Technology Students’ Gymkhana
Indian Institute of Technology Kharagpur

Open Soft 2019 Problem Statement

7 March, 2019

Developing a Legal Search Engine for Law Experts and Commoners

Motivation

With the advancement of the Web and an increasing number of legal documents being available digitally, it is now intractable for legal practitioners to manually find relevant information (prior cases, related acts/statutes etc, law reports etc.) that would assist an ongoing case. Although there exist legal search systems like Manupatra, Westlaw India, the systems are not very efficient (e.g., only allowing keyword search). On discussing with legal experts, we found out that they still mostly rely on Google search to find relevant documents. Thus, a domain-specific search system is required for legal search and retrieval, especially in the Indian context. This project is about developing such a system. The system should ideally cater to both domain experts (e.g., lawyers who want to find prior cases to cite during an ongoing case) and common people (who are facing a problem and want some legal opinion).

Objective

India follows the *Common Law* system, which says that the decision of a case will be based on the decisions given to *similar prior cases.* Hence, there are two primary sources of law : (1) Statutes/Acts, which are the laws made by the legislature, and (2) Precedents or previous case judgements, which contain solutions to similar legal problems, not indicated in the Statutes/Acts. Precedents help a lawyer understand how the Court has dealt with similar scenarios in the past, and prepare the legal reasoning accordingly. Hence lawyers have to go through hundreds of prior cases.

The search engine we are planning to develop, will assist a lawyer and a common man as follows:

A lawyer will be helped in finding relevant prior-cases that he/she may need to argue an ongoing case. On providing the basic keywords that describe the ongoing case, the system shall return relevant prior case documents along with some basic information related to the documents, that will help the lawyer in having a quick glimpse of relevant prior cases.

A common man, who will type in a natural language query, will be helped by retrieving acts, sections of acts and relevant cases, which will potentially help the common man have a basic idea before seeking legal advice.

Background about legal documents and Acts

In this section, we give a brief background about legal documents and Acts in the Indian judiciary, which will help in understanding the problem statement.

Legal Documents

There is a wide variety of legal documents like patents, risk assessment documents, court case judgements, contracts, etc. In this problem, we will concentrate on *court case documents, specifically, documents from the Supreme Court of India*.

A case document is represented by a *case title*. The case title is essentially the two parties involved. The format is *PartyName1 Vs. PartyName2* (eg., Rohan Sharma Vs. Rita Verma). A party can either be a living person or a collective entity (eg., Union of India, State of West Bengal, Tata Steel, etc.). The document broadly has two parts: metadata information and textual content.

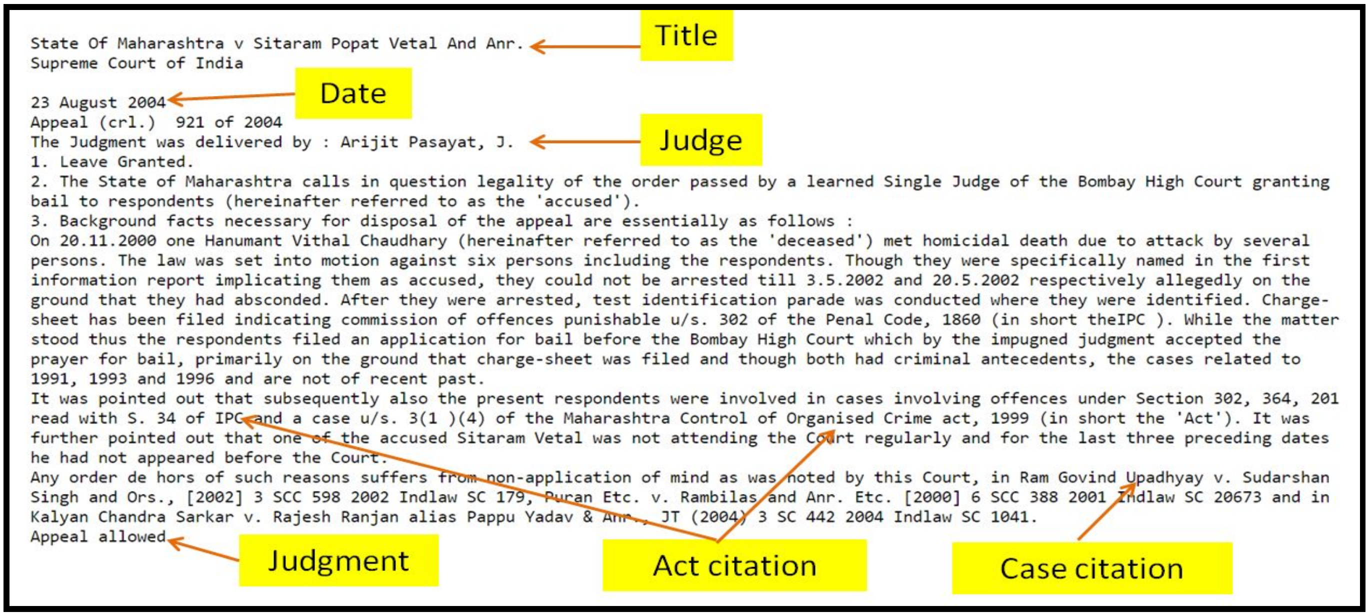
Metadata contains information like Date, Court, Judge, Appeal Numbers, etc.
Date : Dates are provided at the beginning of the case document, as a metadata information.
Name of Judge : The judge name is present at the beginning of the case document, as a metadata information
[Metadata contains some other fields as well, but they are not relevant to this particular problem.]

