

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





[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 |

PROBLEM

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



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

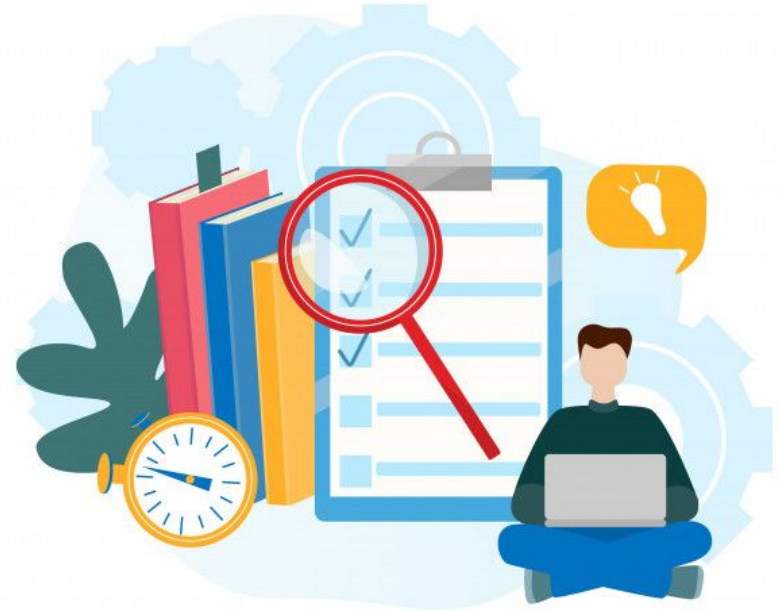
SOLUTION

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 conceptsfrom the body of a judgement



- **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

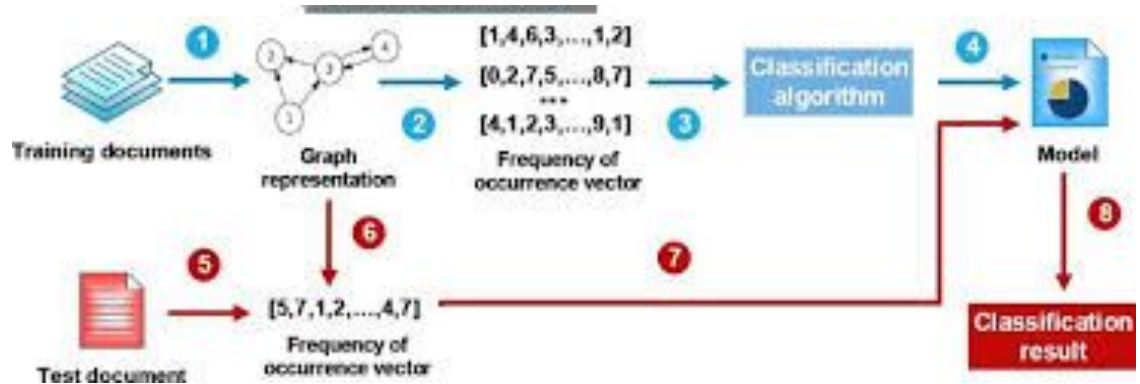


RELATED WORKS

- **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



Uniqueness of this research!!!!

- No single method to identify different formats in reports
- Difficult to find related works on legal domain
- 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).

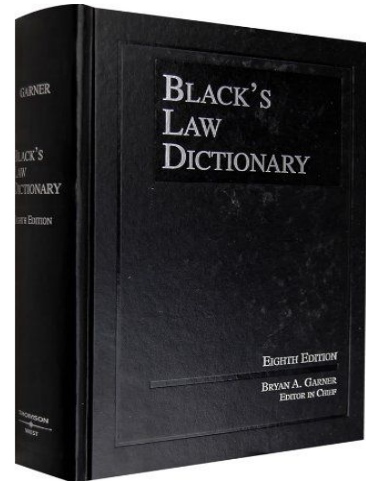
Adopting these canons of construction, this action cannot be maintained. I am of the view that the finding that, at the date the action was filed, the appellant was in arrears of rents for October 1963,

- 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**

- 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 **Text-Rank** and **TF-IDF** method to extract them.

Text-Rank method

- Based on PageRank
- Graph based ranking model

TF-IDF method

- Uses two different matrices - term frequency and inverse term frequency
- Can use to **extract** words that are unique to a given article

RESULTS

- **TF IDF Method**

| Document Type | Number of Correct Keywords | Number of Wrong Keywords |
|---------------|----------------------------|--------------------------|
| NLR | 110 | 143 |
| Supreme Court | 51 | 39 |

