COL764: Project Milestone 1

Topic Selection and Brief Statement

Dipanshu Patidar 2018CS50405
Manoj Kumar 2018CS50411
Team: Kota Factory

Legal Search

Abstract Idea:

In countries like India following the Common Law system, there are two primary sources of law – Statutes (established laws) and Precedents (prior cases). Statutes deal with applying legal principles to a situation (facts/scenario/circumstances which lead to filing the case). Precedents or prior cases help a lawyer understand how the Court has dealt with similar scenarios in the past, and prepare the legal reasoning accordingly.

When a lawyer is presented with a situation (that will potentially lead to the filing of a case), it will be very beneficial to him/her if there is an automatic system that identifies a set of related prior cases involving similar situations as well as statutes/acts that can be most suited to the purpose in the given situation. Such a system shall not only help a lawyer but also benefit a common man, in a way of getting a preliminary understanding, even before he/she approaches a lawyer. It shall assist him/her in identifying where his/her legal problem fits, what legal actions he/she can proceed with (through statutes) and what were the outcomes of similar cases (through precedents).

Problem Statements:

The main goal of our problem is to identify the most relevant cases related to the given situation. This is achieved by the following subtasks:

1. The topic words are extracted from the given situation.
2. Then topic words are considered as a query to search the relevant prior cases by using the retrieval model.

To implement such a model, we divide the problem into two tasks:

Task 1: Identifying relevant prior cases for a given situation
Task 2: Identifying most relevant statutes for a given situation

Information Retrieval Techniques:

- For task 1, we plan to implement any of the TF-IDF-based topic extraction or TextRank-based topic extraction method.
- For task 2, we plan to implement any of the following:
    a. Vector space model
    b. Probabilistic model
    c. Language model

Dataset and problem source:

The problem is taken from the AILA tasks in 2019. [link: https://sites.google.com/view/fire-2019-aila/]

The dataset is taken from:
[https://zenodo.org/record/4063986#.X3dBUMIzbX4]