ANALYSIS AND SUMMARIZATION OF LEGAL CASE REPORTS

Eseoghene Emuraye • Sharmin Kantharia • Hemanth Koganti

Introduction

- Legal documents have extensive theory making them difficult to study and analyze quickly.
- Hence, it is important to find easier ways to summarize large collection of legal texts.
- Using catchphrases, tokenized sentences and citations present in a legal document, we can analyze legal case reports through summarization.
- Such analysis can present insights on the different types of cases, their similarities and their differences.

Objectives

GOAL

Citations-based clustering and summarization of legal case reports to increase the efficiency with which legal cases are analyzed.

Main Objectives:

1. To cluster legal case reports based on similar citations.
2. To summarize case reports in each cluster to increase efficiency with which legal cases are analyzed.
3. Evaluate the accuracy of the generated summaries.

Data Source

- **Name** – Legal Case Reports Dataset
- **Source** – UCI Machine Learning Repository (Originally from the FCA website)
- **Number of Cases** – 3890
- The dataset contains Australian legal cases from the Federal Court of Australia (FCA) [2006 – 2009]

NOTE

The case reports analyzed in this project are taken from the Federal Court of Australia. While the format and legal proceedings may vary in each country, the techniques suggested can be applied to legal documents in the United States, given they have citations included in them.

Methodology

Input: Case reports
Output: Catchphrases, Sentences and Citations

Input: Extracted Citations
Output: Clusters of extracted sentences and catchphrases of case reports

Input: Clusters and Extracted Catchphrases
Output: Generated summaries per cluster

Input: Machine-generated summaries and Extracted Catchphrases
Output: Recall and Precision metric values with respect to ROUGE-N

Text Preprocessing

Extract the following components from the raw case reports and citations files

- Catchphrases
- Sentences
- Citations

Catchphrase Extraction

Original Text

```
<catchphrase "id=c0">interlocutory injunction</catchphrase>
<catchphrase "id=c1">quia timet injunction</catchphrase>
```

Extracted Catchphrases

```
['interlocutory injunction', 'quia timet injunction']
```

Sentence Extraction

Original Text

```
<sentence id="s1">2 CSL carries on business developing, manufacturing, selling and supplying in Australia plasma products, antivenoms and human vaccines.</sentence>
<sentence id="s2">GSKA markets throughout Australia a range of pharmaceutical products including vaccines.</sentence>
<sentence id="s3">3 The products at the centre of this dispute are vaccines used for the prevention of infection caused by Human Papillomavirus ("HPV").</sentence>
```

Extracted Sentences

```
['CSL carries on business developing, manufacturing, selling and supplying in Australia plasma products, antivenoms a
nd human vaccines', 'GSKA markets throughout Australia a range of pharmaceutical products including vaccines']
```

Citations Extraction

Original Text

```
['Dann on behalf of the Amangu People v Western Australia [2006] FCA 1249 (18 September 2006)', 'Harrington-Smith on
behalf of The Wongatha People v State of Western Australia [2002 FCA 184', 'Byron Environment Centre Inc v Arakwal Pe
ople (1997) 78 FCR 1', 'Re An Application for Native Title Determination by the Gunai People (unreported, National Na
tive Title Tribunal, 17 January 1997, French P)', 'Members of the Yorta Yorta Aboriginal Community v State of Victori
a (1996) 1 AILR 482', 'Woodridge v Minister for Land and Water Conservation for the State of New South Wales [2001] F
CA 419 ; (2001) 108 FCR 527', 'Adnyamatha People v South Australia [2003] FCA 1377 ; (2003) 133 FCR 242']
```

Extracted Citations

```
['Dann on behalf of the Amangu People v Western Australia', 'Harrington-Smith on behalf of The Wongatha People v Stat
e of Western Australia [', 'Byron Environment Centre Inc v Arakwal People', 'Re An Application for Native Title Deter
mination by the Gunai People', 'Members of the Yorta Yorta Aboriginal Community v State of Victoria', 'Woodridge v Mi
nister for Land and Water Conservation for the State of New South Wales', 'Adnyamatha People v South Australia']
```

