The Importance of Interaction in Information Retrieval

Bruce Croft
UMass Amherst and RMIT University
Continuing the Interaction Discussion

• Nick Belkin, Gerald Salton Award 2015, “People, Interacting with Information”

• Kalervo Jarvelin, Gerald Salton Award 2018, “Information Interaction in Context”

• Also an important part of the work of Norbert Fuhr (2012) and Sue Dumais (2009)
Two IR Research Communities?

**SYSTEM-ORIENTED**
- Researchers focused on “algorithms”, IR models and system implementation
- Ranking models, text representation, efficiency
- Computer Science viewpoint

**USER-ORIENTED**
- Researchers focused on the users and interfaces of IR systems
- How they use it, why they use it
- Information Science viewpoint
Retrieval models!
Common Ground

• Users have always been a central focus of IR
• Distinguished IR from database research and even AI
• Core concepts of IR are based on people
  • Information needs, relevance, feedback, browsing, evaluation
• Different views on the relative importance of the system
The IR Community Collaborating

Belkin and Croft, 1992
Interaction is Key

- Effective access to information often requires interaction between the user and the system
  - More than a “one-shot” query
  - Both the user and the system should play a role
- Even more effective information access requires a system that actively supports effective interaction
  - Modeling the interaction
  - Becomes more crucial in “limited bandwidth” scenarios such as mobile phones or voice-based systems
Example: Web search
Web Search

• Generally viewed as placing most of the burden for successful search on the user
  • e.g., query reformulation, browsing SERPs

• But, web search engines perform many functions to make browsing more effective
  • Query completion
  • Aggregated ranking
  • Query suggestion

• System has a more passive role in the interaction
Example: Golovchinsky et al, 1999
From reading to retrieval: Freeform ink annotations as queries.
Interacting with text

• User selects and annotates text in documents
  • Annotations then used as the basis for new queries
• Effective retrieval requires the system to use this feedback effectively in query generation and ranking
  • Lee and Croft, Generating queries from user-selected text. IIIX '12.
  • Sorig, Collignon, Fiebrink, and Kando, Evaluation of rich and explicit feedback for exploratory search. CHIIR ‘19.
• System still a passive partner in the interaction
Example: Conversational search
Conversational Search

• Always one of the ultimate goals of IR
• System clearly has an active role in the interaction
• Limited bandwidth of speech and screen means that the system’s role is crucial for success
What am I going to talk about?

• The importance of interaction for information retrieval: past, present and future
  • Historical overview
  • Interaction in question answering
  • Interaction in conversational search
  • Examples from CIIR
  • What needs to be done
A Short History of Interaction in IR

**Time**

- Boolean search systems
- Search strategies
- Cranfield evaluation studies
- Studies of information dialogues
- Natural language queries and ranking
- Browsing
- Relevance feedback
- Expert intermediaries
- Iterative relevance feedback
- Hypertext and links
- Result presentation
- Web search
- Query suggestion
- Clustering and visualization
- Information interaction in context
- Term weighting and highlighting
- Question answering
- Summaries and snippets
- Query transformation
- Information interaction in context
- Exploratory search
- Query log analysis
- Recommendation systems
- Search aggregation
- Conversational search
- Forums and CQA
- Evaluation of interactive systems
- Voice-based search
- Mobile search
Early Days

- Boolean search engines
- Indexing tools and thesauri
- Cranfield evaluation studies
- Search strategies
- Studies of information dialogues

H. M. Brooks and N. J. Belkin. Using discourse analysis for the design of information retrieval interaction mechanisms. SIGIR 83
Bates, M.J. Information Search Tactics. JASIS, 1979
Bates, M.J. The Design of Browsing and Berrypicking Techniques
information interaction and intelligent information provision mechanisms. Journal Inf. Sci. 1986
Understanding Intermediary Interactions

