WUST System at NTCIR-12 Short Text Conversation Task

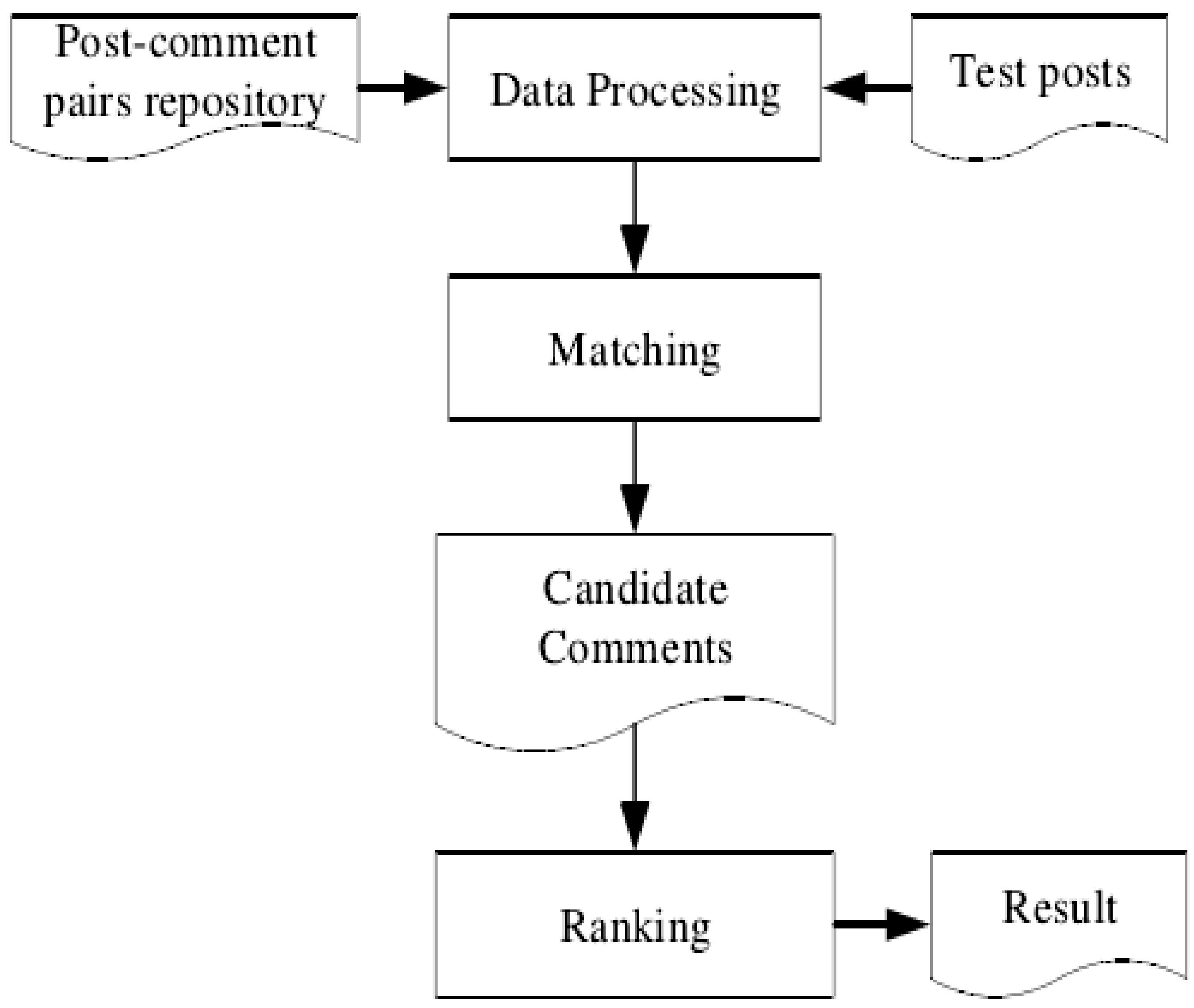
Maofu Liu, Yifan Guo, Yang Wu, Limin Wang
College of Computer Science and Technology, Wuhan University of Science and Technology,
liumaofu@wust.edu.cn, 498966594@qq.com,
wuyang0329@foxmail.com, smile_wlm@163.com

Han Ren
College of Computer Science,
Hubei University of Technology,
Wuhan 430068, P.R. China
hanren@whu.edu.cn

Introduction

- In NTCIR-12, the STC task is a much simplified version of the human-computer conversation: One round of conversation formed by two short texts, with the former being an initial post from users and the latter being a comment given by the computer .
- The STC task consists of two subtasks, i.e. one is Chinese and the other is Japanese. We only concerned with the Chinese subtask in this paper.
- The STC is defined as an IR problem to search for the appropriate comments matching the given query q, derived from the given post, from the dataset.
- In this paper, we adopt the framework to extract topic-words from the given query q, to retrieval candidate responses, and then match and rank the responses to produce the final ranked list.

