☰ **FIRE-2017-IRLeD**   **Track Description**   Dataset & Evaluation Plan   Important Dates   A

Information Retrieval from Legal Documents (IRLeD)

Track description:

In a Common Law System, great importance is given to prior cases. A prior case (also called a *precedent*) is an older court case related to the current case, which discusses similar issue(s) and which can be used as reference in the current case. A prior case is treated as important as any law written in the law book (called *statutes*). This is to ensure that a similar situation is treated similarly in every case. If an ongoing case has any related/relevant legal issue(s) that has already been decided, then the court is expected to follow the interpretations made in the prior case. For this purpose, it is critical for legal practitioners to find and study previous court cases, so as to examine how the ongoing issues were interpreted in the older cases.

With the recent developments in information technology, the number of digitally available legal documents has rapidly increased. It is, hence, imperative for legal practitioners to have an automatic precedent retrieval system. The task of *precedence retrieval* can be modelled as a task of information retrieval, where the current document (or a description of the current situation) will be used as the query, and the system should return relevant prior cases as results.

Generally, legal texts (e.g., court case descriptions) are long and have complex structures. This makes their thorough reading time-consuming and strenuous. So, apart from a precedence retrieval system, it is also essential for legal practitioners to have a concise representation of the core legal issues described in a legal text. One way to list the core legal issues is by keywords or key phrases, which are known as "*catchphrases*" in the legal domain.

Motivated by the requirements described above, we have the following two tasks:

1. **Catchphrase extraction**

2. **Precedence retrieval**

ⓘ

Catchphrases are short phrases from within the text of the document. Catchphrases can be extracted by selecting certain portions from the text of the document. A set of legal documents (Indian Supreme Court decisions) will be provided. For a few of these documents (training set), the catchphrases *(gold standard)* will also be provided. These catchphrases have been obtained from a well-known legal search system Manupatra (www.manupatra.co.in), which employs legal experts to annotate case documents with catchphrases. The rest of the documents will be used as the test set. The participants will be expected to extract the catchphrases for the documents in the test set.

Task 2:

For the precedent retrieval task, two sets of documents shall be provided:

1. Current cases: A set of cases for which the prior cases have to be retrieved.

2. Prior cases: For each "current case", we have obtained a set of prior cases that were actually cited in the case decision. These cited prior cases are present in the second set of documents along with other (not cited) documents.

For each document in the first set, the participants are to form a list of documents from the second set in a way that the cited prior cases are ranked higher than the other (not cited) documents.

NOTE:

1. **A participant team may participate in either or both the sub-tasks.**

2. **Each team can have at most 4 participants.**

Download the dataset here

Please click the following link to download the dataset. The instructions for use is mentioned in the included README file.

[LINK TO THE DATASET]

We are happy to share the dataset. Please cite the overview paper while using this dataset in your research work.

The dataset is also used by the following papers:

- Automatic Catchphrase Identification from Legal Court Case Documents [Mandal et al.-2017][URL=https://dl.acm.org/citation.cfm?id=3133102]

ⓘ