

Dataset & Evaluation Plan

Data:

Task 1 (Catchphrase Extraction):

The data will contain:

- (i) 100 case documents and their corresponding gold standard catchphrases, for training.
- (ii) Another 300 documents as test data. For each of the document in the test data, the participants have to find the catchphrases.

Task 2 (Precedence Retrieval):

Two sets of documents will be provided:

1. Query_docs -- Current cases, formed by removing the links to the prior cases
2. Object_docs -- The prior cases which have been cited by the cases in Query_docs (links to which are removed from the Query_docs) along with some random cases (not among Query_docs)

There will be 200 documents in Query_docs and more than 2000 documents in Object_docs. About half of the documents in Object_docs are cited by at least one of the cases in Query_docs. The other half of the cases in Object_docs were randomly chosen to deliberately make the task challenging. For each of the 200 documents in Query_docs, the task will be to rank the list of 2000+ documents in Object_docs (or a subset), so that the actually cited prior cases are ranked higher than the other documents.

To get the dataset for this track, please send the scanned version of the duly filled-in "organizational-access form" found [here](#) to the email address



FIRE-2017-IRLeD

Track Description

Dataset & Evaluation Plan

Important Dates

Cc