Brooks, Daniels and Belkin, 1986

<table>
<thead>
<tr>
<th>Name of function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Problem State (PS)</td>
<td>Determine position of user in problem treatment process, e.g., formulating problem, problem well-specified, etc.</td>
</tr>
<tr>
<td>2. Problem Mode (PM)</td>
<td>Determine appropriate mechanism capability, e.g., document retrieval</td>
</tr>
<tr>
<td>3. User Model (UM)</td>
<td>Generate description of user type, goals, beliefs, knowledge, etc., e.g., graduate student, thesis etc.</td>
</tr>
<tr>
<td>4. Problem Description (PD)</td>
<td>Generate description of problem type, topic, structure, environment, etc.</td>
</tr>
<tr>
<td>5. Dialogue Mode (DM)</td>
<td>Determine appropriate dialogue type and level for situation, e.g., menu, natural language</td>
</tr>
<tr>
<td>6. Retrieval Strategy (RS)</td>
<td>Choose and apply appropriate strategies to knowledge resource</td>
</tr>
<tr>
<td>7. Response Generator (RG)</td>
<td>Determine propositional structure of response to the user, appropriate to the situation</td>
</tr>
<tr>
<td>8. Explanation (EX)</td>
<td>Describe mechanism operation, restrictions etc. to user as appropriate</td>
</tr>
<tr>
<td>9. Input Analysis (IA)</td>
<td>Convert input from user into structures usable by functional experts</td>
</tr>
<tr>
<td>10. Output Generator (OG)</td>
<td>Convert propositional response to the form appropriate to user, situation, dialogue mode</td>
</tr>
</tbody>
</table>

**Utterance analysis**

<table>
<thead>
<tr>
<th>Utterance number</th>
<th>Subgoal acronym</th>
<th>Utterance analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>18a</td>
<td>RES</td>
<td>User begins to talk about his own country after completing his thesis</td>
</tr>
<tr>
<td>18b</td>
<td>UGOAL</td>
<td>User has the goal of returning to his country</td>
</tr>
<tr>
<td>19</td>
<td>ph</td>
<td>User describes his employment background</td>
</tr>
<tr>
<td>20</td>
<td>RACK</td>
<td>User refers to his state of knowledge on part of the thesis</td>
</tr>
<tr>
<td>21</td>
<td>ph</td>
<td>Plastic communication</td>
</tr>
<tr>
<td>22 &amp; 24a</td>
<td>KNOW</td>
<td>User continues to describe his research topic</td>
</tr>
<tr>
<td>22, 25, 26</td>
<td>ph</td>
<td>Plastic communication</td>
</tr>
<tr>
<td>24b, 27</td>
<td>RES</td>
<td>Intermediary attempts to imitate model of the user's state of knowledge</td>
</tr>
<tr>
<td>29 &amp; 31</td>
<td>ph</td>
<td>User confirms that the two states of knowledge appear to match</td>
</tr>
<tr>
<td>29 &amp; 30</td>
<td>MATCH</td>
<td></td>
</tr>
<tr>
<td>32, 34</td>
<td>KNOW</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>MATCH: KNOW</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>FRAME WORD-FOCUS SHIFT</td>
<td></td>
</tr>
</tbody>
</table>

**Utterances**

- UNo (.) what I'm interested in (.), 18a and I'm doing this, ah um/19
- U(.) in Nova Scotia (.) when I finish 18b and: (.) I
- I spent four years with the government. (.)
- I'm looking for something (.)
- U(outdoor) recreation so I/2
- I handle on what kinds of recreation?
- Yes/25
- U(.) but what I'm looking at now is
- I:
- Uaspects of um (.) recreation/29 by
- I: but you (.) you've (.) got an idea U
- I things go on (laughter) in the for
- Yeah/28
- Uthey/27 (.) I will have to look at
- I: yeah/
- Uaspects of um (.) recreation/29 by
- I but you (.) you've (.) got an idea U
- I things go on (laughter) in the for
- Yeah/34 um: (.) now (.) are you going to (.) look U yeh/33
- I at this on a world (.) world wide level [35] U
Ranking and Result Presentation

- Natural language queries and ranking
- Relevance feedback
- Term weighting and highlighting
- Summaries and snippets
- Clustering and visualization