Textual content: It starts by describing the *facts of the case*, next comes the *arguments* where *precedent cases and relevant acts* are cited, and finally the *judgement* along with the reason for judgement.

Judgement : The last line of the case document will usually contain the final verdict in a 2-3 word phrase. The phrase will contain either “allowed”, “dismissed”, “disposed” or “order accordingly”, e.g., “appeal allowed” , “petition dismissed”.

Figure 1 shows an example case document.

Fig 1 : Example Case Document



```
State Of Maharashtra v Sitaram Popat Vetal And Anr.
Supreme Court of India

23 August 2004
Appeal (crl.)  921 of 2004
The Judgment was delivered by : Arijit Pasayat, J.
1. Leave Granted.
2. The State of Maharashtra calls in question legality of the order passed by a learned Single Judge of the Bombay High Court granting
bail to respondents (hereinafter referred to as the 'accused').
3. Background facts necessary for disposal of the appeal are essentially as follows :
On 20.11.2000 one Hanumant Vithal Chaudhary (hereinafter referred to as the 'deceased') met homicidal death due to attack by several
persons. The law was set into motion against six persons including the respondents. Though they were specifically named in the first
information report implicating them as accused, they could not be arrested till 3.5.2002 and 20.5.2002 respectively allegedly on the
ground that they had absconded. After they were arrested, test identification parade was conducted where they were identified. Charge-
sheet has been filed indicating commission of offences punishable u/s. 302 of the Penal Code, 1860 (in short theIPC ). While the matter
stood thus the respondents filed an application for bail before the Bombay High Court which by the impugned judgment accepted the
prayer for bail, primarily on the ground that charge-sheet was filed and though both had criminal antecedents, the cases related to
1991, 1993 and 1996 and are not of recent past.
It was pointed out that subsequently also the present respondents were involved in cases involving offences under Section 302, 364, 201
read with S. 34 of IPC and a case u/s. 3(1 )(4) of the Maharashtra Control of Organised Crime act, 1999 (in short the 'Act'). It was
further pointed out that one of the accused Sitaram Vetal was not attending the Court regularly and for the last three preceding dates
he had not appeared before the Court.
Any order de hors of such reasons suffers from non-application of mind as was noted by this Court, in Ram Govind Upadhyay v. Sudarshan
Singh and Ors., [2002] 3 SCC 598 2002 Indlaw SC 179, Puran Etc. v. Rambilas and Anr. Etc. [2000] 6 SCC 388 2001 Indlaw SC 20673 and in
Kalyan Chandra Sarkar v. Rajesh Ranjan alias Pappu Yadav & Anr., JT (2004) 3 SC 442 2004 Indlaw SC 1041.
Appeal allowed
```

A case has a **citation ID** assigned by a publisher, through which it is possible to cite the case. A particular case can be assigned different citation ids by different publishers like AIR, SCC, Indlaw etc. In the above example the “case citation” of Ram Govind Upadhyay v. Sudarshan Singh and Ors. has two IDs - (i) SCC 598, and (ii) 2002 Indlaw SC 179 (which are assigned by two different publishers SCC and Indlaw). In this project, we will be using Indlaw citation ids. This Indlaw id for each case is included in the given dataset. Each Supreme Court case document will hence have a unique id of the format *<year> Indlaw SC <number>*.

A case document is usually associated with a **subject** (e.g., Criminal, Rent-Control, etc.) and a set of **catchwords** (legal keywords), which help in understanding basic information of the case without fully reading it. While these subjects / catchwords are not provided in the original court case judgement text, such information is usually assigned manually/algorithmically. For example, the case judgement titled : “Jagannathan Pillai v Kunjithapadam Pillai And Ors.” has
subject : Constitution; Family & Personal; Practice & Procedure; Women & Children
catchwords : Succession & Inheritance
The given dataset (described later) contains subjects and catchwords for all Supreme Court cases.

