

IDENTIFYING KEYWORDS IN LEGAL ARTICLES USING ML TECHNIQUES



Group 01

E/15/077- K.P.W.A.K.K. Dilhani

E/15/211- S.A.I.Maduwanthi

E/15/279- L.S.W.S.Premathilaka

BACKGROUND



PROBLEM

- Large number of Documents
- Need to go through whole document
- Slow
- Prone to mistakes





[Legislation](#)

[Subsidiary](#)

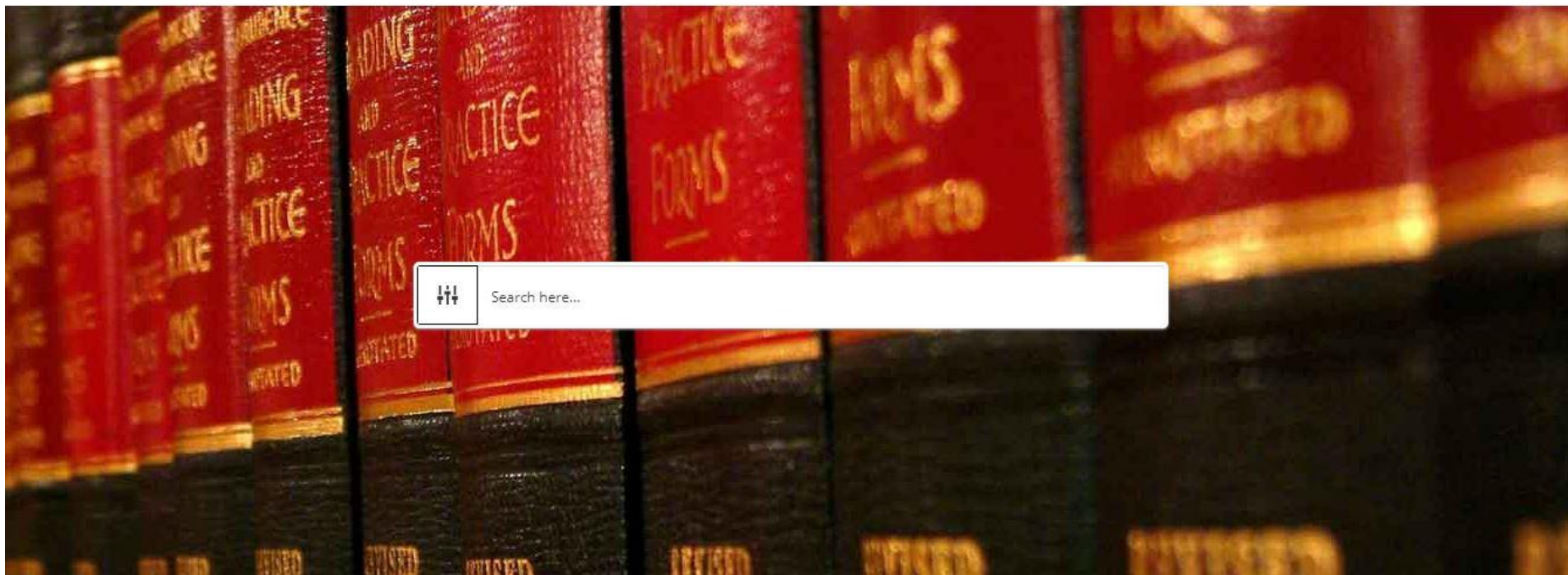
[Agreements](#)

[Case Law](#)

[Conventions](#)

[Articles](#)

[Links](#)



SIGN-IN

SIGN-UP

HELP

FAQ



Enter search terms



Home

Statutes

Caselaw

Statutory Regulations

More...

Browse Statutes

Browse by Year

Browse by Title

Legislative Enactments

Revised Statutes

0011	1000	1070	1773	1789	1799	1806	1821	1823	1835	1839	1840	1841	1842	1843	1844	1845
1846	1848	1852	1853	1856	1861	1862	1864	1865	1866	1867	1869	1870	1871	1873	1876	1877
1880	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897
1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1911	1912	1914	1915	1916
1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933
1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950
1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984

CURRENT ISSUES

- Doesn't give a summarized preview of a document
- Can't search phrases



If manual extraction happened.....

