NTCIR-13 OpenLiveQ Task
Open Live Test for Question Retrieval

Makoto P. Kato, Takehiro Yamamoto (Kyoto University),
Sumio Fujita, Akiomi Nishida (Yahoo Japan Corporation)
Goal

Improve the **REAL** performance of question retrieval systems in a **production** environment

Performance evaluated by **REAL** users

Yahoo! Chiebukuro (a CQA service of Yahoo! Japan)
Task

- Given a query, return a ranked list of questions that can satisfy many REAL users in Yahoo! Chiebukuro (a CQA service)

**Effective for Fever**

**Three things you should not do in fever**

While you can easily handle most fevers at home, you should call 911 immediately if you also have severe dehydration with blue .... Do not blow your nose too hard, as the pressure can give you an earache on top of the cold. ....

10 Answers  Posted on Jun 10, 2016

**Effective methods for fever**

Apply the mixture under the sole of each foot, wrap each foot with plastic, and keep on for the night. Olive oil and garlic are both wonderful home remedies for fever. 10) For a high fever, soak 25 raisins in half a cup of water.

2 Answers  Posted on Jan 3, 2010
OpenLiveQ provides an OPEN LIVE TEST EVIRONMENT

Real users

Ranked lists of questions from participants’ systems are **INTERLEAVED**, presented to real users, and evaluated by their clicks.
• Multileaved comparison methods are used in the online evaluation

• Therefore, submitted runs are evaluated by REAL users’ clicks
Data (tentative)

Plan to provide **valuable data**, which have not been available for universities

- **1,000** training queries
- Top **1,000** search results for each query
- **Clickthrough data**
  - How many users have clicked on each result
  - What types of users have clicked (age, sex, job, etc.)
Challenges

How to satisfy diverse relevance criteria of real users?
While topical relevance has been considered in many IR researches, the other types of relevance criteria (e.g. popularity and freshness) have not.

How to answer ambiguous/underspecified queries?
In community Q&A, there may be more diverse intents behind the same query.

“cheap office”

Q. How to rent a cheap office? OR Q. How to buy cheap MS office?
**Possible Research Topics**

- **Search intent detection**
  - e.g. Estimate important relevance criteria (e.g. popularity) and rank questions by them (e.g. number of user views)

- **User demographic estimation**
  - e.g. Estimate major demographics of the query (e.g. F1) and estimate the relevance mainly for such user groups

- **Eye catching title detection**
  - e.g. Witty titles can be good for “search for fun”
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 2016</td>
<td>Offline training data release</td>
<td>Data will be released for training retrieval systems</td>
</tr>
<tr>
<td>Dec 2016</td>
<td>Registration due</td>
<td>Registration form will be available at the NTCIR-13 site</td>
</tr>
<tr>
<td>Jan-Mar 2017</td>
<td>Offline test run submission</td>
<td>Systems will be evaluated without real users</td>
</tr>
<tr>
<td>Apr-Jun 2017</td>
<td>Online test</td>
<td>Systems will be evaluated with real users</td>
</tr>
<tr>
<td>Jul 2017</td>
<td>Online test result release</td>
<td>Evaluation results will be released to participants</td>
</tr>
</tbody>
</table>

Looking forward to your participation. Thanks!