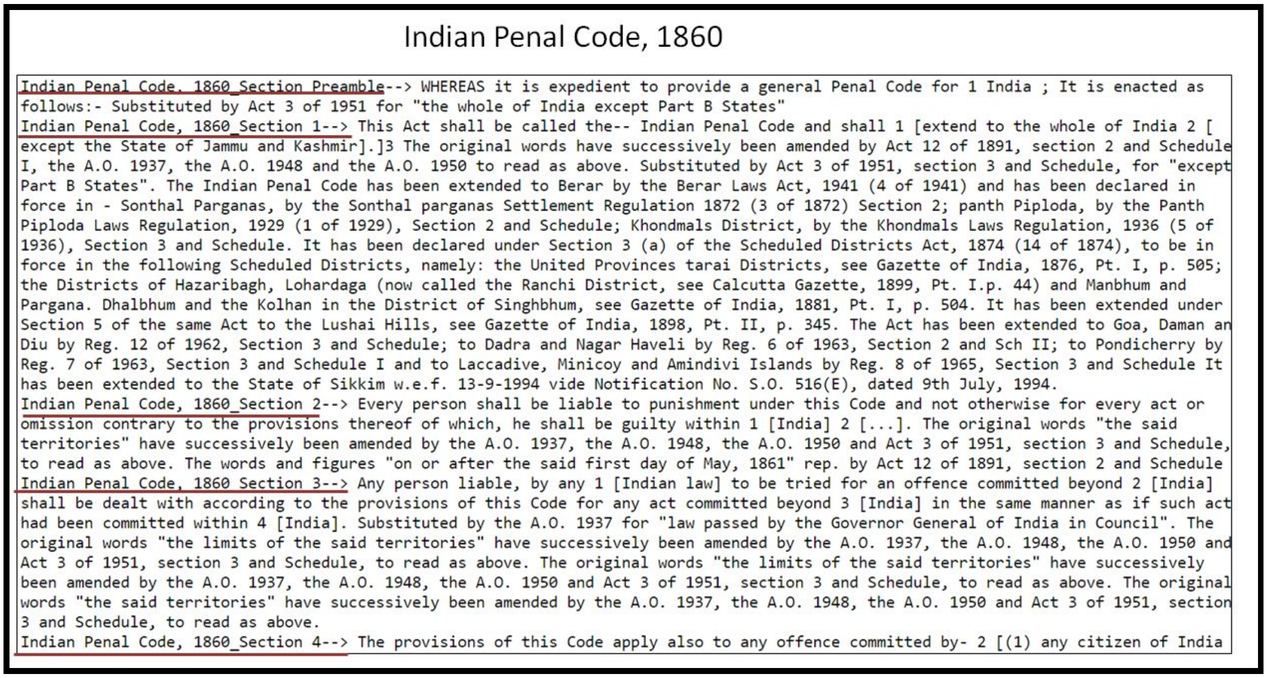
Acts/Legislations

Acts refer to already existing rules/rights that a citizen of a country has. Examples include “Constitution”, “Indian Penal Code, 1860”, “Transfer of Property Act, 1882”, “Dowry Prohibition Act, 1961”, etc.

An Act can be a “central” act (the examples stated above) or a “state” acts which is particular to a state (e.g., “West Bengal State Council of Higher Education Act, 2015”, “Goa Civil Courts Act, 1965”, etc.)

An Act has a textual content. It first has a preamble and sections subsequently. Figure 2 shows the content of an example Act (Indian Penal Code, 1860).