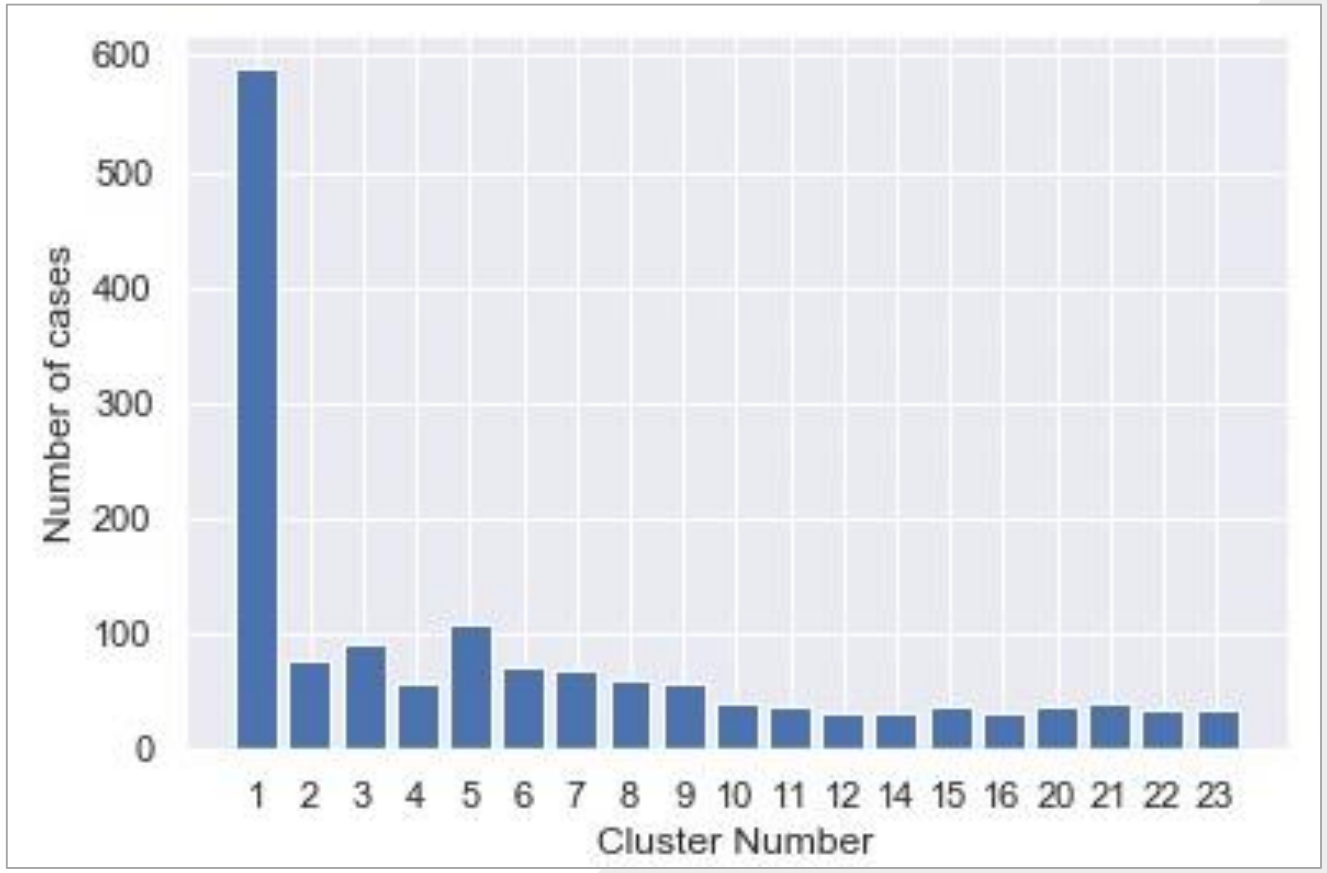
Citations-based Clustering

Directed graph data structure containing extracted citations is generated

Infomap algorithm is run on the graph

Clusters containing the extracted catchphrases and sentences are generated

Clustering



- Out of the total **287** clusters generated, we choose to work with those where the number of cases in each cluster is greater than or equal to 10. We had **48** such clusters.
- The plot here represents the **top 23** clusters, with highest number of cases in them.

Cluster-wise Summarization

Cases in each cluster are merged to form a large body of text.

