Pilot Task: Session Search (SS)

Yiqun Liu*, Jia Chen†, Fan Zhang†, Jiaxin Mao‡, Weizhi Ma†
† Department of Computer Science & Tech., Tsinghua University
‡ Gaoling School of AI, Renmin University of China

yiqunliu@tsinghua.edu.cn, chenjia0831@gmail.com
Motivation

- Multi-query sessions/tasks are common in search process nowadays.
  - Considering the contextual information within sessions has been proved efficient for user intent modeling in IR communities.
  - Existing tasks in NTCIR have not involved session-based retrieval yet.

- Existing relevant tasks
  - TREC Session Track 2011-2014: small-scale datasets, simulated search tasks, only evaluate the last query;

==> NCTIR-16 Session Search Task
Methodology

• To better assess the search effectiveness at both query-level and session-level, we aim to set up two subtasks as follows:

  **Fully Observed Session Search (FOSS):**
  
  ➢ For a k-length session, we provide full session contexts in the first \((k - 1)\) queries.
  
  ➢ NDCG, AP, RBP

  **Partially Observed Session Search (POSS):**
  
  ➢ In this subtask, we truncate all sessions before the last query. For a session with \(k\) queries \((k \geq 2)\), we only reserve the session contexts in the first \(n\) queries, where \(1 \leq n \leq k - 1\).
  
  ➢ RS-DCG, RS-RBP
Methodology

To better assess the search effectiveness at both query- and session-level, we aim to set up two subtasks as follows:

- **Fully Observed Session Search (FOSS):** For a k-length session, we provide full session contexts in the first (k−1) queries. Participants need to rank the candidate documents for the last query of a session. This setting follows TREC Session Tracks to enable ad-hoc evaluation by using metrics such as NDCG, AP, and RBP, etc.

- **Partially Observed Session Search (POSS):** In this subtask, we truncate all session contexts before the last query. For a session with k queries (k ≥ 2), we only reserve the session contexts in the first n queries, where 1 ≤ n ≤ k−1. The value of n varies in different sessions. Participants will need to rank documents for the last k−n queries according to the partially observed contextual information in previous search rounds. Session-level metrics such as RS-DCG and RS-RBP will be adopted for the evaluation of system effectiveness.

The differences between our Session Search task and previous related tasks are listed in the following table.

### Table 1. Differences between our Session Search task and previous related tasks

<table>
<thead>
<tr>
<th></th>
<th><strong>Ours (NTCIR SS task)</strong></th>
<th><strong>TREC Session Tracks</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Sessions</strong></td>
<td>• Chinese: 147,155 (Training) + 1,124 (Validating) + 2,356 (Testing)</td>
<td>• English: 76~1,257</td>
</tr>
<tr>
<td><strong>Session Datasets</strong></td>
<td>• TianGong-ST</td>
<td>• Session Track 2011-2014</td>
</tr>
<tr>
<td></td>
<td>• TianGong-SS-FSD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• An un-released field study dataset</td>
<td></td>
</tr>
<tr>
<td><strong>Document Collection</strong></td>
<td>• A collection provided by TianGong-ST, with more than 297,597 web pages.</td>
<td>• ClueWeb09/ClueWeb12</td>
</tr>
<tr>
<td></td>
<td>• A collection which contains top 50 candidate documents for all queries in the validating and testing set.</td>
<td></td>
</tr>
<tr>
<td><strong>Source/Generation of session data</strong></td>
<td>• Refined from a search log from Sogou.com</td>
<td>• Generated by real search users based on manually designed topics.</td>
</tr>
<tr>
<td></td>
<td>• Extracted from two large-scale field studies.</td>
<td></td>
</tr>
<tr>
<td><strong>Support from log analysis for annotation?</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Support for session-level evaluation?</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Dataset & Resources

- **Training set:**
  - Participants can adopt any training techniques, e.g., adversarial learning, multi-task learning.

- **Validating set:**
  - TianGong-SS-FSD ([http://www.thuir.cn/tiangong-ss-fsd/](http://www.thuir.cn/tiangong-ss-fsd/)), A Chinese-centric dataset collected via a field study. It collected daily search logs as well as explicit feedback from 30 participants for one month.

- **Testing set:** not released yet.
Organizers

Yiqun Liu [yiqunliu@tsinghua.edu.cn] (Tsinghua University)
Jia Chen [chenjia0831@gmail.com] (Tsinghua University)
Fan Zhang [franky94@gmail.com] (Tsinghua University)
Jiaxin Mao [maojiaxin@gmail.com] (Renmin University of China)
Weizhi Ma [mawz@tsinghua.edu.cn] (Tsinghua University)
Schedule

- July 1, 2021: Session Search registration due
- July 15, 2021: Dataset Release
- Sep 1, 2021 - Nov 30, 2021: Formal Run
- Dec, 2021 - Jan, 2022: Relevance Assessment
- Feb 1, 2022: Evaluation Result Release
- Feb 1, 2022: Draft Task Overview Paper Release
- Mar 1, 2022: Draft Participant Paper Submission Due
- May 1, 2022: All Camera-ready Paper Submission due
- Jun 2022: NTCIR-16 Conference & EVIA 2022 in NII, Tokyo, Japan
References


Thank you!