Fig. 2 : Example content of an Act



Indian Penal Code, 1860

```
Indian Penal Code, 1860_Section Preamble--> WHEREAS it is expedient to provide a general Penal Code for 1 India ; It is enacted as
follows:- Substituted by Act 3 of 1951 for "the whole of India except Part B States"
Indian Penal Code, 1860_Section 1--> This Act shall be called the-- Indian Penal Code and shall 1 [extend to the whole of India 2 [
except the State of Jammu and Kashmir].]3 The original words have successively been amended by Act 12 of 1891, section 2 and Schedule
I, the A.O. 1937, the A.O. 1948 and the A.O. 1950 to read as above. Substituted by Act 3 of 1951, section 3 and Schedule, for "except
Part B States". The Indian Penal Code has been extended to Berar by the Berar Laws Act, 1941 (4 of 1941) and has been declared in
force in - Sonthal Parganas, by the Sonthal parganas Settlement Regulation 1872 (3 of 1872) Section 2; panth Piploda, by the Panth
Piploda Laws Regulation, 1929 (1 of 1929), Section 2 and Schedule; Khondmals District, by the Khondmals Laws Regulation, 1936 (5 of
1936), Section 3 and Schedule. It has been declared under Section 3 (a) of the Scheduled Districts Act, 1874 (14 of 1874), to be in
force in the following Scheduled Districts, namely: the United Provinces tarai Districts, see Gazette of India, 1876, Pt. I, p. 505;
the Districts of Hazaribagh, Lohardaga (now called the Ranchi District, see Calcutta Gazette, 1899, Pt. I.p. 44) and Manbhum and
Pargana. Dhalbhum and the Kolhan in the District of Singhbhum, see Gazette of India, 1881, Pt. I, p. 504. It has been extended under
Section 5 of the same Act to the Lushai Hills, see Gazette of India, 1898, Pt. II, p. 345. The Act has been extended to Goa, Daman an
Diu by Reg. 12 of 1962, Section 3 and Schedule; to Dadra and Nagar Haveli by Reg. 6 of 1963, Section 2 and Sch II; to Pondicherry by
Reg. 7 of 1963, Section 3 and Schedule I and to Laccadive, Minicoy and Amindivi Islands by Reg. 8 of 1965, Section 3 and Schedule It
has been extended to the State of Sikkim w.e.f. 13-9-1994 vide Notification No. S.O. 516(E), dated 9th July, 1994.
Indian Penal Code, 1860_Section 2--> Every person shall be liable to punishment under this Code and not otherwise for every act or
omission contrary to the provisions thereof of which, he shall be guilty within 1 [India] 2 [...]. The original words "the said
territories" have successively been amended by the A.O. 1937, the A.O. 1948, the A.O. 1950 and Act 3 of 1951, section 3 and Schedule,
to read as above. The words and figures "on or after the said first day of May, 1861" rep. by Act 12 of 1891, section 2 and Schedule
Indian Penal Code, 1860_Section 3--> Any person liable, by any 1 [Indian law] to be tried for an offence committed beyond 2 [India]
shall be dealt with according to the provisions of this Code for any act committed beyond 3 [India] in the same manner as if such act
had been committed within 4 [India]. Substituted by the A.O. 1937 for "law passed by the Governor General of India in Council". The
original words "the limits of the said territories" have successively been amended by the A.O. 1937, the A.O. 1948, the A.O. 1950 and
Act 3 of 1951, section 3 and Schedule, to read as above. The original words "the limits of the said territories" have successively
been amended by the A.O. 1937, the A.O. 1948, the A.O. 1950 and Act 3 of 1951, section 3 and Schedule, to read as above. The original
words "the said territories" have successively been amended by the A.O. 1937, the A.O. 1948, the A.O. 1950 and Act 3 of 1951, section
3 and Schedule, to read as above.
Indian Penal Code, 1860_Section 4--> The provisions of this Code apply also to any offence committed by- 2 [(1) any citizen of India
```

Data provided for this project

The data is downloadable from the following link:
https://drive.google.com/file/d/1WQw2o70nYHr7B_7yQ5qzdWCG9TlVBIxU/view?usp=sharing
The data contains the following:

1. **Case Documents (CaseDocuments.zip)** : A set of 53,210 Supreme Court case documents from the years 1953--2018. The documents are in .txt format. An example document was shown in Fig. 1

