

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)

<http://ntcir12.noahlab.com.hk/stc.htm>

Twitter: @ntcirstc

February 27, 2015@NTCIR-12 Kickoff

Microblogs: Twitter, Weibo...

Over 40 million users

超过4000万人正在使用

李航博士

华为技术有限公司 诺亚方舟实验室 首席科学家

+ 关注 私信 三

他的主页 他的相册

370 关注 41999 粉丝 1171 微博

微博认证 Lv.22

华为技术有限公司 诺亚方舟实验室 首席科学家

热门 #WSDM2015#会议上，腾讯的报告Semantic Matching in APP Search (by Juchao Zhuo)，百度的报告Boosting Search with Deep Understanding of Contents and Users (by Kaihua Zhu) 讲的都是语义匹配，其实是搜索的核心问题。幻灯片：
网页链接
2月7日 20:02 来自 微博 weibo.com

收藏 转发 181 评论 10 18

最近

- 2015
- 2014
- 2013
- 2012
- 2011

第一条微博

What is STC? (1)

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

@刘知远THU V 🏠

@李航博士 老师的专著《Learning to Rank for Information Retrieval and Natural Language Processing》出第二版了，关注。 <http://t.cn/R7ahdDU>



10月27日 15:18 来自 微博 weibo.com

转发 200 | 评论 18

What is STC? (2)

谢谢知远！新增了LambdaMART等算法的详细介绍。

@刘知远THU  

@李航博士 老师的专著《Learning to Rank for Information Retrieval and Natural Language Processing》出第二版了，关注。<http://t.cn/R7ahdDU>



10月27日 15:18 来自 微博 weibo.com

转发 200 | 评论 18

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?"

@苍井空 V 🏆

头发怎么样？！



2013-7-1 15:44 来自 iPhone客户端

What is STC? (4)

POST: "How's the hair?"

COMMENT by Tetsuya Sakai:
"I don't have any."

Coherent but NOT useful

完全没有了

@苍井空 V 🏠

头发怎么样？！



2013-7-1 15:44 来自 iPhone客户端

2013-7-1 19:13 来自 微博 weibo.com

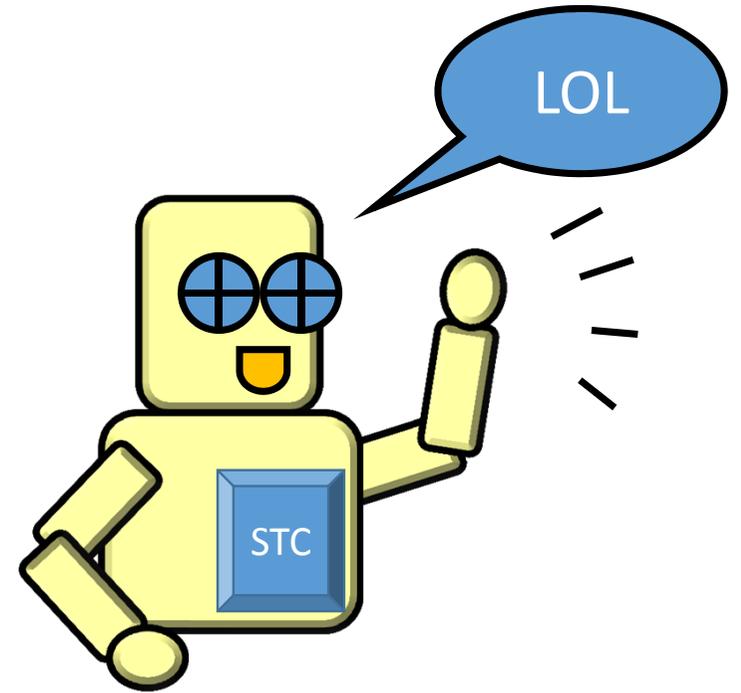
Objectives

- The ultimate objective

Build an open-domain system that can interact naturally with humans

- The objective for NTCIR-12/13

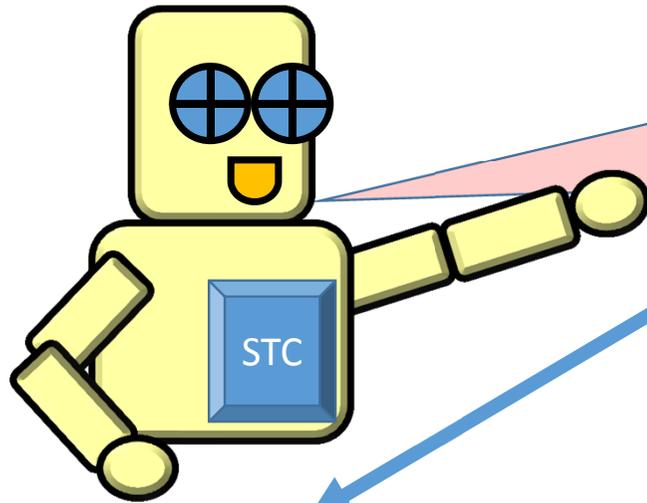
Build an IR system that **effectively reuses past comments** to respond to a post.



