Citation Anchor Text for Improving Precedent Retrieval:
An Experimental Study on Indian Legal Documents

Gaurang Patil [a], Bhoomeendra Singh Sisodiya [a],
P. Krishna Reddy [a], K.V.K. Santhy [b]

[a] IIIT, Hyderabad, Telangana, India
[b] NALSAR, Hyderabad, Telangana, India
Email: pkreddy@iiit.ac.in

Presentation Outline

- **Introduction**
- Related Work
- Proposed Framework
- Experimental Results
- Discussion, Summary, and Future Work

Introduction

- Legal practitioners cite precedents while drafting judgments and arguments.
  - to formulate and reinforce their arguments in court

- Precedent retrieval systems
  - Help reduce legal practitioners' workload
  - Increase the productivity of the judicial stakeholders

- Research Issue: Improving precedent retrieval systems

- Challenges: Judgments are
  - Typically long
  - Discuss multiple legal issues
  - Are written in legalese

About Document Representation

- Document representation affects the performance of the precedence retrieval system
- Development of effective legal document representations is challenging
- Existing research in precedent retrieval has leveraged methods like
  - Meta-data extraction [1], catchphrase extraction [2], event extraction [3], and filtering of paragraphs [4,5,6], etc.
  - Exploiting citation networks to estimate the similarity between two judgments [7,8,9]
- In this paper,
  - We introduce the notion of Citation Anchor Text (CAT), i.e., text surrounding citation, to improve the document representation of the referenced judgment for improving precedent retrieval performance

[1] Rabelo J, Kim MY, Goebel R, Yoshioka M, Kano Y, Satoh K. A Summary of the COLIEE 2019 Competition.New Frontiers in Artificial Intelligence. Springer; 2020.
[2] Mandal A, Ghosh K, Ghosh S, Mandal S. Unsupervised approaches for measuring textual similarity between legal court case reports. Artificial Intelligence and Law. 2021
[3] Joshi A, Sharma A, Tanikella SK, Modi A. U-CREAT: Unsupervised Case Retrieval using Events extrAcTion. In: ACL 2023.
[4] Sisodiya BS, et al. Analysing the Resourcefulness of the Paragraph for Precedence Retrieval. In: Proceedings of the 19th ICAIL '23, ACM; 2023.
[5] Makawana M, Mehta RG. A novel network-based paragraph filtering technique for legal document similarity analysis. Artificial Intelligence and Law. 2023 Oct.
[6] Rabelo J, Kim MY, Goebel R. Semantic-based classification of relevant case law. In: JSAI International Symposium on Artificial Intelligence. Springer; 2022.
[7] Kumar S, Reddy PK, Reddy VB, Singh A. Similarity analysis of legal judgments. In: Proceedings of the Fourth Annual ACM COMPUTE '11; 2011.
[8] Minocha A, Singh N, Srivastava A. Finding Relevant Indian Judgments using Dispersion of Citation Network. In: Proc. of the 24th WWW '15 Companion. 2015.
[9] Bhattacharya P, Ghosh K, Pal A, Ghosh S. Hier-SPCNet: A Legal Statute Hierarchy-based Heterogeneous Network for Computing Legal Case Document Similarity. 43rd SIGIR '20. 2020.

Observation about legal citation

- *Reference:* Martin PW. Introduction to basic legal citation. Legal Information Institute; 2006.
  - *"A reference properly written in 'legal citation' strives to do at least three things, within limited space:*
    - *identify the document and document part to which the writer is referring*
    - *provide the reader with sufficient information to find the document or document part in the sources the reader has available (which may or may not be the same sources as those used by the writer), and*
    - ***furnish important additional information about the referenced material and its connection to the writer's argument to assist readers in deciding whether or not to pursue the reference."***
- When the factual circumstances of the prior case and current case are similar, the decision-making process is often alike [1,2]

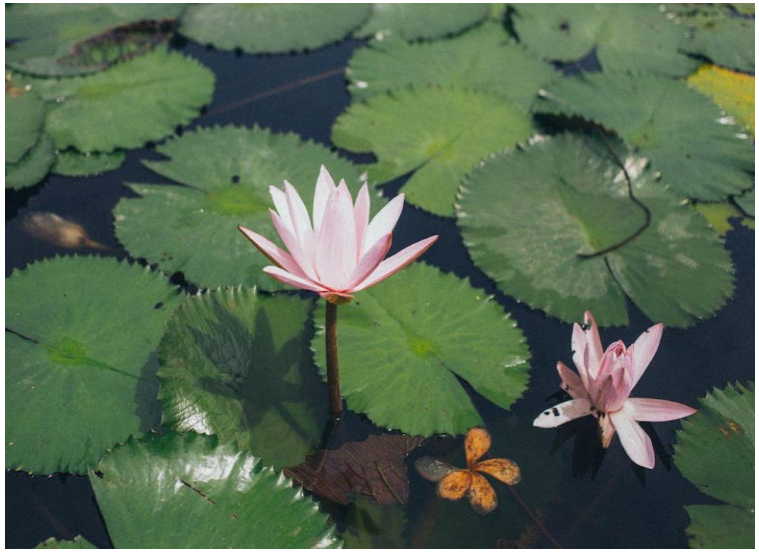
[1] Marshall G. What is binding in a precedent. In: Interpreting precedents. Routledge; 2016. p. 503-17.
[2] Siltala R. A theory of precedent: from analytical positivism to a post-analytical philosophy of law. Hart Publishing; 2000.

