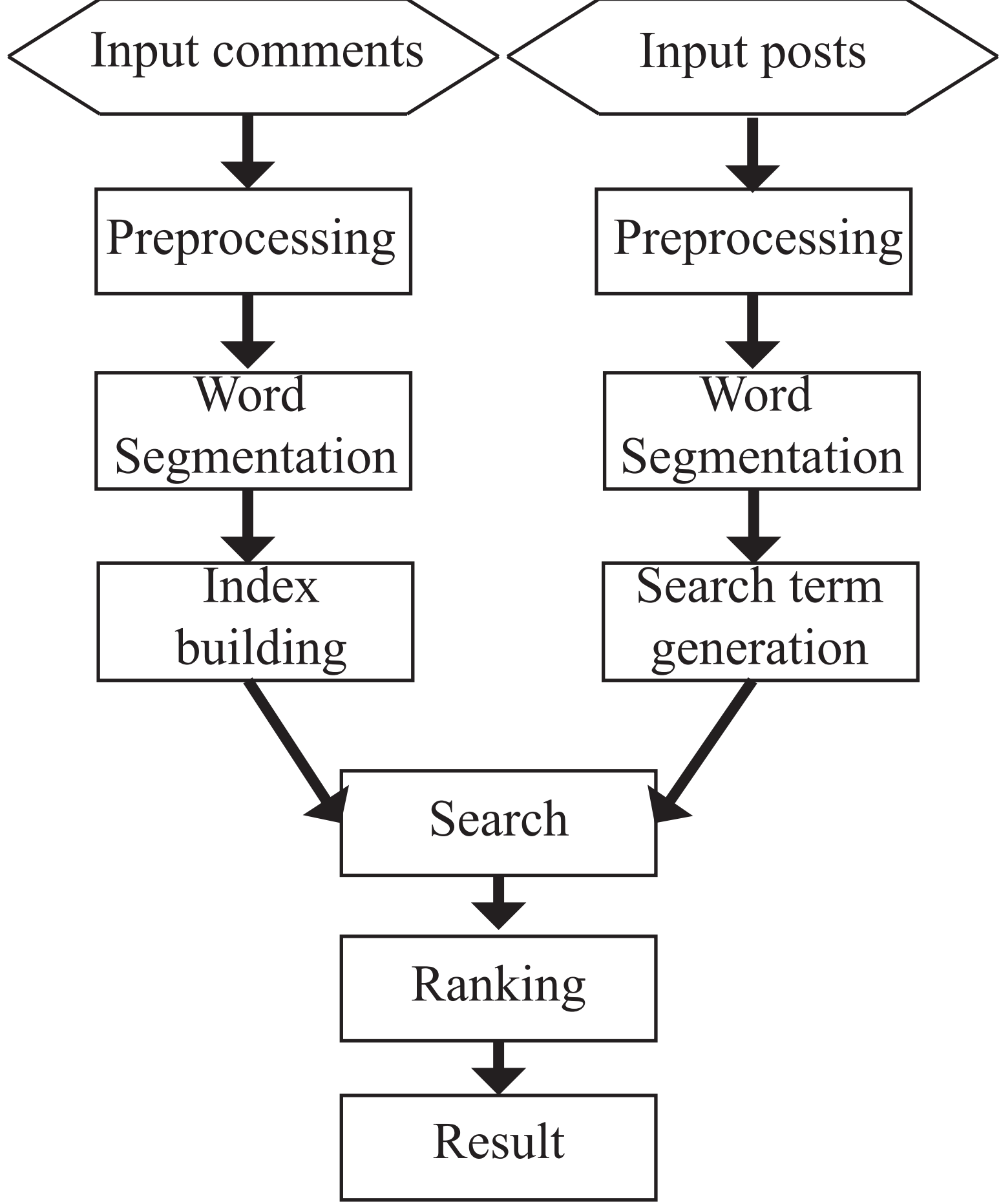# CYUT Short Text Conversation System for NTCIR-12 STC

Shih-Hung Wu, Wen-Feng Shih, Liang-Pu Chen, Ping-Che Yang

Dept of CSIE, Chaoyang University of Technology, Taichung Taiwan(R.O.C.)

Institute for Information Industry, Taipei, Taiwan(R.O.C.)

Abstract

In this paper, we report how we build the system for Chinese subtask in NTCIR12 Short Text Conversation (STC) shared task. Our approach is to find the most related sentences for a given input sentence. The system is implemented based on the Lucene search engine. The result shows that our system can deal with the conversation that involves related sentences.

System Flow

Our main approach to the task is to extract suitable search terms for each post and try to find comments with the search terms in the repository as the candidate responses. Our system then ranks the candidates according to the level or relevance and returns the top ten sentences as the system result for one input post.

