Cheng Luo  
Tsinghua University

Tetsuya Sakai  
Waseda University

Yiqun Liu  
Tsinghua University

Zhicheng Dou  
Renmin University of China

Jingfang Xu  
Sogou Inc.
Why “We Want Web”?

• IR tasks have diversified, and good old-fashioned ad hoc IR tasks have disappeared.
• Ad hoc IR may have reached a plateau in the ‘90s, but we need a stable evaluation environment for breakthroughs!
• In particular, we want to design and improve web search algorithms that are not just black boxes.
Task design

• An ad hoc web search task, which we plan to continue for at least three rounds (NTCIR 13-15).
• Some user behavior data will be provided to improve search quality.
• Evaluation measures: traditional measures like ERR, nDCG, Q-measure plus novel measures (e.g. U-measure which takes into account snippet and document lengths, ones that take into account the informational/navigational distinction)
• English and Chinese subtasks for NTCIR-13 (+Japanese for NTCIR-14)
• Collective failure analysis at a pre-NTCIR-13 workshop
We Want Web@NTCIR-13 (1)

NTCIR-13 (Dec 2017)

- Frozen topic set
- NTCIR-13 fresh topic set
- New runs pooled for frozen + fresh topics
- NTCIR-13 systems
We Want Web@NTCIR-13 (2)

NTCIR-13 (Dec 2017)

- Frozen topic set
- NTCIR-13 fresh topic set
- NTCIR-13 systems

Qrels + std. factors based on NTCIR-13 systems
- NOT released

Qrels + std. factors based on NTCIR-13 systems
- released

Official NTCIR-13 results discussed with the fresh topics
We Want Web@NTCIR-14 (2)

Using the NTCIR-14 fresh topics, compare new NTCIR-14 runs with revived runs and quantify progress.

NTCIR-13 (Dec 2017)

- Frozen topic set
- NTCIR-13 fresh topic set
- NTCIR-13 systems

NTCIR-14 (Jun 2019)

- Frozen topic set
- NTCIR-14 fresh topic set
- NTCIR-14 systems

Qrels + std. factors based on NTCIR-13+14 systems NOT released

Qrels + std. factors based on NTCIR-(13+)14 systems released

Official NTCIR-14 results discussed with the fresh topics
We Want Web@NTCIR-15 (2)

Using the NTCIR-15 fresh topics, compare new NTCIR-15 runs with revived runs and quantify progress.


- frozen topic set
- fresh topic set
- NTCIR-13 fresh topic set
- NTCIR-14 fresh topic set
- NTCIR-15 fresh topic set

Qrels + std. factors based on NTCIR-(13+14+)15 systems released

Official NTCIR-15 results discussed with the fresh topics

NTCIR-13 systems

NTCIR-14 systems

NTCIR-15 systems
We Want Web@NTCIR-15 (3)

How do the standardisation factors for each frozen topic differ across the 3 rounds?

NTCIR-13 (Dec 2017)
- Frozen topic set
- NTCIR-13 fresh topic set
- Qrels + std. factors based on NTCIR-13 systems released
- NTCIR-13 systems

NTCIR-14 (Jun 2019)
- Frozen topic set
- NTCIR-14 fresh topic set
- Qrels + std. factors based on NTCIR-13+14 systems released
- NTCIR-14 systems

NTCIR-15 (Dec 2020)
- Frozen topic set
- NTCIR-15 fresh topic set
- Qrels + std. factors based on NTCIR-13+14+15 systems released
- NTCIR-15 systems

Official NTCIR-15 results discussed with the fresh topics

Qrels + std. factors based on NTCIR-(13+14+)15 systems released
We Want Web@NTCIR-15 (4)

How do the NTCIR-15 system rankings differ across the 3 rounds, with and w/o standardisation?

NTCIR-15 systems ranking

Qrels + std. factors based on NTCIR-13+14+15 systems released

Official NTCIR-15 results discussed with the fresh topics
Highlights of We Want Web@NTCIR

• **Easy** to participate – it’s an ad hoc web task!
• **New** Sogou-Q/Sogou-T (2016) datasets provided to participants!
• **Long-term** monitoring and quantification of improvements
• **User-centric** evaluation explored by leveraging user behaviour data
• **Collective** effort on failure analysis, not just mean performances
Tentative timeline

Jul-Aug 2016 Corpora released to registered participants

Aug-Sep 2016 Designing and constructing topics

Oct 2016-Jan 2017 User behavior data collected for the topics

Feb-Mar 2017 User behavior data released to registered participants

Apr 2017 Task registrations due

May 2017 Topics released; runs received

Jun-Jul 2017 Relevance assessments

Aug 2017 Results released to participants

Sep 2017 Draft participants papers due

Nov 2017 Pre-NTCIR WWW Workshop on Failure Analysis in Beijing

Dec 2017 NTCIR-13 Conference
Organisers

• Cheng Luo, Yiqun Liu (Tsinghua University)
• Tetsuya Sakai (Waseda University)
• Zhicheng Dou (Renmin University of China)
• Jingfang Xu (Sogou, Inc.)

Questions?
ntcirwww@list.waseda.jp
http://www.thuir.cn/ntcirwww