be implemented and redefined in this context. He was the organiser of the last year's SCAI and is a co-founder of the international ACM In-Cooperation Conversational User Interfaces (CUI) conference series started in 2019.

Gustavo Penha (Delft University of Technology): Gustavo is a researcher in the fields of NLP, IR, ML and RecSys. He interned at Amazon's Alexa Shopping, researching explainability for voice product search. His PhD research (4th year at TU Delft) focuses on conversational search and recommendation. Gustavo published in IR, NLP and recommendation venues: ECIR, RecSys, EACL, CHIIR, and is reviewing for ECIR, SIGIR, WWW, CIKM.

Vaishali Pal (University of Amsterdam): Vaishali is a 2nd year PhD candidate at IRLab at the University of Amsterdam, focusing on the areas of NLP and IR. Previously, she interned in the Information Retrieval group at Naver Labs Europe, working on spoken dialogue state tracking, and at Deepset, implementing a dense passage retrieval component for their open-source framework. Her research interests are in NLG, generative QA and conversational systems.

Vaibhav Adlakha (Mila, McGill University): Vaibhav is a 2nd year PhD student at Quebec AI Institute (Mila) and McGill University, focusing on knowledge-grounded conversational agents. During the course of his research, he led multiple data collection tasks. Previously, he interned at NLP Group, Indian Institute of Technology, Delhi, working on information extraction and question-answering over knowledge graphs. He also worked on development of task-oriented chatbots at Samsung Research Institute, Bangalore.

REFERENCES

- [1] Vaibhav Adlakha, Shehzaad Dhuliawala, Kaheer Suleman, Harm de Vries, and Siva Reddy. 2020. TopiOCQA: Open-domain Conversational Question Answering with Topic Switching. In *arXiv preprint arXiv:2110.00768*.
- [2] Mohammad Aliannejadi, Julia Kiseleva, Aleksandr Chuklin, Jeff Dalton, and Mikhail S. Burtsev. 2020. ConvAI3: Generating Clarifying Questions for Open-Domain Dialogue Systems (ClariQ). *CoRR* abs/2009.11352 (2020).
- [3] Avishek Anand, Lawrence Cavedon, Matthias Hagen, Hideo Joho, Mark Sanderson, and Benno Stein. 2020. Conversational Search—A Report from Dagstuhl Seminar 19461. *arXiv preprint arXiv:2005.08658* (2020).
- [4] Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, and Srinivas Chappidi. 2021. Open-Domain Question Answering Goes Conversational via Question Rewriting. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2021).
- [5] Mikhail Burtsev, Aleksandr Chuklin, Julia Kiseleva, and Alexey Borisov. 2017. Search-Oriented Conversational AI (SCAI). In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR '17)*. 333–334.
- [6] Aleksandr Chuklin, Jeff Dalton, Julia Kiseleva, Alexey Borisov, and Mikhail Burtsev. 2018. Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI. In *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*.
- [7] J Shane Culpepper, Fernando Diaz, and Mark D Smucker. 2018. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in IJCNLP (swirl 2018). In *ACM SIGIR Forum*, Vol. 52. ACM New York, NY, USA, 34–90.
- [8] Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. 2020. CAsT 2020: The Conversational Assistance Track Overview. In *Proceedings of the Twenty-Ninth Text REtrieval Conference, TREC 2020, Virtual Event [Gaithersburg, Maryland, USA], November 16-20, 2020 (NIST Special Publication, Vol. 1266)*, Ellen M. Voorhees and Angela Ellis (Eds.). National Institute of Standards and Technology (NIST).
- [9] Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander H. Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, Shrimai Prabhumoye, Alan W. Black, Alexander I. Rudnicky, Jason Williams, Joelle Pineau, Mikhail S. Burtsev, and Jason Weston. 2019. The Second Conversational Intelligence Challenge (ConvAI2). *CoRR* abs/1902.00098 (2019).
- [10] Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. 2022. Neural Approaches to Conversational Information Retrieval. *CoRR* abs/2201.05176 (2022). <https://arxiv.org/abs/2201.05176>
- [11] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 6769–6781.
- [12] Yi Luan, Jacob Eisenstein, Kristina Toutanova, and Michael Collins. 2021. Sparse, Dense, and Attentional Representations for Text Retrieval. *Transactions of the Association for Computational Linguistics* 9 (2021), 329–345.
- [13] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval Augmentation Reduces Hallucination in Conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 3784–3803. <https://doi.org/10.18653/v1/2021.findings-emnlp.320>
- [14] Svitlana Vakulenko, Johannes Kiesel, and Maik Fröbe. 2022. SCAI-QReCC Shared Task on Conversational Question Answering. *CoRR* abs/2201.11094 (2022).
- [15] Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul N. Bennett, Junaid Ahmed, and Arnold Overwijk. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In *International Conference on Learning Representations*.
- [16] Hamed Zamani, Johanne R. Trippas, Jeff Dalton, and Filip Radlinski. 2022. Conversational Information Seeking. *CoRR* abs/2201.08808 (2022).