Accuracy for NLR = 0.4347

Accuracy for Supreme Court = 0.5666

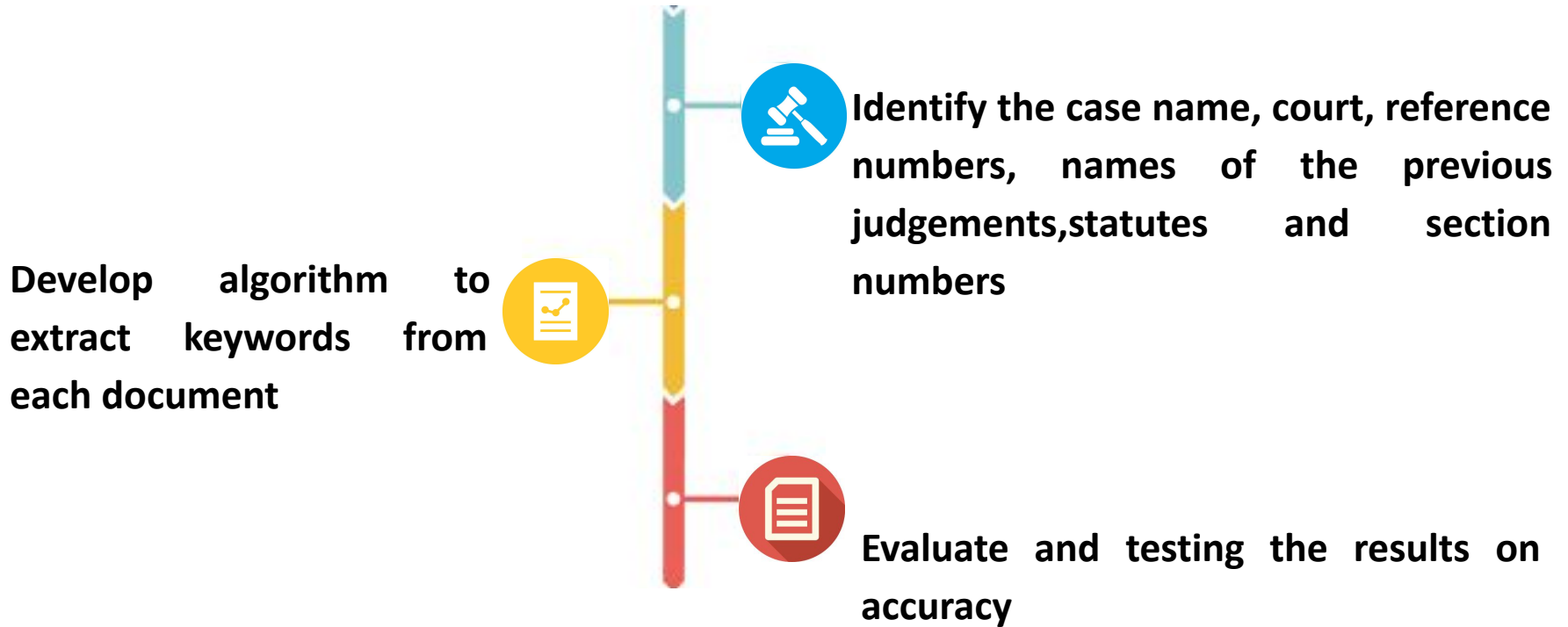
- **Text Rank Method**

| Document Type | Number of Correct Keywords | Number of Wrong Keywords |
|---------------|----------------------------|--------------------------|
| NLR | 186 | 311 |
| Supreme Court | 99 | 151 |

Accuracy for NLR = 0.3742

Accuracy for Supreme Court = 0.396

APPROVED MILESTONES



DIFFICULTIES AND CHALLENGES

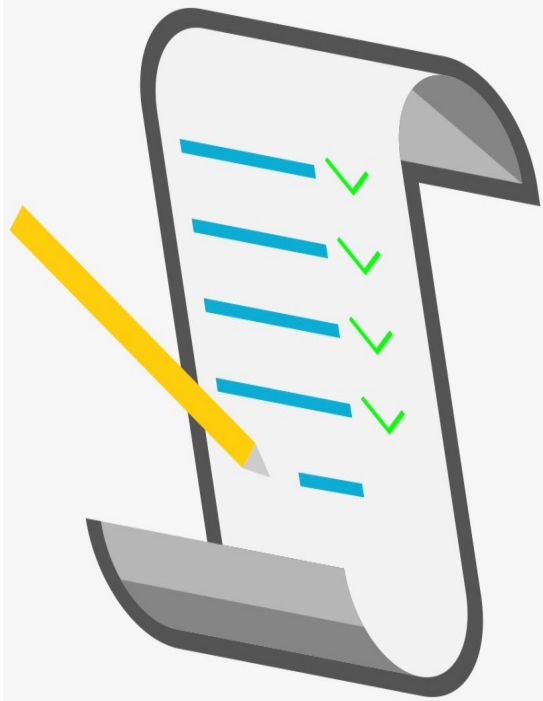
- Different Content arrangement of Case Law files
- No HTML files to work on
- Sometimes scanned PDF files give incorrect data
- Identifying exact judgement date
- Results Evaluation done manually



TIMELINE

| | | W 1 | W 2 | W 3 | W 4 | W 5 | W 6 | W 7 | W 8 | w 9 | W 10 | W 11 | W 12 | W 13 | W 14 | W 15 |
|---|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|
| 1 | Identify the case name,court name,reference numbers, date, judges and their statements. | | | | | | | | | | | | | | | |
| 2 | Develop algorithm to extract keywords from each document | | | | | | | | | | | | | | | |
| 3 | Evaluate and testing the results on accuracy | | | | | | | | | | | | | | | |

CONCLUSION



Considering the results

TF - IDF method is better than TEXT RANK method.

But

TF - IDF also not in ACCEPTABLE Accuracy Level.

DEMONSTRATION....!



THANK YOU...!

Any Questions...?