About Anchor text in the Web Search

- Notably, in the case of the Web, anchor text associated with a hyperlink
  - Often provides a concise alternative description of the web page it points to [1]
  - Has been exploited for indexing web pages [2], images and videos [3], and has resulted in a significant improvement in the performance of search engines [4]

- Example:

```
<ahref="https://www.pexels.com/photo/pink-water-lily-on-body-of-water-1850535/">pink water lily on a body of water</a>
```



[1] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. 1998;30(1):107-17.
[2] Fujii A. Modeling anchor text and classifying queries to enhance web document retrieval. In: Proceedings of the 17th WWW '08; 2008.
[3] Smith JR, Chang SF. Searching for images and videos on the world-wide web. IEEE multimedia magazine. 1996.
[4] Craswell N, Hawking D, Robertson S. Effective site finding using link anchor information. ACM SIGIR '01; 2001.

Contributions

- We consider that there is an opportunity to exploit the text surrounding a citation to capture information related to the referenced judgment
- We exploited the resourcefulness of the text surrounding a citation to
  - Improve the document representation of the referenced judgment
  - Proposed a Preceding citation Anchor Text (PAT)-based document representation approach
  - We conduct experiments on two datasets related to Indian Supreme Court judgments

Presentation Outline

- Introduction
- **Related Work**
- Proposed Framework
- Experimental Results
- Discussion, Summary and Future Work

Related Work: Citation Networks for Precedence Retrieval

- For computing similarity between judgments, the notions of
  - bibliographic coupling (based on the number of common out-citations) and
  - co-citation (number of common in-citations) between the two judgments were exploited [1]
- Exploited dispersion (a term borrowed from network science) to find relevant judgments [2]
- A Network-based paragraph filtering approach was proposed by exploiting co-citation and co-occurrence for better document representation [3]
- Finding similarity by incorporating the statute hierarchy into the citation network [4]

[1] Kumar S, Reddy PK, Reddy VB, Singh A. Similarity analysis of legal judgments. In: Proceedings of the Fourth Annual ACM COMPUTE '11.; 2011.
[2] Minocha A, Singh N, Srivastava A. Finding Relevant Indian Judgments using Dispersion of Citation Network. I WWW '15 Companion. ; 2015.
[3] Makawana M, Mehta RG. A novel network-based paragraph filtering technique for legal document similarity analysis. Artificial Intelligence and Law. 2023 .
[4] Bhattacharya P, Ghosh K, Pal A, Ghosh S. Hier-SPCNet: A Legal Statute Hierarchy-based Heterogeneous Network for Computing Legal Case Document Similarity. ACM SIGIR '20.; 2020.

Related Work: Document Representation

- Efforts are being made with various document representations [1,2]-
  - e.g., the whole judgment, a summary of the judgment, constituent paragraphs of the judgment, and text surrounding the citations (as a representation of the same judgment)
- Event-based representations for a judgment [3]
- A paragraph-based similarity framework [4]
  - Showed that computing similarity at the paragraph level could help establish relevance better than computing similarity at the document level
  - Unigram-bigram-based TF-IDF was used to obtain vector representations for paragraphs and was found effective
- A combination of textual and network-based similarity methods were proposed [5]

[1] Mandal A, Ghosh K, Ghosh S, Mandal S. Unsupervised approaches for measuring textual similarity between legal court case reports. Artificial Intelligence and Law. 2021 Sep;29(3):417-51.
[2] Mandal A, Chaki R, Saha S, Ghosh K, Pal A, Ghosh S. Measuring Similarity among Legal Court Case Documents. In:10th Annual ACM Compute '17, 2017.
[3] Joshi A, Sharma A, Tanikella SK, Modi A. U-CREAT: Unsupervised Case Retrieval using Events extrAcTion. ACL 2023.
[4] Sisodiya BS, Unnam NB, Reddy PK, Das A, Santhy KVK, Reddy VB. Analysing the Resourcefulness of the Paragraph for Precedence Retrieval. In: ICAIL '23. 2023.
[5] Bhattacharya P, Ghosh K, Pal A, Ghosh S. Legal case document similarity: You need both network and text. Information Processing & Management. 2022;59(6):103069.