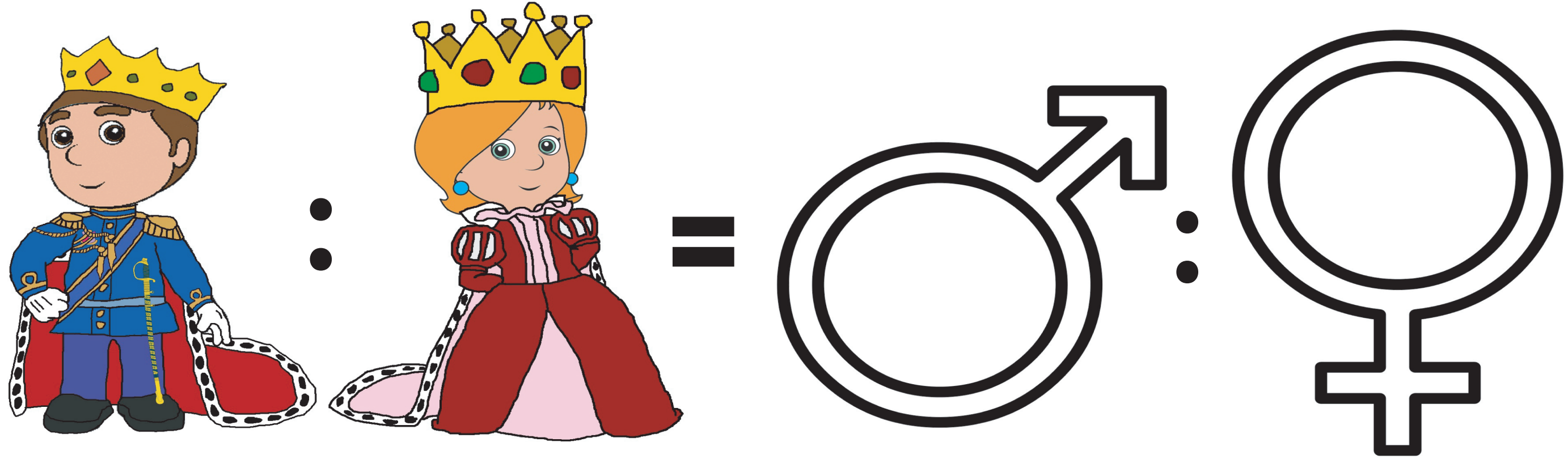| Post	Response Comment
train-post-10010 昨天晚上程序掛掉了。 (Last night, my process died.)	repos-cmmt-1025251760 親身經歷過的人默默無語路過。。。。。(As a person with the same experience, I cannot say anything.)

Formal run

Run	Mean nDCG@1	Mean P+	Mean nERR@10
cyut-C-R1	0.2233	0.3851	0.3608

Analogy Approach

A post A and a response B in the training set can be coupled to find the best response D for a test post C by the analogy A:B=C:D

Post A	train-post-10040 到家了，感觉好温馨啊 (Arriving home, what a warm feeling!)
Response Comment B	repos-cmmt-1035394800 不错，温馨，家的感觉 (Right, worm, the feeling of home)
Post C	test-post-10010 远离城市，才得安宁。 (Far from a city, get peace.)
Response Comment D	repos-cmmt-1000993080 感觉不错 (Feeling good)

Relationship between posts and responses

By observing the training set, we find that there some communication types between the posts and their responses. Here we list 7 major types:

1. Sympathy
   The response shows feelings of pity and sorrow for the misfortune in the post.
2. Confirmation
   The response is confirming the opinion or information in the post.
3. Reply information
   Whey post is a query for some kind of information, the response should be a reply to the question.
4. Sarcasm
   Some response is laughing or deny the idea in the post.
5. Nonsense
   Nonsense some time is a kind of humor; however, literally the response is nonsense.
6. Ask back
   Some responses are questions that ask back the same thing in the posts.
7. Envy
   Some response shows the feeling of envy to the post.

Sympathy		Sarcasm	
C	test-post-10690 手机被偷了怎么办呢？有什么方式还能找回来吗？ (My cell phone is stolen, what can I do? Is there any way that it can be found?)	C	test-post-10460 上帝削弱我的自制力，一定是为了突出另外一种能力。是什么？ (God weakened my ability on self-control, it must be for reinforce my another ability, what can it be?)
D	repos-cmmt-1015174060 感觉很无奈 (Feeling helpless)	D	repos-cmmt-1003212790 油太多了 (Too much oil)
Confirmation		**Nonsense**	
C	test-post-10010 远离城市，才得安宁。 (Far from a city, get peace.)	C	test-post-10130 终于看到蓝天啦！和坏天气saybye！ (See the blue sky finally! Say bye to the bad weather!)
D	repos-cmmt-1000993080 感觉不错 (Feeling good)	D	repos-cmmt-1005597660 可不想成功的人就永远不会成功 (The one who don't want success will never get success.)
Reply information		**Ask back**	
C	test-post-10250 老规矩，周末不谈PPT。有什么乐子可以大家谈谈？ (As old rules, don't talk about PPT on weekend. Anything fun that we can talk about?)	C	test-post-10050 一大早上，各种打哈欠的，有木有 (Early morning, yawning in every way, isn't it)
D	repos-cmmt-1022638290 去网吧通宵 (Stay overnight at Internet café.)	D	repos-cmmt-1000295940 是吗？？？ (Really???)
		Envy	
		C	test-post-10340 重庆小雨，三五人围坐火锅！香吃无语。 (Chongqing rain, three to five people sitting around the pot! Enjoy food.)
		D	repos-cmmt-1000967570 好想吃啊 (Really want to eat.)

Conclusion

The type of posts are also need to explore. There are many posts involved in social events and entertainments. The conversation is not just question answering, it also contains discussions on subjects in many cases. A proper response in these cases should be a confirmation on the opinion or give a counter question.