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.

STC research questions



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]

post

Search and reuse

post-comment repository

post

comment
comment

post

comment
comment

post

comment
comment

post

comment
comment

Data and language scope

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

Table 1: Statistics of the dataset

Retrieval Repository	#posts	39,279
	#comments	1,136,998
	#original pairs	1,136,998
Labeled Data	#posts	225
	#comments	6,017
	#labeled pairs	6,017
Test Data	#query posts	50 (or 100)

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)

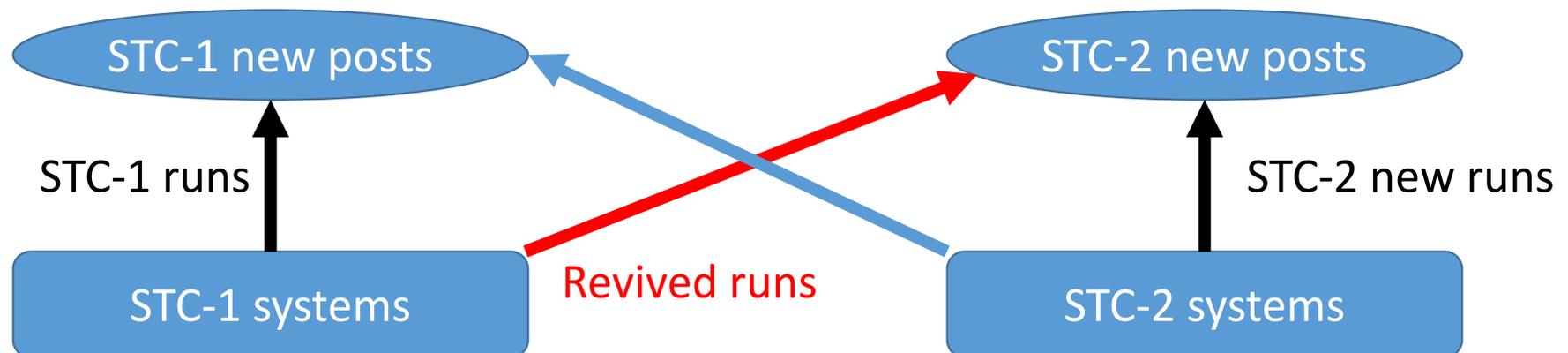
ERR (expected reciprocal rank)