Related Work: Anchor Text

- **Early days of Web**: Anchor text Provides a concise alternative representation of the page it points to and often describes the page better than the page itself [1]
  - retrieval using anchor text is better than using the content of the web page [2]
  - Anchor text has been exploited extensively for improving web search [3,4,5]
- **Legal domain:** Efforts have also been made to identify RFC (Reason for Citing) a particular case by processing text near a citation [6]. It has been used, for example,
  - to build a visualization tool to view the citation network [7]
  - for finding documents that are often cited for a given reason [8]
- Incoming citations to a judgment have been exploited to improve the summaries of landmark judgments [9]
- Question-answering systems leveraging reference information have been explored [10]

[1] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. 1998;30(1):107-17.
[2] Craswell N, Hawking D, Robertson S. Effective site finding using link anchor information. In: SIGIR '01; 2001.
[3] Dai N, Davison BD. Mining Anchor Text Trends for Retrieval. In: Advances in Information Retrieval, : Springer; 2010.
[4] Ma Z, Dou Z, Xu W, Zhang X, Jiang H, Cao Z, et al. Pre-training for Ad-hoc Retrieval: Hyperlink is Also You Need. In: CIKM '21, 2021.
[5] Metzler D, Novak J, Cui H, Reddy S. Building enriched document representations using aggregated anchor text. In: SIGIR '09, 2009.
[6] Humphrey TL, et al., inventors; LexisNexis Group, assignee. Automated system and method for generating reasons that a court case is cited. United States Patent US 6,856,988; 2005 Feb 15.
[7] Zhang P, Koppaka L. Semantics-based legal citation network. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law. ICAIL '07; 2007.
[8] Zhang P, Silver HR, inventors; LexisNexis Inc, assignee. Systems and methods for identifying documents based on citation history. United States Patent US 9,201,969; 2015 Dec 1.
[9] Bindal P, Kumar V, Bhatnagar V, Sirohi P, Siwal A. Citation-Based Summarization of Landmark Judgments. In: D Pawar J, Lalitha Devi S, editors. Proceedings of the 20th ICON. Goa, India: (NLPAI); 2023.
[10] Tran OT, Ngo BX, Le Nguyen M, Shimazu A. Answering Legal Questions by Mining Reference Information. In: New Frontiers in Artificial Intelligence. Springer; 2014.

Presentation Outline

- Introduction
- Related Work
- **Proposed Framework**
- Experimental Results
- Discussion, Summary and Future Work

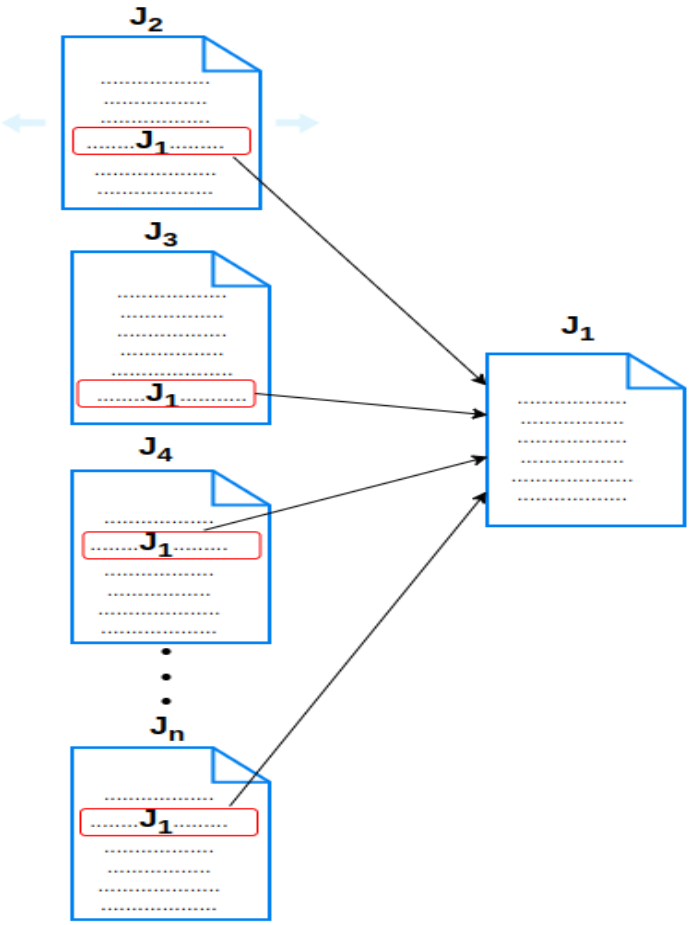
Basic Idea

- Let judgment J_1 be cited by a set C of judgments:

  $C = \{J_2 , J_3 ,. . . , J_n \}$

- The title of J_1 appears in each judgment of C

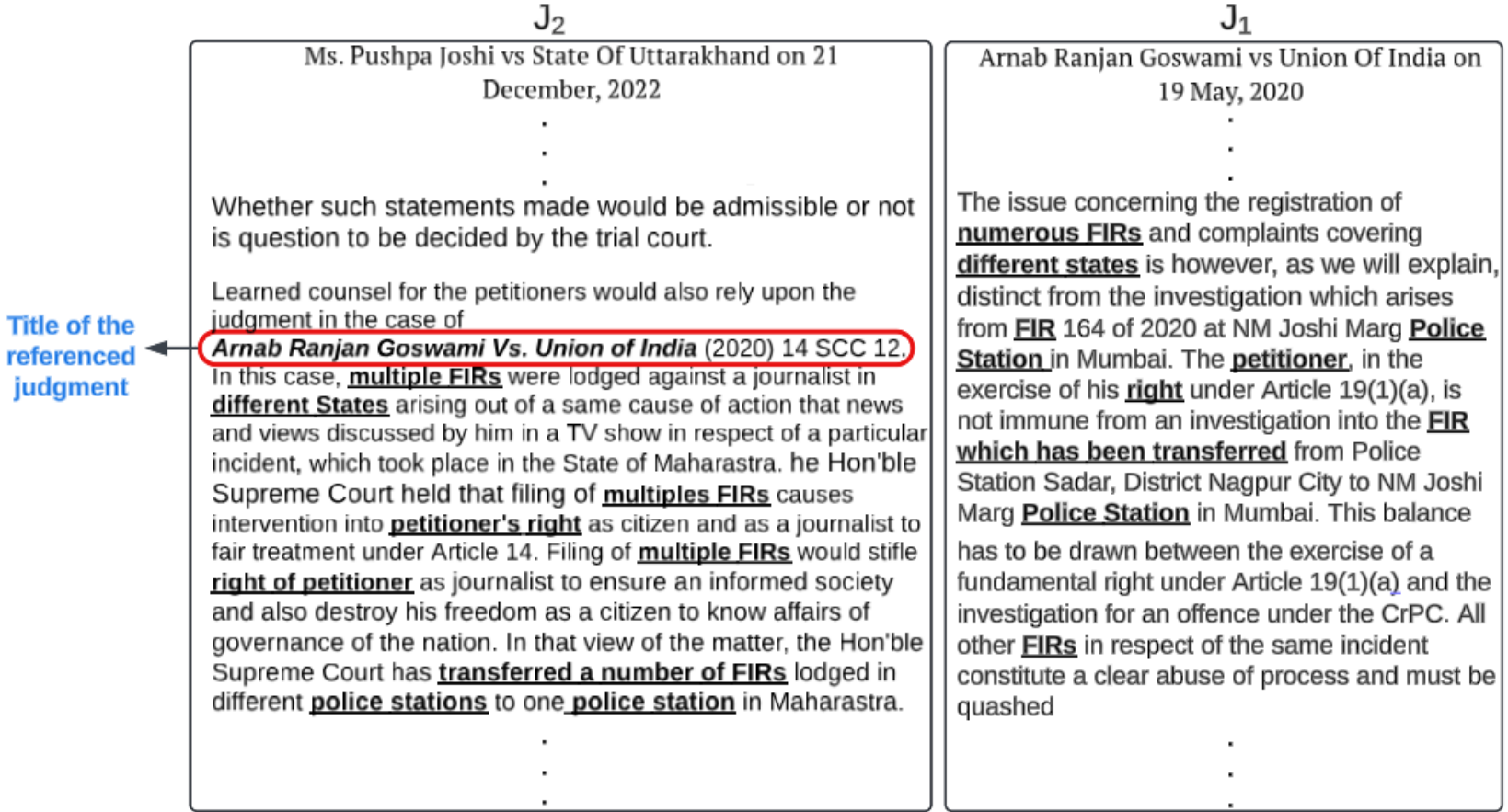
- The set of text units surrounding the title of J_1 in C could be utilized to improve the representation of J_1