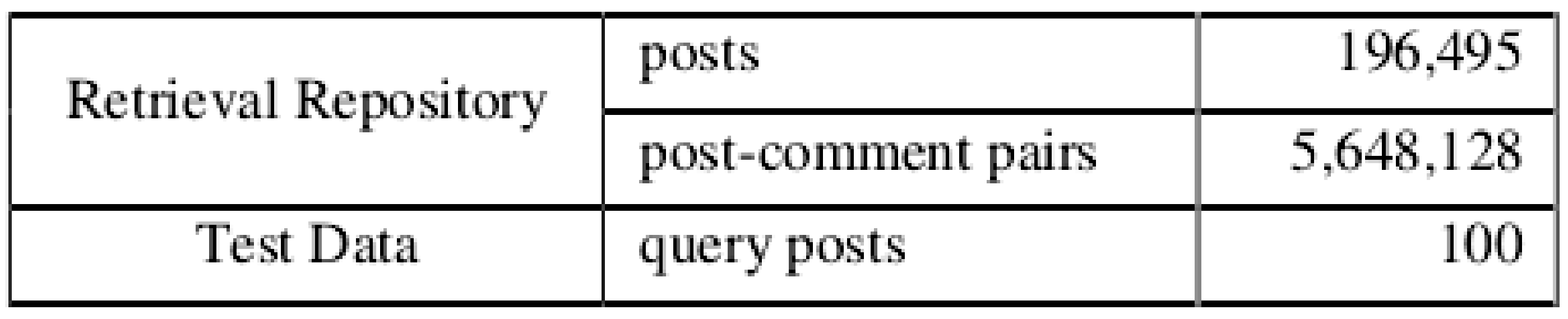
System Architecture



1. Data Preprocessing

- First, the system inputs the queries derived from the test posts, and then segments the Chinese words and removes the stop words. Here we use ICTCLAS for Chinese word segmentation.
- Second, the system processes the post-comment pair repository and builds the inverted index table of the posts and words separately.
- Finally, the model produces the topic words set T, and retrieves a number of candidate posts by the set T for each query.

Table1. Statistics of the dataset for Chinese STC task



Retrieval Repository	posts	196,495
	post-comment pairs	5,648,128
Test Data	query posts	100

.

2. Matching

- The main work of this stage is to find the posts being similar to the query q and use their comments as the candidates.
- Here we use a simple VSM for measuring the similarity between the query q and the candidate post p. The score is more close to 1 if the two texts are more similar.
- We choose the top-N most similar posts and produce the set $P_q^{reduced}$. The system further picks post-comment pairs

3. Ranking

- First, the system measures the similarity between q and each of comments in post-comment pairs.
- Then, the system uses a linear ranking function for further evaluate and returns the comments with top-10 comments to the given post.

Experiments

- The official evaluation results are listed in Table 2.

Table2. Formal run experiment official results

Run	WUST_C_R1
Mean nDCG@1	0.0567
Mean nDCG@1	0.1218
Mean nERR@10	0.0980

- In Chinese STC task, our system only depends on the VSM and the given topic words by ICTCLAS, which does not care the global context. As a result, our system fails to produce the desired result.
- We matched only 38 test posts so the average score is very low. One reason is that our model aims to match the test post and the post-comment pairs when they have common words for measuring query-posts and query-comments similarities. When the candidate comments and the query have no similarity in surface , the system does not work.

Conclusions

- In this paper, we have described our model based on VSM for STC task in Chinese. We also analyzed our result submitted and then adjusted parameters which outperformed the former. recognition approach based on linguistic phenomena.
- We feel that there are two important ways to improve the efficiency of our model for STC task. We need to enhance the accuracy by combining with other models, such as the topic-words, and to consider matching between query and response in terms of semantic relevance, speech act, and entity association.
- In the future, besides on the basis of information-retrieval, we also would like to generate the appropriate and human-like response derived from what we searched from the post-comment pair repository.