Simplifying user interaction and providing information
# Ranking and Interaction

<table>
<thead>
<tr>
<th>Query Alteration Process</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-Search</strong></td>
<td></td>
</tr>
<tr>
<td>1. Repeated Concepts</td>
<td>User chooses query terms to be repeated for emphasis</td>
</tr>
<tr>
<td>2. Thesaurus Display</td>
<td>User chooses terms obtained from thesaurus display to update query (with or without time restrictions)</td>
</tr>
<tr>
<td>3. Word Frequency</td>
<td>User looks at display of word frequency information before updating query</td>
</tr>
<tr>
<td>4. Source Document</td>
<td>User looks at display of source document before updating</td>
</tr>
<tr>
<td><strong>Post-Search</strong></td>
<td></td>
</tr>
<tr>
<td>5. Title Display</td>
<td>User looks at titles of first five retrieved documents before updating</td>
</tr>
<tr>
<td>6. Abstract Display</td>
<td>User looks at abstracts of first five retrieved documents</td>
</tr>
<tr>
<td>7. Relevance Feedback</td>
<td>Query is updated automatically using relevance judgments supplied by user following an initial search</td>
</tr>
<tr>
<td><strong>Combined Methods</strong></td>
<td></td>
</tr>
<tr>
<td>8. Abstract plus Thesaurus</td>
<td>User looks at pre- and post-search information</td>
</tr>
</tbody>
</table>

Lesk and Salton, 1969. Interactive search and retrieval methods using automatic information displays

<table>
<thead>
<tr>
<th>Demands on Computer</th>
<th>Demands on User</th>
<th>Precision Improvement: Over Word Stem Match</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Recall</td>
<td>High Recall</td>
</tr>
<tr>
<td>A. Fully Automatic word stem match automatic thesaurus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td>none</td>
<td>—</td>
</tr>
<tr>
<td>normal</td>
<td>none</td>
<td>+4%</td>
</tr>
<tr>
<td>B. Pre-Search Interaction thesaurus display source document display</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal +</td>
<td>medium-high</td>
<td>+6%</td>
</tr>
<tr>
<td>normal +</td>
<td>medium</td>
<td>+8%</td>
</tr>
<tr>
<td>C. Post-Search Interaction title display abstract display relevance feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>medium</td>
<td>+13%</td>
</tr>
<tr>
<td>high</td>
<td>very high</td>
<td>+17%</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>+10%</td>
</tr>
</tbody>
</table>
Relevance Feedback Interactions

• Positive document examples
• Negative document examples
• Positive passage examples
• Positive and negative terms in documents
• Batch and incremental document feedback
Example: Golovchinsky et al, 1999
From reading to retrieval: Freeform ink annotations as queries.
Text Highlighting


Summaries and Snippets

You have just typed in the following query:

"Alternatives to Postscript."

A brief explanation about the query:

To be relevant, a document must identify one or more of the following: an alternative to Postscript, a page description language.

The following articles have been retrieved in relation to that query:


   As published in the Chicago Reader, an alternative source, Rocket A. Rocket, offers a new alternative strategy to go beyond the traditional strategy. With this alternative, the strategy is to publish books and sell them to the general public.

   Total sales in the first year were $1.5 million, which is more than Rocket's overall revenue. The books are sold for $20 each, which is more than Rocket can offer for a single book. Rocket uses this strategy to gain recognition and market share.

   [Click here to get the full article.]


   There was support from teachers with at least a bachelor's degree. The teacher who taught the class was a former teacher and is currently a high school teacher. The teacher who taught the class is a former teacher who has a bachelor's degree from the University of California, Berkeley. The teacher who taught the class is a former teacher who taught high school or university. The teacher who taught the class is a former teacher who taught high school or university.

   [Click here to get the full article.]

Google patent, 2005.

SIGIR 2019
Clustering in Research


Clustering in Commercial Systems

Browsing and Guided Assistance

1. Iterative search and dialogues
2. Expert intermediaries
3. Hypertext and links
4. Information interaction in context
5. Exploratory search

Active, dynamic system support for interaction
Identifying need to support more complex activities
There can be three parts to your statement (all optional):
1. Your reaction to the reference just shown (if any).
   This must come first:
   “Yes” or “No”
2. A selection from the names (authors) or terms shown, by number.
   The statement signifies rejection of all numbers that follow it.
3. New names or terms (terms preferably in quotes). The element should be separated by commas.