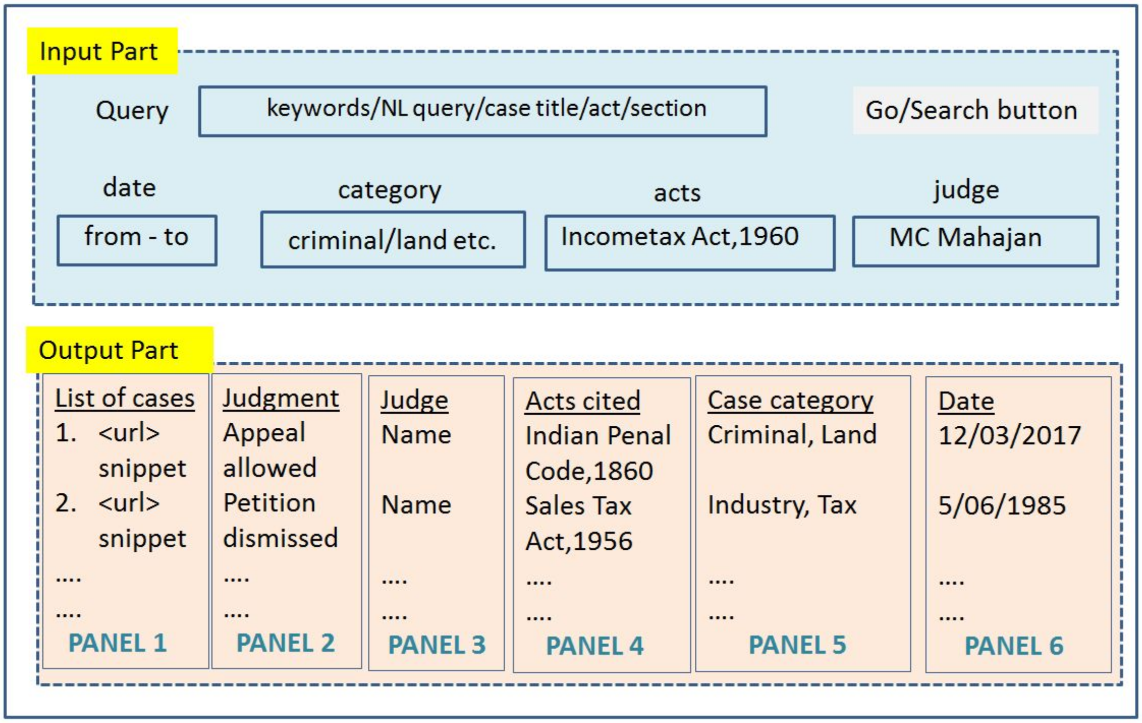
2. **Case ID (doc_path_ttl_id.txt)** : Each line in this file is of the format :
   <filename> --> <case title> --> <ID>
   Eg.: 2002_S_732 --> Ram Govind Upadhyay v Sudarshan Singh and Others --> 2002 Indlaw SC 179
   This line means that, the file with name “2002_S_732” has the case document for “Ram Govind Upadhyay v Sudarshan Singh and Others” whose unique citation id is “2002 Indlaw SC 179”
   This citation id should be used while linking documents present as citations inside a case document.

3. **Case subject Information (subject_keywords.txt)** : A text file containing the subject, catchwords, and related information about the cases. The format is :
   filename --> case title --> subject $$$ catchwords
   Eg., 1987_K_80 --> Jagannathan Pillai v Kunjithapadam Pillai And Ors. --> Constitution; Family & Personal; Practice & Procedure; Women & Children$$$ Succession & Inheritance
   This line means that, the file with name “1987_K_80” has the case document for “Jagannathan Pillai v Kunjithapadam Pillai And Ors.” which has
   subject : Constitution; Family & Personal; Practice & Procedure; Women & Children
   catchwords : Succession & Inheritance
   Note that some documents may not have either subject or keywords

4. **List of acts (actlist.txt):** A list of ~11K acts (both central and state acts) present in the Indian legal system.

5. Full textual content of the Acts (**Acts.zip**)
   Each act is in a txt file, with format as was shown in Fig. 2

Fig. 3 : A prototype image of the web-based search engine for Indian legal documents



Input Part

Query: keywords/NL query/case title/act/section

Go/Search button

date: from - to

category: criminal/land etc.