Case Citation

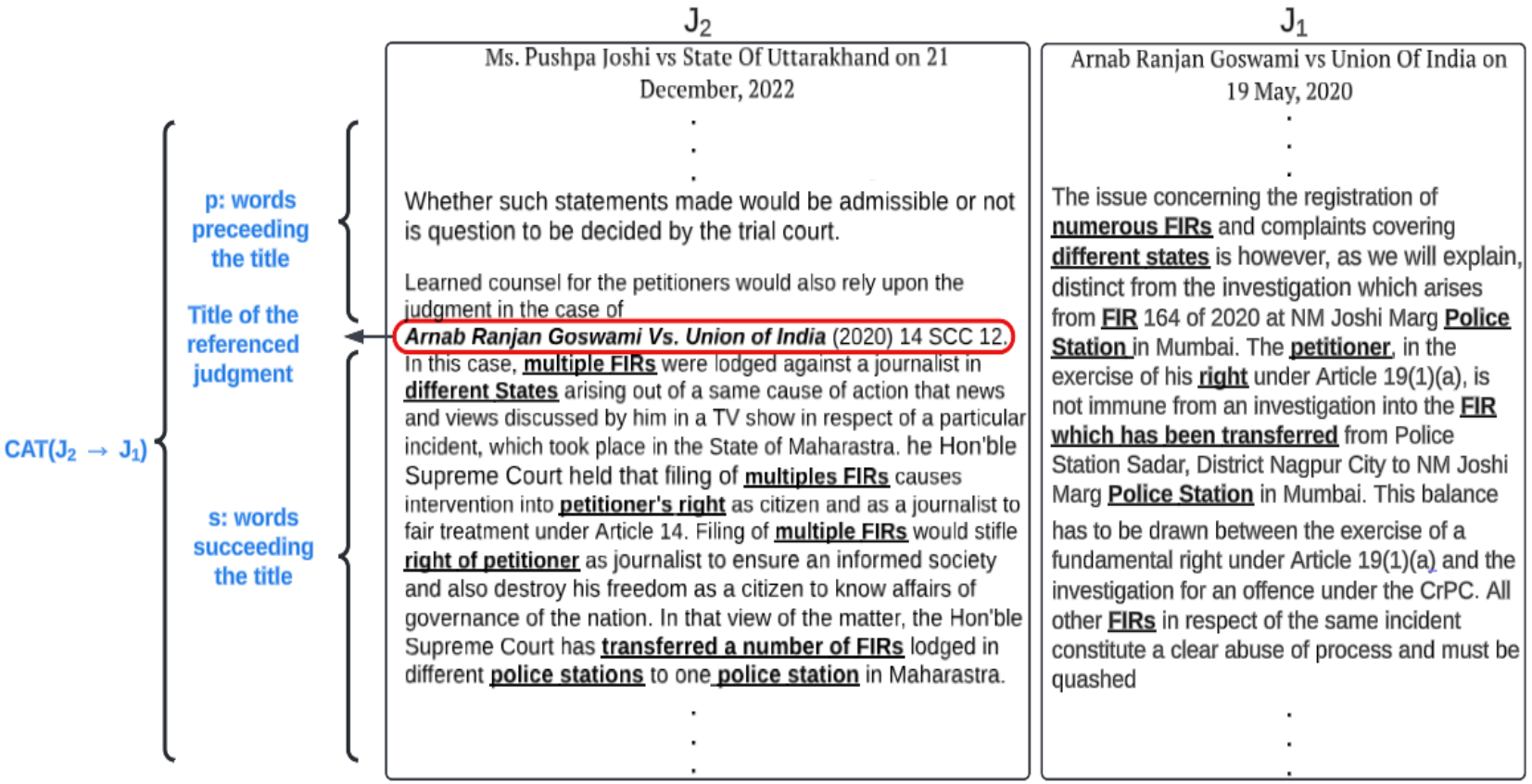
Definition 1. *Case citation. Let J_x and J_y be two judgments, where x and y are integers. We say 'J_y cites J_x' (denoted as $J_y \rightarrow J_x$) if J_y contains a reference to J_x.*

- Example: Judgment J_2 cites J_1 . The title of J_1 can be found in J_2. We denote this case citation as '$J_2 \rightarrow J_1$'



Citation Anchor Text (CAT)