Examples: ‘posture’, ‘circulatory system’
   Yes, not 11, 12
   No, 7, 13, 4
   ‘heart rate’
   Yes
Press enter key when you are ready to proceed


A Positive End-Expiratory Pressure—Nasal-Assist Device (PEEP—NAD) for treatment of respiratory distress syndrome.; Tummons, Anesthesiology, 38, 592–5, June 73
   Yes, 13, not 6

We are not making progress.
Please reconsider this document:
1. R Markello, 2. R Schuder, . . .
   .
   .
   18. respiration, 19. surgery, operative, 20. time factors, 21. ventilation-perfusion ratio
   Yes, 1, not 19, 20
**I^3R**

- Designed to structure a search session based on interactions with a “expert intermediary”
- Inspired by Belkin’s work and research on multiple search strategies and representations

Croft and Thompson, 1987
I^3R: A new approach to the design of document retrieval systems
I³R Interface
Information Interaction in Context

Exploratory Search

• Supporting complex search processes beyond “one-shot” retrieval

Marchionini, 2006. Exploratory search: From finding to understanding
Web Search and SERPs

- “Ten blue links”
- Query log analysis
- Query suggestion
- Query transformation
- Search aggregation

Providing diverse sources of information to the user

Papers by Dumais, Teevan, White on user behavior, including “sessions”
Example: Web search
Evaluation

- Expected Search Length and RF measures
- TREC interactive track
- TREC session track
- NDCG and variations
- User behavior models and simulations
- User studies and crowdsourcing

Difficult to evaluate system actions beyond ranking and user actions beyond clicking
Questions and Answers

1. Question answering
2. Recommendation systems
3. Answer retrieval
4. Voice-based search
5. Forums and CQA
6. Mobile search
7. Conversational systems


Allan et al., 2012. SWIRL: Conversational Answer Retrieval.

Surdeanu, Ciaramita, Zaragoza, 2008. Learning to rank answers on large online QA collections.
QA and Interaction

• Longer questions give more context for answers
  • but were thought to require too much user effort
• Answers are more precise than “relevance”
  • different models for evaluation and feedback
  • better basis for modeling interaction?
• SERPs and diversity
  • not appropriate for answers?
  • snippets vs. answers
• CQA data reflects human-to-human, mostly single-turn, interaction
  with potentially complex information needs
• Forum data reflects multi-turn, multi-party, conversational interaction
Bandwidth and Interaction

• Mobile devices and voice-based systems limit the bandwidth for interaction
  • mostly on output

• SERPs no longer possible

• Question-answer paradigm more concise and potentially more accurate

• QA interaction requires more active role by system
  • Selecting responses, asking clarifying questions, obtaining feedback about wrong answers
  • Multi-turn “conversational” retrieval
Conversational Answer Retrieval
(from SWIRL 2012)

• Open-domain, natural language text questions

• Dialogue would be initiated by the searcher and proactively by the system

• Dialogue is about questions and answers, including history, with the aim of refining the understanding of questions and improving the quality of answers

• Answers extracted from the corpus (or corpora) being searched, and may be at different levels of granularity, depending on the question

• Evaluated as an open-domain IR task, in contrast to conversational chat or template-based conversation
Research Challenges for CAR

• Tasks
  • Breaking down the research required into manageable pieces

• Test Collections
  • Creating test collections that capture aspects of conversational retrieval for training and testing

• Evaluation
  • Creating (or agreeing on) measures that can be used for evaluating multi-turn, conversational interactions directed at addressing information needs
Tasks

• Retrieving similar questions
• Retrieving good answers
• Predicting next questions
• Response retrieval
• Hybrid generation/retrieval of responses
• Choosing clarifying questions
• Conversational recommendation
• Conversational question answering
• Modeling intent in search conversations
• Intent-based response retrieval
• Intent-based generation/retrieval

Increasing system modeling of history and context of the search dialog (cf. Belkin and I^3R)
Answer Retrieval

- **Paper:** Yang, Ai, Guo, and Croft. 2016. aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model.

- **Test Collection:** TREC QA, Yahoo CQA

- **Evaluation:** MAP, MRR

- **Model:**

![Diagram](image-url)
Response Retrieval

- **Paper**: Yang, Qiu, Qu, Guo, Zhang, Croft, Huang, and Chen. 2018. Response Ranking with Deep Matching Networks and External Knowledge in Information-seeking Conversation Systems.
- **Test Collection**: UDC, MSDialog, AliMe
- **Evaluation**: MAP, Recall@1, 2, 5
- **Model**: [Diagram showing the model architecture]
Response Retrieval

**Response Retrieval**

**QA Dialog Title:** Windows Update Failure
Dialog Tags: Windows, Windows 10, Windows update, recovery, backup, PC

**USER:** I have Windows10, version 1511, OS Build 10586.116. For the past year I have tried to upgrade from this without success. Upgrade download OK but on installing only get to 85 - 93% and then on restart install previous version of windows (the 1511 version), I have Windows update assistant installed. Any help or advice on this would be most welcome.