Summaries are generated for each cluster using Text Rank Algorithm (Gensim package)

Summarization

Original Text

```
[['ON EXTENSION OF RECEIVER ORDERS TO TRUST PROPERTIES   Introduction  1 On 27 March 2006 the Australian Securities and Inve
stments Commission (ASIC) commenced proceedings in this Court seeking the appointment of receivers to the property of certai
n officers and former officers of companies in the Westpoint Property and Finance Group (the first to fourth defendants)',
  'It also sought the appointment of receivers to the property of certain companies in the Group',
  'The application was brought under s 1323 of the Corporations Act 2001 (Cth) (the Act)',
  'On 20 April 2006 orders were made that receivers be appointed to the property of each of the defendants other than Bowesc
o Pty Ltd which already had a receiver appointed under an existing security',
  'Orders were also made for the disclosure by each of the first to fourth defendants of their assets and liabilities, bank
accounts, debtors and property, real and personal',
  'Orders in relation to the fifth to eighth defendants required like disclosure of their assets and liabilities',
  'Similar orders were made later in respect of the ninth defendant',
  'ASIC now seeks to amend the receiver orders in respect of the first, third, fourth and fifth to ninth defendants',
```

Generated Summary

```
[['Constructive or Creative Insolvency?',
  "Part 5.3A of ch 5 of the Corporations Act 2001 (Cth) (the Act) makes provision for inter alia , the administration of a c
ompany's affairs with a view to the execution of a deed of company arrangement after its approval at a meeting of creditor
s",
  "Its origins may be traced to amendments made to the Act's predecessor, the Corporations Law (repealed) by the Corporate L
aw Reform Act 1992 (Cth) (Corporate Law Reform Act 1992)",
  'In the Explanatory Memorandum circulated by the then Attorney-General when introducing the bill which became the Corporat
e Law Reform Act 1992 , it was noted (paras 14, 15 and 21) that the proposed Pt 5.3A would implement recommendations made in
the Law Reform Commission\'s Report No 45 in respect of the Commission\'s General Insolvency Inquiry, popularly known as "th
e Harmer Report"',
  "In the Harmer Report (Volume 1, para 52 and para 53) the following observations are made in relation to the then state of
Australia's corporations law:  Conservative legislation  The Commission is also concerned that, apart from conclusions that
might be suggested by statistical evidence, the legislative approach to corporate insolvency in Australia is most conservati
ve",
  'There is very little emphasis upon or encouragement of a constructive approach to corporate insolvency by, for example, f
ocussing on the possibility of saving a business (as distinct from the company itself) and preserving employment prospects',
  'Creative alternatives to insolvency  Constructive or creative insolvency is not a myth',
  'However, it requires suitable procedures that encourage and offer a reasonable prospect of achieving that result',
  'A constructive approach to corporate insolvency requires the preservation, if practical and possible, of the property and
```

Evaluation

We evaluate the accuracy of the summaries generated, using **Recall-Oriented Understudy for Gisting Evaluation or ROUGE-N** metric. Here we keep in mind the 2 types of entities (summaries) used in the evaluation –

- Generated Summary – machine-generated summary
- Reference Summary – extracted catchphrases for each case

The following evaluation metrics are calculated –

RECALL

Recall with respect to ROUGE score indicates how much of the reference summary is captured by the machine-generated summary.

$$\text{Recall} = \frac{Number\ of\ overlapping\ words}{Total\ words\ in\ reference\ summary}$$

PRECISION

Recall with respect to ROUGE score indicates how much of the generated summary is in fact relevant.

$$\text{Precision} = \frac{Number\ of\ overlapping\ words}{Total\ words\ in\ generated\ summary}$$