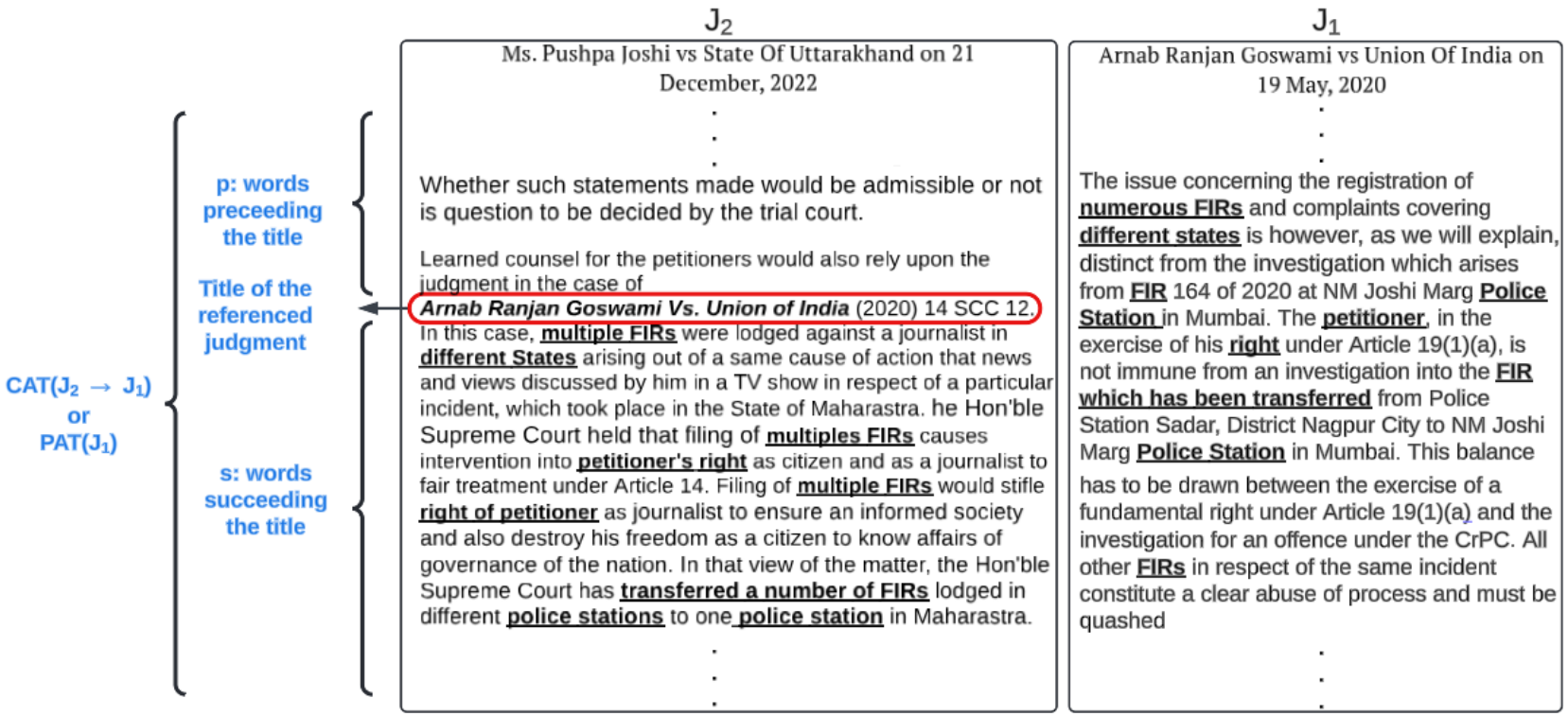
- Given $J_y \rightarrow J_x$, we refer to the text surrounding the title of J_x, which is present in J_y as CAT.
  - Here, CAT consists of the text on both sides of the title and the text in the title itself.



Preceding citation Anchor Text (PAT)

- We define the notion of PAT by associating the referenced judgment with CAT.

Definition 3. *Preceding citation Anchor Text (PAT). Let $J_y \rightarrow J_x$ be a case citation and p and s be integers. Then, $CAT(J_y \rightarrow J_x)$ is referred to as $PAT(J_x)$.*

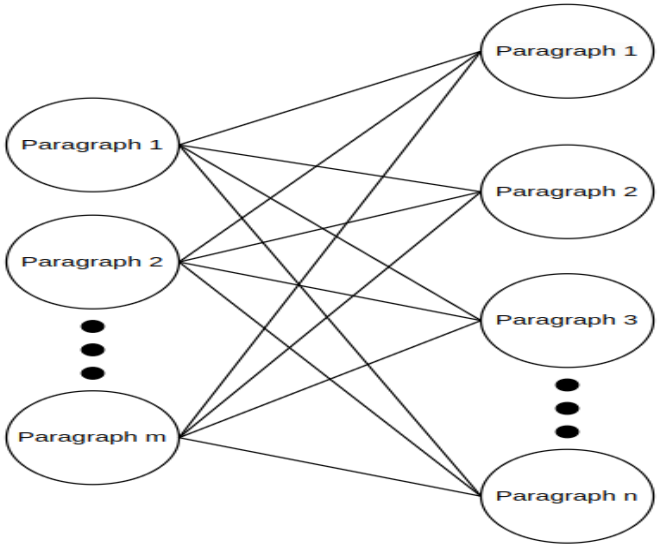


Observation: $PAT(J_1)$ captures certain keywords (bold and underlined) that are related to J_1

Representing a Document

- Vector representations for text can be obtained using techniques like TF-IDF, BM25 (Best Match 25), and other embedding techniques
- We can represent the document as
  - a whole judgment text (JT): The text of J_x is represented as a single vector.
  - a set of judgment paragraphs (JP): J_x is split into a set of its constituent paragraphs, representing each paragraph as a vector.

- Paragraph-based similarity: We compute cosine similarities between all possible pairs of paragraphs in two documents.
  - We obtain a similarity matrix of size m x n
  - We consider the average of top-k values in the matrix as the final similarity score.