David

**Responses**

**AGENT James (Microsoft MVP - Windows Client):**
Response: There's not a doubt in my mind that those Norton "leftovers" is your troubleshooter here - but now that the Norton Removal Tool has been deprecated and especially since the new-fangled Norton Remove and Reinstall tool doesn't get rid of the leftovers, a manual upgrade or a clean install of Microsoft Win10 appears to be your only possible resolution here. Feel free to give Norton/Symantec a piece of your mind!

Term Match: Magenta  Semantic Match: Blue  Correspondence Match: Red

<table>
<thead>
<tr>
<th>Context</th>
<th>Method</th>
<th>y^2</th>
<th>Top-1 Ranked Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>[User] I open Excel and it automatically formats my dates into American formatting. I have changed and saved the formatting to NZ style. However everyday I pull the document out of office 365 it reverts back to the American format. How do I stop this? [Agent] Is it one file or all files in Excel? [User] It does seem to be all Excel files. How do I change the global date format setting?</td>
<td>SMN</td>
<td>0</td>
<td>Go to Settings -&gt; System -&gt; Tablet Mode..., Change setting as indicated in the snapshot below.</td>
</tr>
<tr>
<td></td>
<td>DMN-KD</td>
<td>1</td>
<td>That is a Windows setting. Go to Control Panel &gt; Regional settings. This will change date settings for all applications.</td>
</tr>
<tr>
<td></td>
<td>DMN-PRF</td>
<td>1</td>
<td>That is a Windows setting. Go to Control Panel &gt; Regional settings. This will change date settings for all applications.</td>
</tr>
</tbody>
</table>
Hybrid Response Generation/Retrieval

- **Paper:** Song, Li, Nie, Zhang, Zhao, and Yan. 2018. An Ensemble of Retrieval-Based and Generation-Based Human-Computer Conversation Systems.

- **Test Collection:** Wiebo, Tieba, Twitter/Foursquare (Ghazvininejad et al, A Knowledge-Grounded Neural Conversation Model. In AAAI ’18)

- **Evaluation:** Bleu, Rouge-L, human

- **Model:**

  ![Diagram of Hybrid Response Generation/Retrieval System]

<table>
<thead>
<tr>
<th>Category</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval</td>
<td>literal human utterances; various expressions with great diversity;</td>
<td>not tailored to queries; bottleneck is the size of repository</td>
</tr>
<tr>
<td>Generation</td>
<td>tailored for queries; highly coherent</td>
<td>insufficient information; universal sentences</td>
</tr>
</tbody>
</table>
# Hybrid Response Generation/Retrieval