10 min for one Document

For **10, 000** Documents

For a person who **works 8 hrs**



Need **209 Days** to upload the files to the system

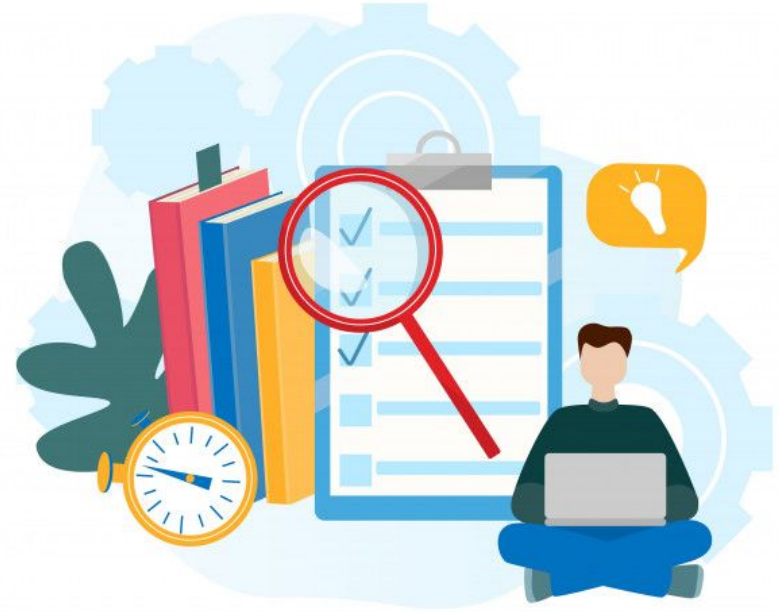
SOLUTION

A MACHINE LEARNING

based Model to

IDENTIFY and **EXTRACT**

necessary details...!!!



OBJECTIVES

- **IDENTIFY**

- the names of the previous judgements
- names of the statutes, section numbers
- key legal concepts

from the body of a judgement And

- **LINK**

them to relevant parts in previously uploaded informations



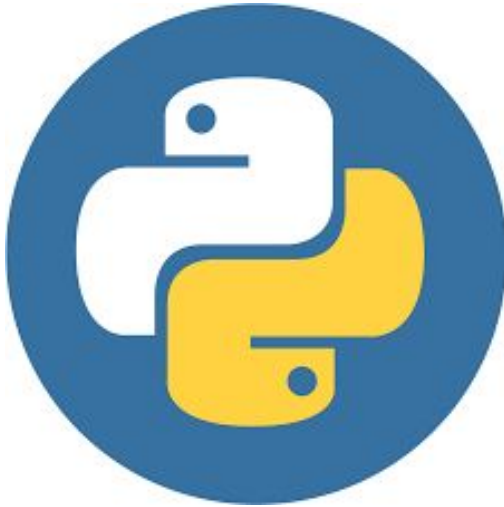
- **EXTRACT**



- Set of keywords in a particular document
- Court name
- The date and the year of the judgement
- Judges who are involved in the judgement and their roles
- Lawyers who are involved in the judgement and their roles