Evaluation

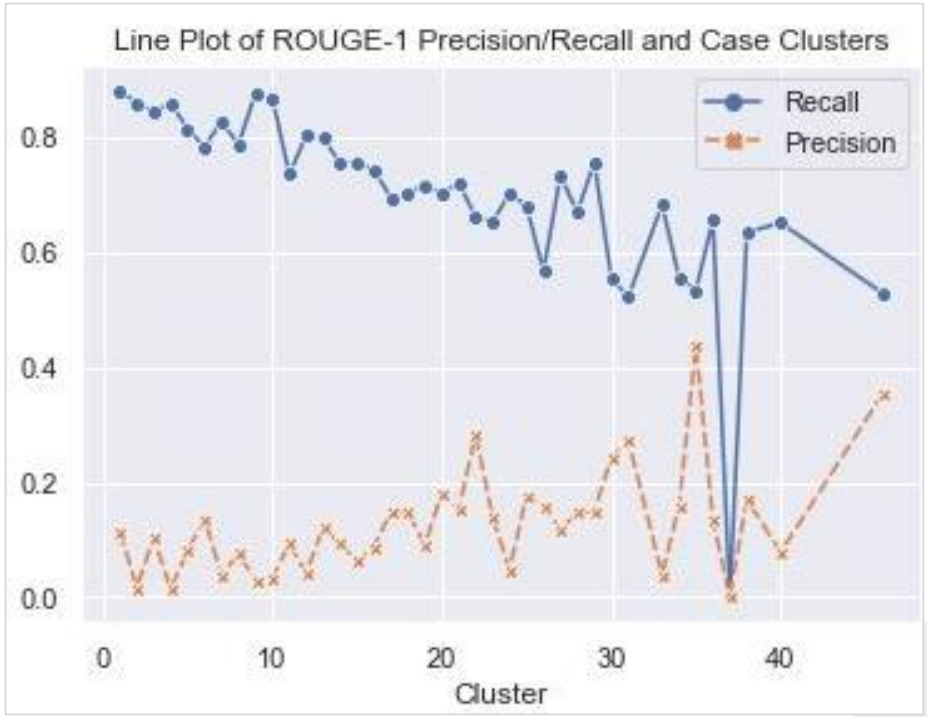


Figure 1: Recall and Precision for ROGUE-1

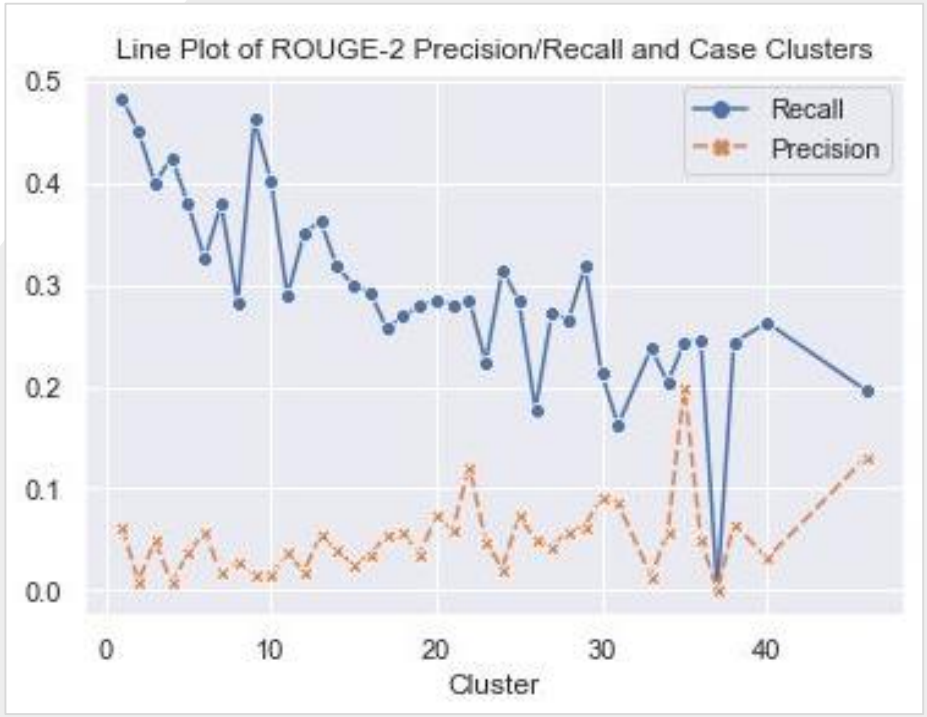


Figure 2: Recall and Precision for ROGUE-2

In both plots, we can observe that the generated summaries capture a good amount of the reference summaries judging from the observed values of recall across all clusters (> 0.5). However, the generated summaries captures some of the text that are not relevant, judging from the observed values of precision across all clusters (< 0.3).

Challenges

- The documents present in some clusters were large and thereby increased the computation time.
- Converting clustered cases to the appropriate data type to fit in the summarizer in the gensim package was a hindrance.
- It was difficult to ascertain the reason why cluster 37, despite having 16 cases, did not generate any summary when fed into the summarizer.

Conclusion

A great deal of work has been done to generate the summaries for each cluster. Judging by an overall recall score greater than 0.5, it can be understood that the extractive summarization techniques, give a good representation of cases based on similar citations. Through the technologies implemented in this project, it can be understood that given citations, we can generate clusters and analyze the texts in each cluster. Hence, similar concepts can be applied to legal cases in any constitution.

Future Work

Legal texts have characteristics different from news articles and other texts, specifically in the vocabulary and ambiguity.

- Given more time, it would be interesting to identify the topics labels present in different clusters of legal case reports by applying the Latent Dirichlet Allocation (LDA) algorithm to generate different topic labels per cluster allowing comparison across clusters to check for similarities or differences.
- There is also a possibility to explore and apply abstractive summarization techniques to generate summaries for the same dataset and compare the results of the 2 summarizations methods.
- Legal jargon can also be extracted for better understanding the language of a case, using keywords generated in through the TextRank algorithm.