

Home
Study
Research
Collaborate
News and events
People
About

ACM ICTIR '22 Best Student Paper Award, Madrid, Spain

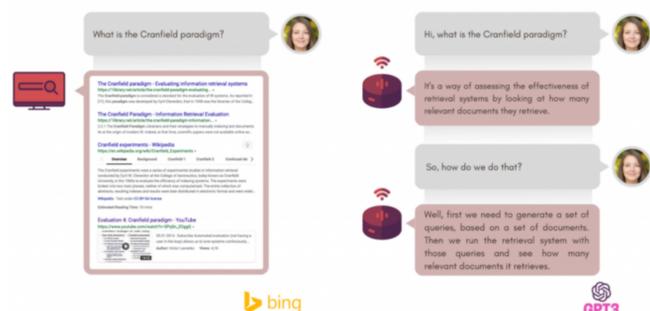
7 October 2022

We are delighted to announce that Xiao Fu won the Best Student Paper Award at the ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR 2022). This is the premier conference on theoretical information retrieval. This paper was co-authored with his supervisors, [Emine Yilmaz](#) and [Aldo Lipani](#). The awarded paper, "Evaluating the Cranfield Paradigm for Conversational Search Systems," links the scientific literature on the offline evaluation of search engines such as Google, Bing, and DuckDuckGo, originally developed in the 50s by Cyril Cleverdon at the University of Cranfield to the emerging research on conversational search systems such as Alexa and Siri (aka digital assistants).



The award (on the left) and Xiao presenting the paper at ICTIR (on the right).

An important research topic in this domain is the evaluation of these information systems before users are actually exposed to them. This topic, known as offline evaluation, has been studied a lot by the Information Retrieval community when talking about search engines but lacked a proper evaluation when talking about conversational search systems. This form of evaluation is fundamental for both industry and academia. This setting, in fact, gives the ability to research groups in academia to evaluate their approaches without having to employ a user base and to research groups from industry to discard bad systems before exposing them to users and risking a negative reaction.



The difference between a search system (on the left) and a conversational search system (on the right).

However, the sequential and interactive nature of conversations makes the application of traditional Information Retrieval methods require stronger assumptions. When building a test collection for search, it is fair to assume that the ratings provided by an annotator correlate well with the ratings perceived by an actual user of the search engine. However, when building a test collection for conversational search, we do not know if it is fair to assume that the rating provided by an annotator correlates well with the rating perceived by an actual user of the conversational search system.

The main question addressed by this paper is: Can we rely on assessors reading a conversation as we trust the assessment provided by the users interacting with the conversational system?

Answering this question is paramount because it underpins and guide

1. the development of more realistic user simulators for the evaluation of conversational systems, and
2. the design of more reliable and robust evaluation measures for conversational search systems.

The results in this paper show that there is a fair agreement between the two assessment methods (reading and interacting) and that these two kinds of assessments share similar conversational patterns.

Relevant Links

- The presentation of the paper is available [here](#).
- The paper is available [here](#).



Related News

- The Democratisation of Cat Models**
27 October 2022
- Online course entitled 'Engineering based fragility and vulnerability assessment' delivered**
28 September 2022

[UCL's first net-zero carbon building gets shortlisted at the 2022 Building Awards](#)
27 September 2022

UCL facilities

- About UCL
- Faculties and departments
- Library
- Museums and Collections
- UCL Bloomsbury Theatre
- UCL Shop

UCL locations

- Maps and buildings
- UCL and London
- UCL Global
- UCL East

Connect with UCL

- Alumni
- Businesses
- Media Relations
- Jobs
- Support us

