# 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 <span style="color:crimson">can satisfy many REAL users in Yahoo! Chiebukuro (a CQA service)</span>**

**Effective for Fever** **← INPUT** Q&A 🔍 検索

## 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**
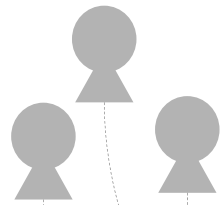
**← OUTPUT**

## 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
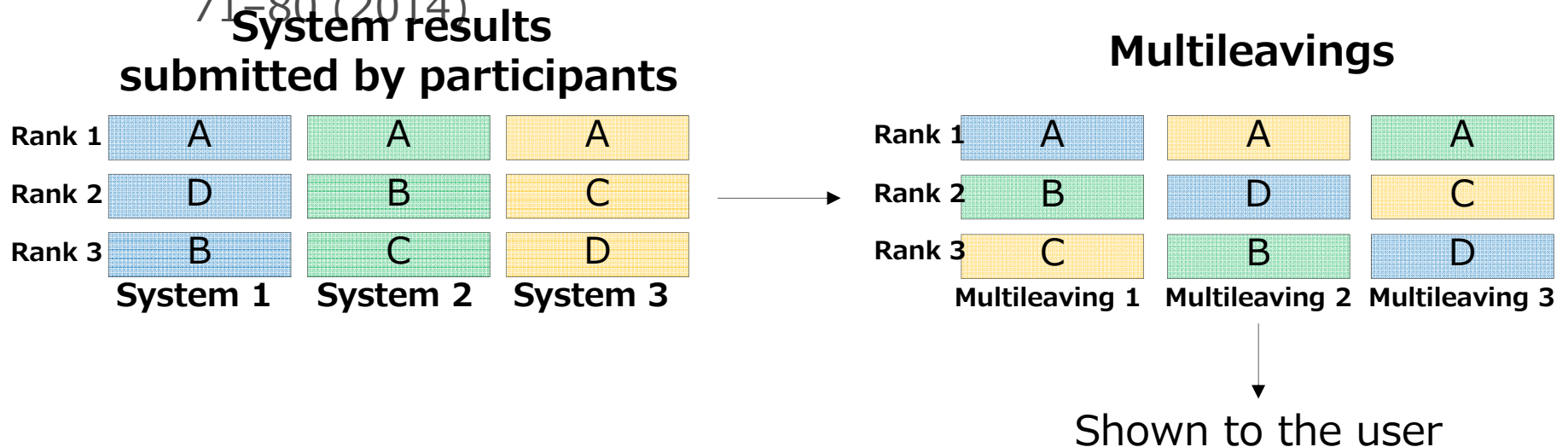
**Ranked lists of questions from participants' systems are INTERLEAVED, presented to real users, and evaluated by their clicks**

# Multileaving-based Evaluation

- ## Multileaved comparison methods are used in the online evaluation

  – Schuth, A., Sietsma, F., Whiteson, S., Lefortier, D., de Rijke, M.: Multileaved comparisons for fast online evaluation. In: CIKM. pp. 71-80 (2014)

**System results submitted by participants**

**Multileavings**

| | System 1 | System 2 | System 3 |
|---|---|---|---|
| Rank 1 | A | A | A |
| Rank 2 | D | B | C |
| Rank 3 | B | C | D |

| | Multileaving 1 | Multileaving 2 | Multileaving 3 |
|---|---|---|---|
| Rank 1 | A | A | A |
| Rank 2 | B | D | C |
| Rank 3 | C | B | D |

Shown to the user

- ## 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
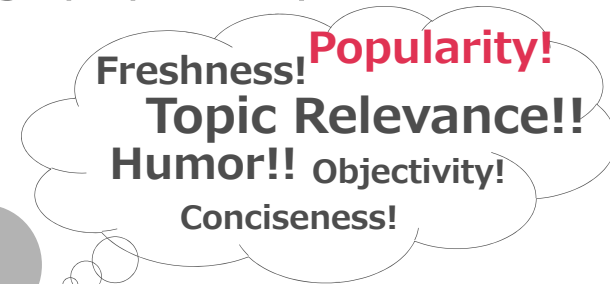  - What types of users have clicked (age, sex, job, etc.)

## 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

**What are major spots in Kyoto?**

Freshness! **Popularity!**
**Topic Relevance!!**
**Humor!!** Objectivity!
Conciseness!

## 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

| Effective for Fever | ✕ | Q&A ▾ | 🔍 検索 |

- **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"

## Schedule

| Date | Event | Details |
|------|-------|---------|
| Oct 2016 | **Offline training data release** | Data will be released for training retrieval systems |
| Dec 2016 | **Registration due** | Registration form will be available at the NTCIR-13 site |
| Jan-Mar 2017 | **Offline test run submission** | Systems will be evaluated without real users |
| Apr-Jun 2017 | **Online test** | Systems will be evaluated with real users |
| Jul 2017 | **Online test result release** | Evaluation results will be released to participants |

**Looking forward to your participation. Thanks!**