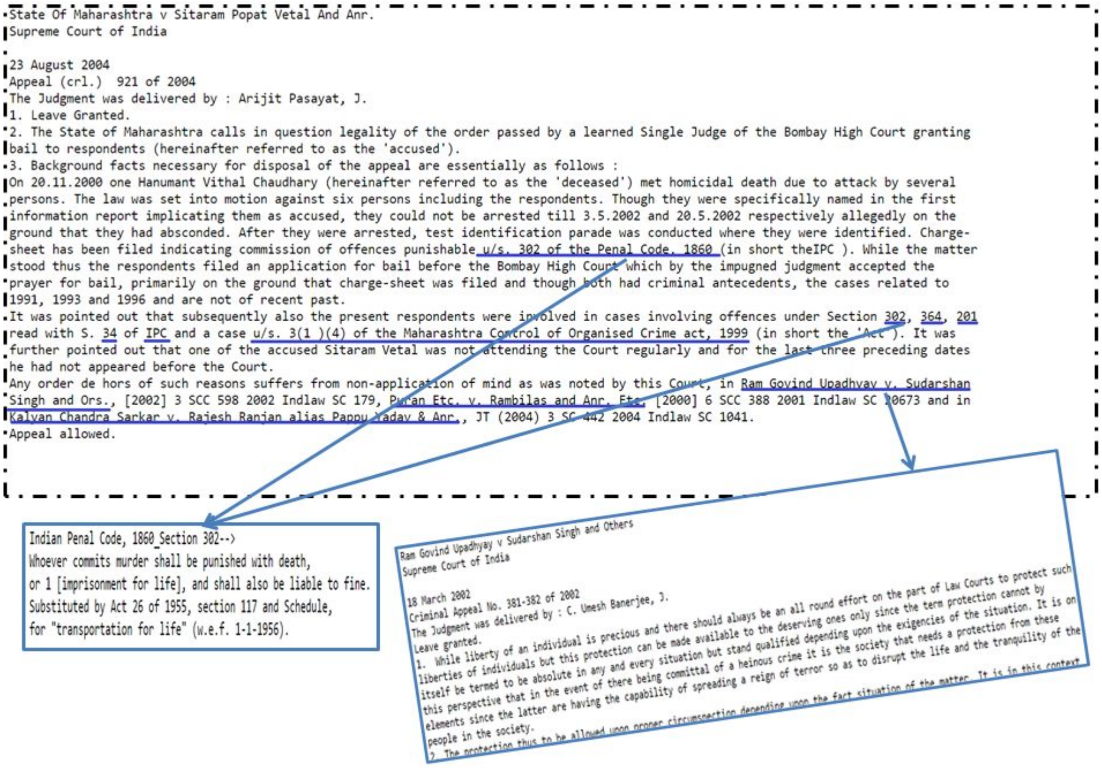
acts: Incometax Act,1960

judge: MC Mahajan

Output Part

List of cases	Judgment	Judge	Acts cited	Case category	Date
1. <url> snippet	Appeal allowed	Name	Indian Penal Code,1860	Criminal, Land	12/03/2017
2. <url> snippet	Petition dismissed	Name	Sales Tax Act,1956	Industry, Tax	5/06/1985
....	....	....	....	....	....
....	....	....	....	....	....
PANEL 1	PANEL 2	PANEL 3	PANEL 4	PANEL 5	PANEL 6

Fig. 4 : Full case document visualization :
popups windows and extracted informations highlighted



State Of Maharashtra v Sitaram Popat Vetal And Anr.
Supreme Court of India

23 August 2004
Appeal (crl.) 921 of 2004
The Judgment was delivered by : Arijit Pasayat, J.
1. Leave Granted.
2. The State of Maharashtra calls in question legality of the order passed by a learned Single Judge of the Bombay High Court granting bail to respondents (hereinafter referred to as the 'accused').
3. Background facts necessary for disposal of the appeal are essentially as follows :
On 20.11.2000 one Hanumant Vithal Chaudhary (hereinafter referred to as the 'deceased') met homicidal death due to attack by several persons. The law was set into motion against six persons including the respondents. Though they were specifically named in the first information report implicating them as accused, they could not be arrested till 3.5.2002 and 20.5.2002 respectively allegedly on the ground that they had absconded. After they were arrested, test identification parade was conducted where they were identified. Charge-sheet has been filed indicating commission of offences punishable u/s. 302 of the Penal Code, 1860 (in short theIPC). While the matter stood thus the respondents filed an application for bail before the Bombay High Court which by the impugned judgment accepted the prayer for bail, primarily on the ground that charge-sheet was filed and though both had criminal antecedents, the cases related to 1991, 1993 and 1996 and are not of recent past.
It was pointed out that subsequently also the present respondents were involved in cases involving offences under Section 302, 364, 201 read with S. 34 of IPC and a case u/s. 3(1)(4) of the Maharashtra Control of Organised Crime act, 1999 (in short the 'Act'). It was further pointed out that one of the accused Sitaram Vetal was not attending the Court regularly and for the last three preceding dates he had not appeared before the Court.
Any order de hors of such reasons suffers from non-application of mind as was noted by this Court, in Ram Govind Upadhyay v. Sudarshan Singh and Ors., [2002] 3 SCC 598 2002 Indlaw SC 179, Puran Etc. v. Rambilas and Anr. Etc. [2000] 6 SCC 388 2001 Indlaw SC 20673 and in Kalyan Chandra Sarkar v. Rajesh Ranjan alias Pappu Yadav & Anr., JT (2004) 3 SC 442 2004 Indlaw SC 1041.
Appeal allowed.

Indian Penal Code, 1860_Section 302-->
Whoever commits murder shall be punished with death,
or 1 [imprisonment for life], and shall also be liable to fine.
Substituted by Act 26 of 1955, section 117 and Schedule,
for "transportation for life" (w.e.f. 1-1-1956).