<table>
<thead>
<tr>
<th>Query</th>
<th>Utterance (Translated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieved-1</td>
<td>This mobile phone’s photo effect is pretty good.</td>
</tr>
<tr>
<td>Retrieved-2</td>
<td>I really have a crush on it.</td>
</tr>
<tr>
<td>multi-seq2seq</td>
<td>Go for it.</td>
</tr>
<tr>
<td>seq2seq</td>
<td>Rushing for it rather than having a crush on it.</td>
</tr>
<tr>
<td>Ha-ha.</td>
<td></td>
</tr>
<tr>
<td>Query</td>
<td>Can I see the house tomorrow afternoon?</td>
</tr>
<tr>
<td>Retrieved-1</td>
<td>You can call me!</td>
</tr>
<tr>
<td>Retrieved-2</td>
<td>You can see the house on weekends.</td>
</tr>
<tr>
<td>multi-seq2seq</td>
<td>You can see the house on weekends, please call me in advance</td>
</tr>
<tr>
<td>seq2seq</td>
<td>OK.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context</th>
<th>System Output Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donated to the [ENTITY] last night and now I have to listen to</td>
<td></td>
</tr>
<tr>
<td>automated phone calls. It’s enough to make me want to cancel.</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td></td>
</tr>
<tr>
<td>r/g</td>
<td></td>
</tr>
<tr>
<td>Ground Truth</td>
<td></td>
</tr>
<tr>
<td>r/g</td>
<td></td>
</tr>
<tr>
<td>I didn’t get it. I didn’t.</td>
<td></td>
</tr>
<tr>
<td>Seq2Seq</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
<tr>
<td>I’m sorry to hear that. Please dm us your email address so we can look</td>
<td></td>
</tr>
<tr>
<td>into this. Thanks!</td>
<td></td>
</tr>
<tr>
<td>Seq2Seq-Facts</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
<tr>
<td>It’s a known issue in LA county. I just got an email from my dm and</td>
<td></td>
</tr>
<tr>
<td>tech and it should be good in 30 minutes or so.</td>
<td></td>
</tr>
<tr>
<td>Retrieval</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>We’re listening and would like to know more and help with your experience.</td>
<td></td>
</tr>
<tr>
<td>Please follow us so I can dm you our contact info. [ENTITY]</td>
<td></td>
</tr>
<tr>
<td>HNCM-RS</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>We’re sorry to hear this. Please dm us if you need assistance.</td>
<td></td>
</tr>
<tr>
<td>HNCM-RS</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
<tr>
<td>We’re sorry to hear this. Please dm us your contact info so we can look</td>
<td></td>
</tr>
<tr>
<td>into this.</td>
<td></td>
</tr>
</tbody>
</table>
Choosing Clarifying Questions

- **Paper:** Aliannejadi, Zamani, Crestani, and Croft, 2019. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations.
- **Test Collection:** Qulac (TREC Web track, crowdsourcing)
- **Evaluation:** MRR, P@1, nDCG@1, 5, 20
- **Model:**

\[
score = \gamma(\phi_T(t), \phi_H(h), \phi_Q(q), \eta(t, h, q), \sigma(t, h, q))
\]
Conversational Recommendation

- **Paper:** Zhang, Xu, Yang, Ai, and Croft, 2018. Towards Conversational Search and Recommendation: System Ask, User Respond
- **Test Collection:** Amazon product dataset
- **Evaluation:** MAP, MRR, nDCG
- **Model:**

![Diagram of conversational recommendation system with modules for user, search, and question representation.]
Conversational Recommendation

Can you find me a mobile phone on Amazon?
Sure, what operating system do you prefer?
I want an Android one.
OK, and any preference on screen size?
Better larger than 5 inches.
Do you have requirements on storage capacity?
I want it to be at least 64 Gigabytes.
And any preference on phone color?
Not particularly.
Sure, then what about the following choices?

I don't like them very much...
OK, do you have any preference on the brand?
Better be Samsung or Huawei.
Any requirement on price?
Should be within 700 dollars.
OK, then what about these ones?

Great, I want the first one, can you order it for me?
Sure, I have placed the order for you, enjoy!
Conversational Question Answering

- **Paper:** Qu, Yang, Qiu, Croft, Zhang, and Iyer, 2019. BERT with History Answer Embedding for Conversational Question Answering.
- **Test Collection:** QuAC dataset
- **Evaluation:** F1, HEQ-Q, HEQ-D
- **Model:**

