Call for Task Participation

NTCIR-12 Pilot Task:
Short Text Conversation (STC)

Lifeng Shang, Zhengdong Lu, Hang Li (Huawei Noah’s Ark Lab, Hong Kong)
Tetsuya Sakai (Waseda University, Japan)
Twitter: @ntcirstc

February 27, 2015@NTCIR-12 Kickoff
Microblogs: Twitter, Weibo...
What is STC? (1)

POST: “Dr. Hang Li’s Learning to Rank for IR and NLP second edition released! Follow!”
What is STC? (2)

POST: “Dr. Hang Li’s Learning to Rank for IR and NLP second edition released! Follow!”

COMMENT by Hang Li: “Thanks ZhiYuan! I’ve added detailed explanations of the LambdaMART algorithms etc. “

Coherent AND useful
What is STC? (3)

**POST:** “How’s the hair?”
What is STC? (4)

**POST:** “How’s the hair?”

**COMMENT by Tetsuya Sakai:**
“I don’t have any.”

*Coherent but NOT useful*
Objectives

• The ultimate objective
  Build an open-domain system that can interact naturally with humans

• The objective for NTCIR-12/13
  Build an IR system that effectively reuses past comments to respond to a post.

Coherence: the post-comment pair makes sense as a consecutive short text exchange between two people.
Usefulness: the comment contains information or an opinion that might be useful to the author of the post.
Given a new post, can a coherent and useful comment be returned by searching a post-comment repository? What are the challenges and limitations of this IR-based STC approach? [Ji14]
Data and language scope

We also provide English machine translations of the Chinese posts and comments.

Table 1: Statistics of the dataset

<table>
<thead>
<tr>
<th></th>
<th>Retrieval Repository</th>
<th>Labeled Data</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>#posts</td>
<td>39,279</td>
<td>225</td>
<td>50 (or 100)</td>
</tr>
<tr>
<td>#comments</td>
<td>1,136,998</td>
<td>6,017</td>
<td></td>
</tr>
<tr>
<td>#original pairs</td>
<td>1,136,998</td>
<td>6,017</td>
<td></td>
</tr>
<tr>
<td>#posts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#comments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#labeled pairs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of test posts will be determined using topic set size design [Sakai14CIKM, Sakai14EVIA]
Task design and evaluation measures

• Ad hoc IR design: given a “new” post, retrieve coherent and useful comments from repository.
• Pooling and graded relevance assessments
  L2: coherent and useful
  L1: coherent but not useful
  L0: not coherent (and therefore not useful either)
• Evaluation measures (basically one good comment is enough):
  G@1 (normalised gain at rank 1)
  ERR (expected reciprocal rank)
  P+ (Similar to Q-measure, suitable for navigational intents)
  [Sakai14PROMISE]
Plans for STC-2@NTCIR-13

• Follow the INTENT-2 “revived run” model [Sakai13INTENT]
• STC-1 participants will keep their systems in the fridge
• When they come back at STC-2, they use both STC-1 and STC-2 systems to handle the STC-2 posts
• Compare STC-1 and STC-2 systems on the STC-2 test collection
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 27, 2015</td>
<td>NTCIR-12 kickoff</td>
</tr>
<tr>
<td>Oct 31, 2015</td>
<td>NTCIR-12 task registration deadline</td>
</tr>
<tr>
<td>Nov 2, 2015</td>
<td>STC test topics released</td>
</tr>
<tr>
<td>Nov 30, 2015</td>
<td>STC run submission deadline</td>
</tr>
<tr>
<td>Dec 2015-Jan 2016</td>
<td>STC relevance assessments + evaluation</td>
</tr>
<tr>
<td>Feb 1, 2015</td>
<td>STC results sent to participants + STC draft overview released</td>
</tr>
<tr>
<td>Mar 1, 2015</td>
<td>NTCIR-12 participants’ draft papers due / Task organisers’ feedback</td>
</tr>
<tr>
<td>May 1, 2015</td>
<td>NTCIR-12 all camera ready papers due</td>
</tr>
<tr>
<td>Jun 7-10, 2015</td>
<td>NTCIR-12 conference</td>
</tr>
</tbody>
</table>

We will give you training data as soon as you register! Sooner the better!
Twitter: @ntcirstc
Prospective participants and budget

- Prof. James Kwok’s group at Hong Kong University of Science and Technology
- Prof. Xiaoyan Zhu’s group at Tsinghua University
- Prof. Jun Zhao’s group at Institute of Automation, Chinese Academy of Sciences
- Prof. Bin Wang’s group at Institute of Information Engineering, Chinese Academy of Sciences
- Prof. Qingchao Chen’s group at Harbin Institute of Technology Shenzhen
- Prof. Xuanjing Huang’s group at Fudan University
- Prof. Jussi Karlsgren’s group at Gavagai

Huawei will cover the relevance assessment cost. No seeding funding from NTCIR required.
Related tasks

• **TREC Microblog** (2011-) [Lin13]
  Data: twitter, NOT distributed to participants
  Tweets2011: only IDs distributed, data downloaded individually
  Tweets2013: Evaluation as a Service (access through APIs)
  Ad hoc search etc. Evaluation based on binary relevance

• **NTCIR Community Question Answering** (2010) [Ishikawa10]
  Data: Japanese Yahoo! Answers (Chiebukuro)
  Given a Q and its responses, rank the responses (which is the best answer?). Evaluation using G@1 etc.
References