TECHNOLOGIES

Python



Machine learning



GIT

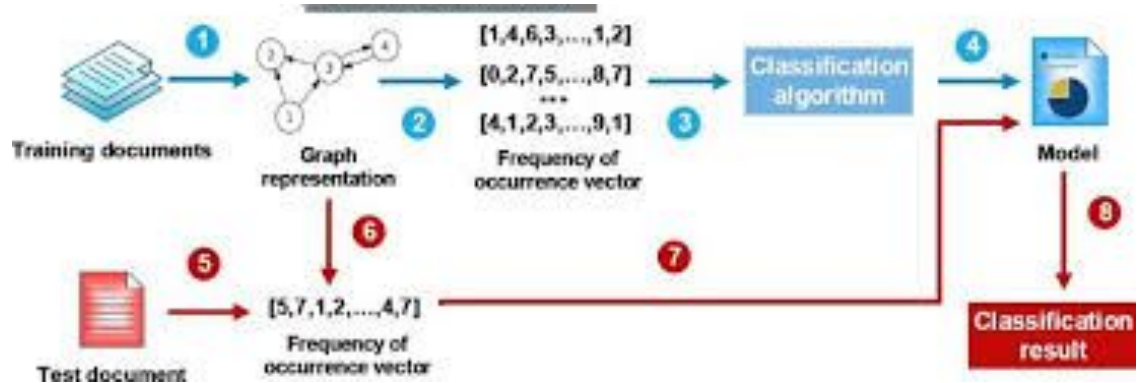


RELATED WORKS

- **Supervised learning method**
 - Implementation of algorithm by using linguistic knowledge and keyword extraction procedure
 - Improvement of KEA keyphrase extraction algorithm
- **Unsupervised learning method**
 - Design of keyphrase extraction algorithm for a single document using sentence clustering.
 - Focused on keyword extraction based on entropy difference between the intrinsic and extrinsic modes

- **Graph-based method**

- Proposed an algorithm for automatic indexing by co-occurrence graphs constructed from metaphors(key graph)
- Introduced a stochastic graph based method for extracting the most important sentences



Similarities and Differences of Related Works

SIMILARITIES	DIFFERENCES
Identify single keywords and keyphrases	No single method to identify different formats in reports
	Difficult to find related works on legal articles
	Variations of citing methods in case laws

METHODOLOGY

There are set of **special words and phrases** surrounding the previous judgment.

The case of *Kelner v. Baxter and Others*⁽²⁾ had set down the principle that when a person contracts on behalf of a non-existent company he was personally liable.

In the same case (page 180) reference had *ex parte Hartop*⁽³⁾ at 352 where Lord Erskine stated mere fact of a person professing to sign a contract of or as an agent for another *will not per se prevail as a contracting party to attach to the former*".

In the case of *Furnivall v. Coombes*⁽⁴⁾ a clause to responsibility was included and was held to be

It was not the intention of the Legislature to impose new obligations on the tenant. Even if a doubt is entertained, the Courts will lean to a construction that an enactment is not intended to impose a serious new obligation, but only to provide new or better means of enforcing an existing obligation (vide *Finch v. Bannister*¹; *Gaby v. Palmer*²; Craies on Statute Law (5th Edn.) p. 111). While the words in a statute should be construed according to the context (Craies ibid p. 150, 160), it is a sound rule of construction to give the same meaning to the same words occurring in different parts of an Act of Parliament (vide *Courtauld v. Legh*³).

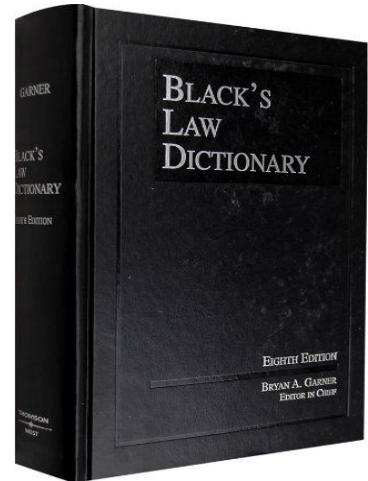
Adopting these canons of construction, this action cannot be maintained. I am of the view that the finding that, at the date the action was filed, the appellant was in arrears of rents for October 1963,

- Some of those patterns are,

The case of Klener **v.** Baxter and others
vide Finch **v.** Bannister
; Gaby **v.** Palmer
Vide Courtauld **v.** Legh
held in Peter **vs.** James that
the judgement of Peter **v.** James
in Peter **vs.** James

- There will be a certain **pattern of words** which will precede and follow, with the name of the **statutes** too.