Proposed Framework

By considering the notion of Preceding anchor text (PAT), we list the following approaches :

Document Representation	Approach	Description
Concatenated Preceding citation Anchor Text Units	CPATU	• If J_x has n citations, there will be n units of PAT • The set of n units of PAT is called as Preceding citation Anchor Text Units (PATU) • The concatenation of all n units in PATU is called CPATU
Judgment Text + Concatenated Preceding citation Anchor Text Units	JT-CPATU	• We append CPATU of J_x to the text of J_x • Represent the resultant document as a single entity

Proposed Framework….

Document Representation	Approach	Description
Only Preceding citation Anchor Text Units	PATU	• Represent J_x using its PAT units alone. No concatenation is involved.
Combination of Judgment Paragraphs - Preceding citation Anchor Text Units	JP-PATU	• We represent J_x as a union of the paragraphs of J_x and PAT units of J_x
Hybrid	Hybrid	• Let the similarity of query representation w.r.t. JP be s_1 and similarity w.r.t. PATU be s_2 with the weighting factor $\alpha \in [0, 1]$ • The final similarity score: $\alpha \cdot s_1 + (1 - \alpha) \cdot s_2$

Presentation Outline

- Introduction
- Related Work
- Proposed Framework
- **Experimental Results**
- Discussion, Summary and Future Work

Experimental Results

- We conduct experiments on two datasets:
  - FIRE IRLeD 2017 dataset [1]
  - FIRE AILA 2019 dataset [2]
- IRLeD dataset
  - Consists of 2000 prior cases and 200 test queries, along with relevant labels/answers
  - A query comprises a judgment (a.k.a. current case)
  - The task is to find relevant judgments (a.k.a. prior cases) for each query

Table 1. Dataset Details

Dataset	No. of Prior Cases	No. of Test Queries	% of Prior Cases Successfully Mapped	Mean PAT units per Mapped Case
IRLeD-based	2000	200	94%	11.16
AILA-based	2914	40	74.3%	4.35

[1] Mandal A, Ghosh K, Bhattacharya A, Pal A, Ghosh S. Overview of the FIRE 2017 IRLeD Track: Information Retrieval from Legal Documents. In: FIRE (Working Notes); 2017. p. 63-8.
[2] Bhattacharya P, Ghosh K, Ghosh S, Pal A, Mehta P, Bhattacharya A, et al. FIRE 2019 AILA Track: Artificial Intelligence for Legal Assistance. In: Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation. FIRE '19. ; 2019.

Extraction of PAT

- Extracted PAT from the Indian Supreme Court judgment dataset.

- Vectorization: Unigram-bigram-based TF-IDF, since TF-IDF has been proven effective for estimating similarity at the document level [1] as well as paragraph-level [2]
  - TF-IDF is fitted on the dataset corpus
  - Cosine similarity is used for similarity computation

[1] Mandal A, Ghosh K, Ghosh S, Mandal S. Unsupervised approaches for measuring textual similarity between legal court case reports. Artificial Intelligence and Law. 2021 Sep;29(3):417-51.
[2] Sisodiya BS, Unnam NB, Reddy PK, Das A, Santhy KVK, Reddy VB. Analyzing the Resourcefulness of the Paragraph for Precedence Retrieval. In: ICAIL '23. NY, USA: ACM; 2023.

Performance Metrics

- Evaluation is done based on metrics recommended by the creators of the two datasets:
  - Precision@10 (P@10)
    - (# of relevant ones in top-10 retrieved documents) /10
  - Recall@100 (R@100)
    - (# of relevant ones in top-100 retrieved documents) / (total # of relevant documents)
  - **Mean Average Precision (MAP)**
  - Mean Reciprocal Rank (MRR)
  - Binary Preference-based measure (BPREF)
- Both consider MAP as the most important metric

Performance Parameters

- Performance parameters:
  - *k*: Number of paragraph pairs considered for computing similarity score in paragraph-level approach (top-k)
  - α : Weighting factor used to weight similarity scores from JP and PATU in the Hybrid approach