![Diagram](image)
Conversational Question Answering

<table>
<thead>
<tr>
<th>#</th>
<th>ID</th>
<th>R</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Q₁</td>
<td>U</td>
<td>What is relevant about Lorrie’s musical career?</td>
</tr>
<tr>
<td></td>
<td>A₁</td>
<td>A</td>
<td>... her first album on that label, Leave the Light On, was released in 1989.</td>
</tr>
<tr>
<td>2</td>
<td>Q₂</td>
<td>U</td>
<td>What songs are included in the album?</td>
</tr>
<tr>
<td></td>
<td>A₂</td>
<td>A</td>
<td>CANNOTANSWER</td>
</tr>
<tr>
<td>3</td>
<td>Q₃</td>
<td>U</td>
<td>Are there any other interesting aspects about this article?</td>
</tr>
<tr>
<td></td>
<td>A₃</td>
<td>A</td>
<td>made her first appearance on the Grand Ole Opry at age 13,</td>
</tr>
<tr>
<td>4</td>
<td>Q₄</td>
<td>U</td>
<td>What did she do after her first appearance?</td>
</tr>
<tr>
<td></td>
<td>A₄</td>
<td>A</td>
<td>... she took over his band at age 16 and began leading the group ...</td>
</tr>
<tr>
<td>5</td>
<td>Q₅</td>
<td>U</td>
<td>What important work did she do with the band?</td>
</tr>
<tr>
<td></td>
<td>A₅</td>
<td>A</td>
<td>leading the group through various club gigs.</td>
</tr>
<tr>
<td>6</td>
<td>Q₆</td>
<td>U</td>
<td>What songs did she played with the group?</td>
</tr>
<tr>
<td></td>
<td>A₆</td>
<td>A</td>
<td>CANNOTANSWER</td>
</tr>
<tr>
<td>7</td>
<td>Q₇</td>
<td>U</td>
<td>What are other interesting aspects of her musical career?</td>
</tr>
<tr>
<td></td>
<td>A₆</td>
<td>A</td>
<td>To be predicted ...</td>
</tr>
</tbody>
</table>
Modeling Intent in Search Interactions


- **Test Collection:** MSDialog, UDC

- **Evaluation:** Accuracy, precision, recall, F1

- **Model:**

![Diagram showing model structure with layers labeled as Dense Layer, Conv Layer, and Embedding Layer. Features include Q, NF, PF, FD, GG, More Conv, Pool, Dropout, and U1, U2, U3, U4, U5.]}
## User Intent Taxonomy

<table>
<thead>
<tr>
<th>Code</th>
<th>Label</th>
<th>Description</th>
<th>Example</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>QQ</td>
<td>Original Question</td>
<td>The first question by a user that initiates the QA dialog.</td>
<td>If a computer is purchased with win 10 can it be downgraded to win 7?</td>
<td>13</td>
</tr>
<tr>
<td>RQ</td>
<td>Repeat Question</td>
<td>Posters other than the user repeat a previous question.</td>
<td>I am experiencing the same problem ...</td>
<td>3</td>
</tr>
<tr>
<td>CQ</td>
<td>Clarifying Question</td>
<td>Users or agents ask for clarification to get more details.</td>
<td>Your advice is not detailed enough. I'm not sure what you mean by ...</td>
<td>4</td>
</tr>
<tr>
<td>FD</td>
<td>Further Details</td>
<td>Users or agents provide more details.</td>
<td>Hi. Sorry for taking so long to reply. The information you need is ...</td>
<td>14</td>
</tr>
<tr>
<td>FQ</td>
<td>Follow Up Question</td>
<td>Users ask follow up questions about relevant issues.</td>
<td>Thanks. I really have one simple question -- if I ...</td>
<td>5</td>
</tr>
<tr>
<td>IR</td>
<td>Information Request</td>
<td>Agents ask for information of users.</td>
<td>What is the make and model of the computer? Have you tried installing ...</td>
<td>6</td>
</tr>
<tr>
<td>PA</td>
<td>Potential Answer</td>
<td>A potential answer or solution provided by agents.</td>
<td>Hi. To change your PIN in Windows 10, you may follow the steps below: ...</td>
<td>22</td>
</tr>
<tr>
<td>PF</td>
<td>Positive Feedback</td>
<td>Users provide positive feedback for working solutions.</td>
<td>Hi. That was exactly the right fix. All set now. Tx!</td>
<td>6</td>
</tr>
<tr>
<td>NF</td>
<td>Negative Feedback</td>
<td>Users provide negative feedback for useless solutions.</td>
<td>Thank you for your help, but the steps below did not resolve the problem ...</td>
<td>4</td>
</tr>
<tr>
<td>GG</td>
<td>Greetings/Gratitude</td>
<td>Users or agents greet each others or express gratitude.</td>
<td>Thank you all for your responses to my question ...</td>
<td>22</td>
</tr>
<tr>
<td>JK</td>
<td>Junk</td>
<td>There is no useful information in the post.</td>
<td>Emojis. Sigh .... Thread closed by moderator ...</td>
<td>1</td>
</tr>
<tr>
<td>O</td>
<td>Others</td>
<td>Posts that cannot be categorized using other classes.</td>
<td>N/A</td>
<td>1</td>
</tr>
</tbody>
</table>
Intent-Aware Response Retrieval