Ram Govind Upadhyay v Sudarshan Singh and Others
Supreme Court of India

18 March 2002
Criminal Appeal No. 381-382 of 2002
The Judgment was delivered by : C. Umesh Banerjee, J.
Leave granted.
1. While liberty of an individual is precious and there should always be an all round effort on the part of Law Courts to protect such liberties of individuals but this protection can be made available to the deserving ones only since the term protection cannot by itself be termed to be absolute in any and every situation but stand qualified depending upon the exigencies of the situation. It is on this perspective that in the event of there being committal of a heinous crime it is the society that needs a protection from these elements since the latter are having the capability of spreading a reign of terror so as to disrupt the life and the tranquility of the people in the society.
2. The protection thus to be allowed upon proper circumspection depending upon the fact situation of the matter. It is in this context

Problem Statement

Input: A query to the legal search engine (to be developed) can be of the following types:
Type 1 query: A set of legal/non-legal keywords (e.g., culpable homicide, murder)
Type 2 query: Name of a particular Act / section of an Act (e.g., IPC 302, Indian Institutes of Management Act, 2017)
Type 3 query: Title of a particular case (e.g., “State Of Maharashtra v Sitaram Popat Vetal And Anr.”)
Type 4 query: A query in natural language (e.g., “I ordered some materials online on Shoppers stop. Amount was deducted from my account. But shoppers stop have not delivered the materials and also they are not ready to refund my money which I have already paid. I want to take strict action against them.” or “owner of property for 20 years, tenant did not pay rent, tenant not vacating property”)

Optional filters: The following filters should be applicable over a search (for a given query):

1. Time range (e.g., January 2000 to June 2006 ; before 1990 ; after 2016) - only relevant cases in this time period should be returned,

2. Category (criminal, tax, property-related, etc.): case documents of the entered category should only be shown. Consider both subjects and keywords to be the category of a case.
3. Judge name (e.g., “C. Umesh Banerjee”, “Y. Chandrachud”, etc.) - relevant judgements delivered by the particular judge should only be retrieved
4. Acts (Public Provident Fund Act, 1968 , Payment of Bonus Act, 1965 etc.) - relevant documents citing those acts should only be returned

Note : The “Category” and “Acts” filter can have 1 or more inputs, for eg., there can be 1 or more categories entered, and/or 1 or more acts entered. You can have an upper limit restriction for these filters.

An example scenario :
Query (of type 1) : “murder, property”
Filters :
date: 01/01/2010 - 31/12/2016 ;
acts: Code of Criminal Procedure, 1973 + Married Women's Property Act, 1874 ;
Judge: A. Singh
For this query and filters, the system must retrieve cases related to “murder” and “property” that were decided during the time period “Jan 2010 - Dec 2016” and which cited the acts “Code of Criminal Procedure, 1973” or “Married Women's Property Act, 1874” and that were ruled by judge “A. Singh”

Output :
The output should have the features shown in Fig. 3. There will be one portion of the interface for taking the input query (along with filters), and another portion for displaying the search results. We envision that the part for displaying the search results will be partitioned into various panels (specifically, six panels shown in Fig. 3), for displaying various types of information relevant to a particular type of query.

Note that, Figure 3 is just a prototype, and you are free to design the interface in your own way.

Desired output for Type 1 and Type 4 queries :

In Panel 1 :
A ranked list of relevant prior cases, along with the following informations :

1. The title of the case should be shown as an url. The anchor text of the url should be the case title and the citation ID.
2. A short summary of the case (similar to a snippet shown on Web search engines)

The url, if clicked, should show the full text (maybe as a separate window, or in another panel) with the following informations highlighted (example representation in Fig. 4):

a. The sections of acts (Section 12 of the Consumer Protection Act,1986) / acts cited (Consumer Protection Act,1986) from this case
b. Hyperlinks to these sections/acts, which if clicked on, will provide the full text of the sections/acts
c. Highlight precedent cases cited. Note that not all case citations inside the case document will have a citation ID. If a citation ID is provided, you are required to hyperlink it. If there is no citation ID, you are still required to highlight it.
d. Hyperlinks to the precedent cases

The other panels should show various information about each relevant case shown in Panel 1 (in the same ranked order as in Panel 1).
In Panel 2 : The final judgment
In Panel 3 : Name of the judge who gave the verdict
In Panel 4 : Acts cited by the case
In Panel 5 : Case category
In Panel 6 : Date of the case

Note that, all six panels should show different information pertaining to the same case, and these cases should be ranked in decreasing order of relevance to the query.

Desired output for Type 2 query :

1. Link to the full text of the section/act that has been queried
2. Details of some cases that have cited the queried section / act (the panels show information about these cases, as stated earlier)

Desired output for Type 3 query :

Information about the particular case that has been queried. In case multiple cases are found relevant, all of them must be shown in decreasing order of relevance.

1. Highlighted full text document of relevant case (Fig. 4)
2. The information stated above for Type 1 queries should be shown here as well, for the relevant document(s).