- Use **Black's law online legal dictionary** to identify the key concepts in the judgments
- Develop **comprehensive lists of word patterns** that precede and follow the wanted informations.
- **Machine learning algorithm** should identify and extract the wanted details by looking at these pre-identified preceding and following words



- There are some words which are **specific to a particular document**

But

not a previous judgment, statute, legal concept, judge or lawyer name etc.

ALAWATUGODA RATEMAHATMEYA v. KIRIWANTE.

P. C., Nuwara Eliya, 8,928.

Forest Ordinance, No. 10 of 1885, chapter IV.—Prosecution under rules of 3rd February, 1887—Proof in such cases—Validity of judgment—Criminal Procedure Code, s. 372.

In a prosecution for clearing (for chena cultivation) a land at the disposal of the Crown without a permit, in breach of a rule framed under chapter IV. of the Ordinance No. 10 of 1885, it is necessary to prove that the land is not one within a reserved or village forest ; that it is at the disposal of the Crown ; that it is a chena ; that its extent and boundaries are so and so ; and that the accused cleared it.

A judgment of a criminal court should specify the offence with which the accused is charged, in terms of section 372 of the Criminal Procedure Code.

THE charge against the accused in this case was that he cleared for chena cultivation a land known as Komarikagalawahena (situated at Thenpila in Walapane), at the disposal of the Crown, without a permit from the Government Agent or Assistant Government Agent within whose jurisdiction the land was situated, in breach of clause 1 of the rules dated 3rd February, 1887, framed under chapter IV. of Ordinance No. 10 of 1885, and the Police Magistrate, after evidence heard, delivered judgment as follows : “ This is Crown land under the Ordinance 12 of 1840. “ Defendant is convicted and fined two rupees and fifty cents.”

On appeal (taken with leave of the Court below), *Wendt* appeared for accused appellant.

The Supreme Court quashed the conviction and remitted the case for further evidence.

For example, following words have some weighted importance in this legal document

‘CHENA CULTIVATION’, ‘RESERVED FOREST’, ‘ORDINANCE’

- They can also be considered as **keywords!!!**
- Use **TF-IDF** method to extract them.

EVALUATION

- In this phase we have shown that,
 - Identify the sentences consists of names of the previous judgements from the body of a judgements
 - Extract the keywords from a particular document.
 - Identify the case name
 - Identify the Court and reference numbers.
- Hope to evaluate the keyword extraction by using more keyword extraction methods and by combining them.



WHERE WE ARE NOW....!



DIFFICULTIES AND CHALLENGES

- Different Content arrangement of Case Law files
- No HTML files to work on
- Sometimes scanned PDF files give incorrect data
- Identifying exact judgement date
- Variations of citing previous cases



CONCLUSION

- **Answers on the questions of keyword identifying process**
 - How are the legal documents different from others,
 - What makes a legal document relevant
 - What features of a legal document matters the most
 - What tools and estimators work best for predictions in the legal domain
- **Our plan**
 - To implement a machine learning model searching documents using a single keyword or keyword phrase.

FUTURE WORKS

Develop an API with efficient Machine learning model to,

1. Identify **previous judgements names** exactly and **link** to the previous judgements which have already uploaded
2. Identify the names of the **statutes, section numbers** and link those names and section numbers to the relevant sections.
3. Identify the **key legal concepts** discussed in the body of the judgement
4. Identify the **judges and their roles** in the case



Work completion as planned...

What was planned	Completion
First build a web-based system to store these documents and keywords	Identify the sentences consists of names of the previous judgements from the body of a judgements
Then facilitate accurate and efficient searching when they are searched by keywords/ combinations of key words.	Extract the keywords from a particular document.
Finally develop a mechanism to automatically identify keywords when a document is given	Identify the case name,Court and reference numbers.

Work plan for second phase - CO425

- Identify the exact names of the previous judgements
- Identify the names of the statutes, section numbers.
- Identify the Judges who are involved in the judgement and their roles.
- Identify the Lawyers who are involved in the judgement and their roles.

Link that identified details to relevant sections which have previously uploaded

Identify the names of the legal concepts

Develop the API

THANK YOU..!

Any Questions...?