- Paper: ...
- Test Collection: UDC, MSDialog
- Evaluation: MAP, $R_n@k$
- Model:
# Intent-Aware Response Retrieval

<table>
<thead>
<tr>
<th>ID</th>
<th>Role</th>
<th>Utterances</th>
<th>Intent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utterance-1</td>
<td>User</td>
<td>Windows downloaded this update &quot;2018-02 Cumulative Update for Windows 10 .....&quot; But during the restart it says &quot;we couldn’t complete the update, undooing changes&quot;. So what can I do to stop this? Thanks</td>
<td>OQ</td>
</tr>
<tr>
<td>Utterance-2</td>
<td>Agent</td>
<td>Is there any other pending updates? Try Download Troubleshooter for Win 10.</td>
<td>IR/ PA</td>
</tr>
<tr>
<td>Utterance-3</td>
<td>User</td>
<td>Yes, pending updates the same one. I already used the built in troubleshooter, it did fix some 3 issues, but doing a restart the problem persists. Can I stop updates from installing this particular one? Thanks.</td>
<td>PA/ FQ</td>
</tr>
<tr>
<td>Utterance-4</td>
<td>User</td>
<td>Not sure if related but I just saw that Malicious Software Removal of March did not install .....</td>
<td>FD</td>
</tr>
<tr>
<td>Response-1 (Correct)</td>
<td>Agent</td>
<td>Try run troubleshooter and then restart your PC. If problem persist, open start and search for Feedback and open Feedback Hub app and report this issue.</td>
<td>PA</td>
</tr>
<tr>
<td>Response-2 (Wrong)</td>
<td>Agent</td>
<td>Glad to know that you fixed the issue, and as I said downloading the &quot;Show or hide updates&quot; troubleshooter and restarting the PC will help you. Thank you for asking questions and providing feedback here!</td>
<td>GG</td>
</tr>
</tbody>
</table>
What Next?

• Intent-aware hybrid generation and retrieval of responses
• Incorporating NLP comprehension and inference models
• Studying explicit vs. implicit dialog models for search
• User studies of answer interaction and visualization...
• Developing better evaluation methodologies for interactive conversational retrieval
• Developing large test collections...
• Developing other modes of interaction
Summary

• Both user- and system-oriented IR researchers have recognized the importance of interaction
• Search systems are increasingly modeling and participating actively in the search process
• Conversational answer retrieval is driving progress in this direction, and much remains to be done
• As we move to multi-modal (and multi-party) interactive search, modeling the search context (history, goals, etc.) and the dialog will be the basis for effective user-system collaborations
SIGIR 2022?
THANK YOU
Answer Interaction

QA Test Collections

- TREC QA: 1.5K factoid questions with 60K paired potential answer sentences
- Yahoo L6 Webscope: 4.5M questions and associated answer passages from CQA service (Manner Questions subset: 150K “how” questions)
- WikiQA: 3K factoid questions with 30K answer sentences from associated Wiki page
- MS MARCO: 1M factoid questions from Bing log with 9M “companion” passages and 180K manually generated answers
- SQUAD: 100K manually generated questions with associated answers that are text spans in 530 Wikipedia articles
- WebAP: 8K text span answer passages (av. 45 words) from relevant documents for 80 TREC Gov2 questions
- Yahoo nfL6 subset: 85K non-factoid question and answer pairs
- WikiPassageQA: 4K non-factoid queries and answer passages created from 860 Wikipedia pages
- ANTIQUE: 2.5K questions from nfL6 with more complete relevance judgments
Conversation Test Collections

- Ubuntu (UDC): 1M conversations from technical support chat logs
- QuAC: 14K crowdsourced QA dialogs based on Wikipedia articles
- MSDialog: 35K conversations from MS technical support forum, 2K labelled with utterance intent
- AliMe: 63K context-response pairs from commercial online help chatbot (Chinese)
- Qulac: 10K crowdsourced clarifying question-answer pairs related to 200 TREC topics (see talk at conference)
- Amazon: Simulated product purchase conversations based on product facets
- MSMARCO Conversational Search: 45M user sessions containing 340K unique queries
- TREC CAST: New TREC track building on MSMARCO, others