Related Technical challenges

1. **Document Indexing** : You have to index the case documents for faster retrieval. Please refer to [1]. Decide indexing on what fields will be efficient, given the query and the different filters. You can index the documents using legal keywords, e.g., using a legal dictionary such as https://thelawdictionary.org/letter/a/. However, note that the system should also be able to retrieve semantic matches. For instance, if a query contains “stab”, the system should be able to retrieve documents containing “murder”, “homicide”, etc. Especially, expect queries of type 4 (given by a layman, who has little or no legal knowledge) to have very low match with legal keywords. In this situation, the system should still be able to retrieve documents that are *semantically related* to the query.

2. **Spelling correction and resolution** : If the query is an act / section of act, or a case title, an efficient spell checking module should be used to handle minor typos and abbreviations.
   Eg. 1 : The query “State Of Maharashtra vs Sitaram Popat Vetal And Another” or the query “Maharashtra v. Sitaram Popat Vetal And Anr.” should be able to retrieve the case document titled “State Of Maharashtra v Sitaram Popat Vetal And Anr.”
   Eg. 2 : The query “Ashray Adhikar Abhiyan Vs. UOI and ors” should be able to retrieve the case document titled “Ashray Adhikar Abhiyan v Union of India and others”
   Eg. 3 : The query “IPC section 302” should be able to retrieve Section 302 of the Act Indian Penal Code, 1860

3. **Information extraction from the documents** : You should extract the following information from the full text of the case documents:
   a. Date: provided at the beginning of the case document

   b. Judgement: The last line of the case document will usually contain the final verdict in a 2-3 word phrase. The phrase will contain either “allowed”, “dismissed”, “disposed” or “order accordingly”. Eg. “appeal allowed” , “petition dismissed”.

      Note that, Indian legal documents are not very systematically written, so you may sometimes find this judgement at the beginning of the document (after the metadata information). Design a flexible algorithm for this.

   c. Name of Judge: The judge name is present at the beginning of the case document, as a metadata information. Again, since Indian legal documents lack consistency, a judge name may be written in different formats, eg., with or without abbreviations. Try to normalize these names, as best as possible.

   d. Snippet (Summary generation) : generate a summary of the case document using an unsupervised summarization algorithm. Refer [2] for details. The summary should be upto 100 words. It should well represent the facts of the case (initial part of the case document) and the reason for the final judgement (concluding part of the case document).

4. **Disambiguation of Acts:** There can be multiple variants of an act with same name but different dates. For eg.,

   a. Code of Criminal Procedure, 1898 ; Code of Criminal Procedure, 1973
   b. Code of Civil Procedure, 1882 ; Code of Civil Procedure, 1908
   c. 10 variants of the Kerala Finance Act

   Sometimes a case document may only refer to an act, without mentioning the date (eg., section 145 of the Code of Criminal Procedure). In such cases, link the citation to the most recent version of the corresponding Act.
   Also, a case may sometimes refer to a *subsection* of an Act, e.g., “Section I(j) of IPC”. This citation should be linked to the text of Section I of IPC. In other words, subsection information can be ignored.

5. **Handling abbreviations:** Some popular Acts are often referred to using abbreviations. For instance:

   CrPC: Code of Criminal Procedure, Criminal Procedure Code
   CPC: Code of Civil Procedure, Civil Procedure Code
   IPC: Indian Penal Code
   Constitution: Constitution of India 1950
   TADA: Terrorist and Disruptive Activities (Prevention) Act, 1987
   FEMA: Foreign Exchange Management Act, 1999

6. **BONUS: Query Suggestion:** Though this is not a required feature, it will be better if the system can suggest alternate queries for a given query. E.g., if someone types a query "unintended murder", the suggestions can be "accidental killing", “culpable homicide”, etc.

Submission Instructions

1. The system should be a purely web-based platform. There should not be any OS dependency or library specific dependencies. You must finally submit a URL, where one can use the system.
2. The interface should be easy-to-use, since it will be used by people who are not experts in computer engineering (e.g., lawyers, and common people). For instance, you can include a user manual in the system.
3. Provide a soft copy of the source code, interface, etc. which you have developed. Source codes should be neatly formatted, commented and modular. The codes should be written in a way that, individual modules can be modified later or new methods can be inserted with ease. For instance, if later it is decided to use a different algorithm to generate the snippet, it should be easy to make this change.
4. A technical report vis-à-vis a clear description of each step you have carried out. Please clearly state any reasonable assumption if you make in your implementation. The reference to all tools you have used should be stated explicitly with their sources.
5. **The report should also contain instructions to install the system on a new web server, and to run the code.**

References

1. Indexing: Chapter 1 of https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf
2. Summarization: : https://rare-technologies.com/text-summarization-in-python-extractive-vs-abstractive-techniques-revisited/