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



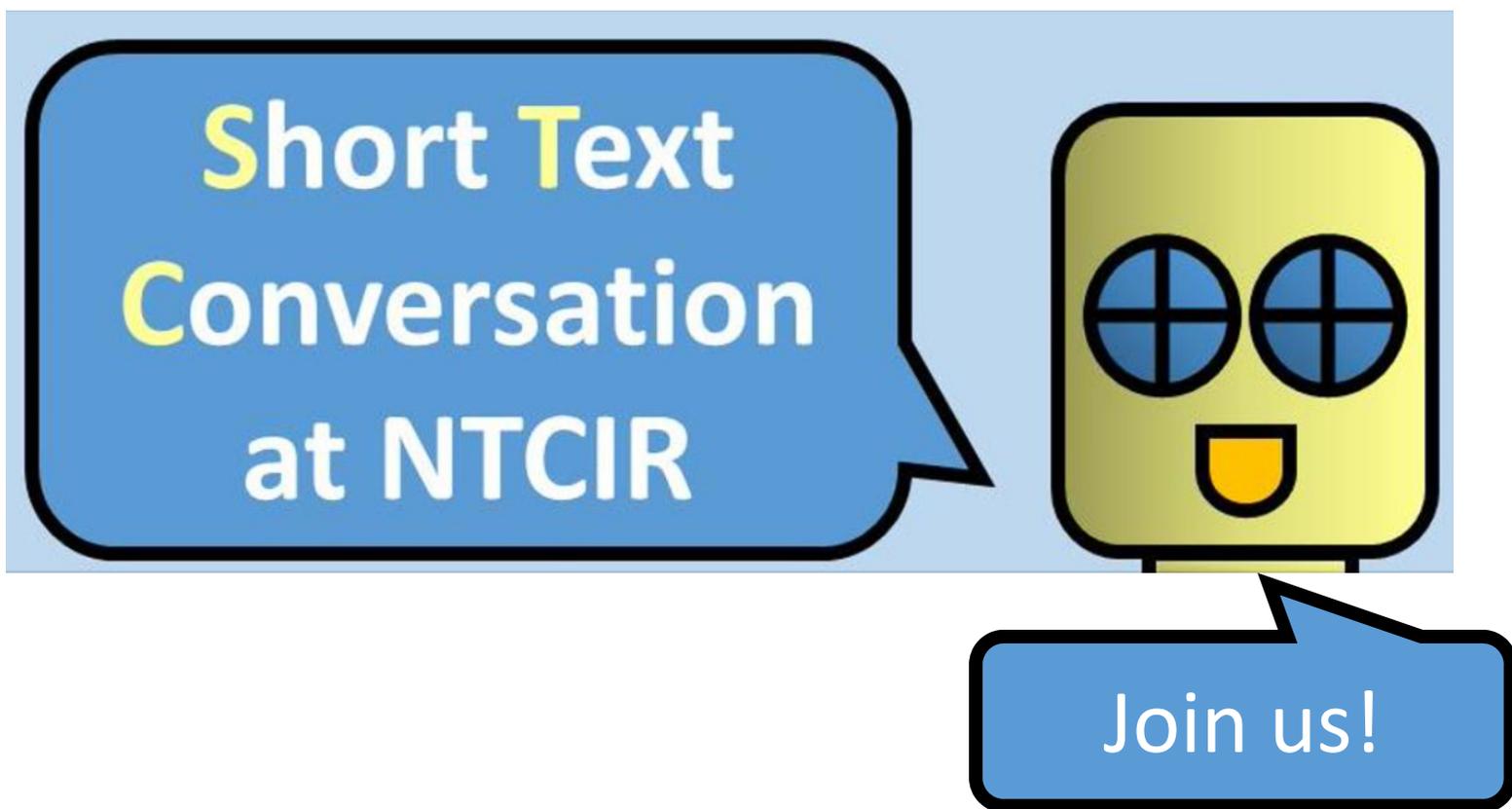
Schedule

We will give you training data as soon as you register!
Sooner the better!

Feb 27, 2015	NTCIR-12 kickoff
Oct 31, 2015	NTCIR-12 task registration deadline
Nov 2, 2015	STC test topics released
Nov 30, 2015	STC run submission deadline
Dec 2015-Jan 2016	STC relevance assessments + evaluation
Feb 1, 2015	STC results sent to participants + STC draft overview released
Mar 1, 2015	NTCIR-12 participants' draft papers due / Task organisers' feedback
May 1, 2015	NTCIR-12 all camera ready papers due
Jun 7-10, 2015	NTCIR-12 conference

<http://ntcir12.noahlab.com.hk/stc.htm>

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. Qingchai Chen's group at Harbin Institute of Technology Shenzhen
- Prof. Xuanjing Huang's group at Fudan University
- Prof. Jussi Karlgren'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

[Ishikawa10] Ishikawa, D., Sakai, T. and Kando, N.: Overview of the NTCIR-8 Community QA Pilot Task (Part I): The Test Collection and the Task, Proceedings of NTCIR-8, pp.421-432, 2010.

[Ji14] Ji, Z., Lu, Z. and Li, H.: An Information Retrieval Approach to Short Text Conversation, 2014.
<http://arxiv.org/abs/1408.6988>

[Lin13] Lin, J and Efron, M.: Overview of the TREC-2013 Microblog Track, Proceedings of TREC 2013, 2013.

[Sakai13INTENT] Sakai, T. et al.: Overview of the NTCIR-10 INTENT-2 Task, Proceedings of NTCIR-10, pp.94-123, 2013.

[Sakai14CIKM] Sakai, T.: Designing Test Collections for Comparing Many Systems, Proceedings of ACM CIKM 2014, pp.61-70, 2014.

[Sakai14EVIA] Sakai, T.: Topic Set Size Design with Variance Estimates from Two-Way ANOVA, Proceedings of EVIA 2014, pp.1-8, 2014.

<http://www.f.waseda.jp/tetsuya/CIKM2014/ir0030-sakai.pdf>

[Sakai14PROMISE] Sakai, T.: Metrics, Statistics, Tests, PROMISE Winter School 2013: Bridging between Information Retrieval and Databases (LNCS 8173), 2014.

<http://research.microsoft.com/en-us/people/tesakai/metrics.pdf>