FIRE 2017 IRLeD track run submission instructions

Instructions common for both the tasks:

* Runs are to be submitted as plain text attachments to a mail to fire2017irled@gmail.com
* If a team is participating in both the tasks, a separate mail for each task should be sent.
* The **Subject** of each mail containing a run should be:

<div align="center">

FIRE 2017 IRLeD run submission for Task 1

(or)

FIRE 2017 IRLeD run submission for Task 2

</div>

* Each team should choose a teamid (a string without any whitespace) for itself. This teamid will be used to report the performance of the team.
Some example teamids are: iiest_flt17, iitkgp_fire_legal_track_2017, etc.

* Each team can submit **up to three (03) runs for each task**, which will be useful if you wish to compare the performance of different strategies.

* Each run is to be submitted as a separate text file attached in your mail. Each run should also have **a distinct runid**, which will be used to report the performance of each individual run.
It is advised that a runid should include the teamid. For instance, if the team with teamid "iiest_flt17" is submitting three runs, the runids can be: iiest_flt17_1, iiest_flt17_2, iiest_flt17_3.

* Along with the runs, each team should submit a write-up describing their methodology in detail for all runs. The write-up can be in any format, either in plain text or pdf.

* We will consider three types of runs, based on the amount of manual intervention in different stages such as query formation and retrieval:
  1. Fully automatic where no step involves manual intervention
  2. Semiautomatic where there is some manual intervention, but manual intervention is limited to the query formation stage
  3. Manual where there is manual intervention in both the query formation and the document retrieval stages

Please indicate in the write-up the type of each submitted run (automatic / semiautomatic / manual). This is to ensure fair comparison among different types of runs. The same team can submit different types of runs, if they wish to.

Task Specific Instructions:

Task 1:

* It is expected that for each test document, your strategy will give a ranked list of **catchphrases**, along with a relevance / importance score for each catchphrase with respect to the document. The submitted run should reflect this ranking, so that we can evaluate the top K results for each document (e.g., K = 100). **For each document, you should report all catchphrases which your strategy identifies, in a ranked order starting with the most relevant / important one.**

* An individual file containing a run should be formatted as follows:
Each line should have:
1. The chosen runid followed by **||** (two consecutive pipe characters)
2. The document name for which the catchwords are being extracted, followed by **||** (two consecutive pipe characters)
3. A comma-separated list of <catchphrase:score> pairs, where a catchphrase and its score are separated by a colon. Note that the catchphrases can contain spaces within them.

Some sample lines in the above format are:

```
iiest_flt16_1||case_100_statement||first catchphrase:0.658,second catchphrase:0.369
iiest_flt16_1||case_101_statement||first catch:0.858,second catch:0.789,third
catch:0.789
```

Please ensure that the above format is followed for all submitted runs, otherwise our evaluation scripts cannot be used. Also, ensure that all catchphrases judged relevant to a particular document appear in ranked order, with the most relevant / important catchphrase at the beginning. Also note that a run file should contain catchphrases relevant to all the test documents for which you wish to submit results.

Task 2:

For Task 2 we are following an evaluation procedure similar to what is used by TREC, including using the standard TREC evaluation scripts *(available at http://trec.nist.gov/trec_eval/)*.

* It is expected that for each current case, your strategy will give a ranked list of prior cases, along with a relevance score for each prior case with respect to the current case. The submitted run should reflect this ranking, so that we can evaluate the top K results for each current case (e.g., K = 100).

Note that, for each current case, you should report all prior cases which your strategy finds relevant, in a ranked order. We will check the ranks of the prior cases that were actually cited from the current case, in your ranked list.

* An individual file containing a run should be formatted according to the traditional TREC format, which is as follows.
Each line should have six (06) fields separated by one space character. The fields are:
1. a current case id, e.g., current_case_0001, current_case_0002, ..., current_case_0200
2. the specific string "Q0"
3. a prior case id, which is judged relevant to the current case id in the first field
4. a relevance based rank for the prior case id, corresponding to the current case id (the rank can start from 0 or 1)
5. a relevance score for the prior case id, corresponding to the current case id
6. the runid of this run.

Some sample lines in the above format are:
```
current_case_0001 Q0 prior_case_0059 1 0.626000 iiest_flt16_1
current_case_0001 Q0 prior_case_0023 2 0.366700 iiest_flt16_1
```

Ensure that the above format is followed for all submitted runs, otherwise the TREC evaluation scripts cannot be used. Also, ensure that all prior cases judged relevant to a particular current case appear in ranked order, with the most relevant prior case at the top. Also note that a run file should contain prior cases judged relevant to all current cases for which you wish to submit results.

If you have any further queries, feel free to mail to fire2017irled@gmail.com