

Submitting Runs to STC@NTCIR-12 [Japanese subtask]

(This instruction is modified based on the run submission instruction of INTENT-2@NTCIR-10)

(Further modified for the Japanese subtask)

Run Names

Run files should be named as follows. Make your <teamID> is exactly what you registered.

<teamID>- J-R[priority].txt

e.g., Huawei-J-R1.txt (J means Japanese; 1 means this run has the highest priority to be included in the pool for relevance assessment)

Number of Runs

IMPORTANT! A participating team can submit up to 5 runs.

NOTE: Please don't do this: you have only one algorithm, but you submit two runs by trying two different parameter settings: e.g. param=0.8 for Run 1, param=0.2 for Run 2. THIS IS LIKE TUNING WITH THE TEST DATA! (Exceptions: if, for example, param=0.0 means switching off a certain component of your system, then yes you might want to try both param=0.0 and param=0.8 etc.)

Tweet Ranking Run Submission Format

For those of you who are familiar with the TREC run format, our Tweet Ranking run format is basically the same. The only difference is that we require a one-sentence description of the run in line 1 of each runfile. Line 1 of each runfile must be of the form:

<SYSDESC>[insert a short description in English here]</SYSDESC>

The rest of the runfile should contain lines of the form:

[TweetID] 0 [RetrievedTweetID] [Rank] [Score] [RunName]\n

Use a white space as the field separator. Here TweetID is the ID (id_str) of an input tweet and RetrievedTweetID is the ID (id_str) of a tweet in the repository set. Note that the second field "0" is a dummy field. For example, a Japanese tweet ranking run named Huawei-J-R1 should look like this:

```
613587908235112448 0 496282184613761024 1 27.73 Huawei-J-R1
613587908235112448 0 496279274530152448 2 25.15 Huawei-J-R1
613587908235112448 0 480913653370077184 3 21.89 Huawei-J-R1
```

IMPORTANT! Return no more than 10 tweets per input tweet.

NOTE: The list of TweetIDs are provided as a text file, each line containing one TweetID. e.g.,

```
613587908235112448
573067096029978624
566605228036747264
...
```

GUIDELINE FOR TWEET RETRIEVAL

- Tweets are deleted on a daily basis. In order to make formal-run a fair one for those who have recently joined the task, we decided not to include deleted tweets in the evaluation set.
So, please do not retrieve tweets that do not exist as of Feb. 4th 2016. The list of tweets that have been deleted will be made available.
- For retrieving tweets, participants can use any information except for the replies to the TweetIDs in the test data by searching Twitter. This is like using the reference data.

Checking Your Runs

Each run file should be checked for formatting errors using the [CheckIntent](#) script. We will not evaluate badly formatted run files.

First, try

```
% CheckIntent-STC.pl <runfile>
```

and if a problem is detected, execute

```
% CheckIntent-STC.pl <runfile> <fixedrunfile>
```

to fix the problems. Then remember to check and rename your runfiles, e.g.

```
% diff <fixedrunfile> <runfile>
```

```
% mv <fixedrunfile> <runfile>
```

Uploading Your Runs

1. Create a single file for your team called "<teamID>-STC.tar.gz". This file should contain ALL of your run files, PLUS a plain text file called "<teamID>-STC.txt" which contains a complete list of your runs, one run file name per line. Thus if you are submitting five runs, the text file should contain exactly five lines.
2. Upload the gzipped tar file to our run submission site. The URL will be announced to registered participants only.