Table 2. Performance Parameters

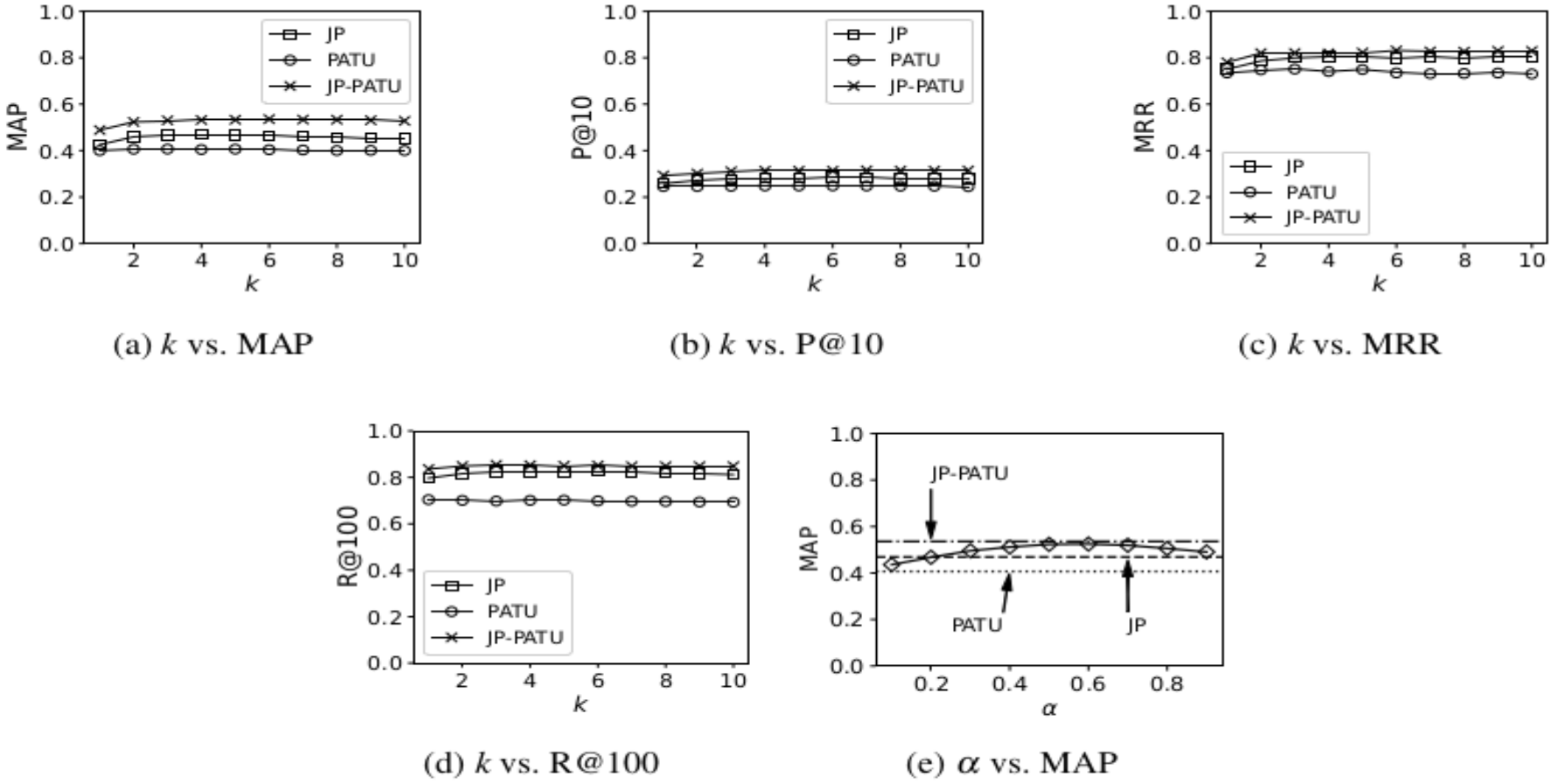
Dataset	Variable	Default	Variation
IRLeD - based	k	$k_{JP} = 4$	1 to 10, with a step of 1
		$k_{PATU} = 2$	1 to 10, with a step of 1
		$k_{JP-PATU} = 6$	1 to 10, with a step of 1
	α	0.7	0.1 to 0.9, with a step of 0.1
AILA - based	k	$k_{JP} = 1$	1 to 10, with a step of 1
		$k_{PATU} = 3$	1 to 10, with a step of 1
		$k_{JP-PATU} = 1$	1 to 10, with a step of 1
	α	0.8	0.1 to 0.9, with a step of 0.1

Experiments on IRLeD-Based Dataset

Table 3. IRLeD based dataset: Performance of document level approaches. The best performance is in bold. Asterisk (*) indicates a statistically significant improvement w.r.t. baseline JT (t-test, p-value < 0.05)

Document Representation	MAP	P@10	MRR	R@100
JT (Baseline)	0.3237	0.2105	0.6285	0.7500
CPATU	0.3031	0.1960	0.6100	0.6690
JT-CPATU	**0.3673***	**0.2330***	**0.6763***	**0.7870***

- Even though CPATU exhibits a relatively lower MAP than JT, the MAP of CPATU is reasonably higher. This is a notable result as the judgment is represented by only its PAT units.
- Improvements obtained by JT-CPATU across all metrics are statistically significant w.r.t. baseline JT.
  - This indicates that JT-CPATU captures certain nuances that are not captured by JT
  - **We attribute this result to the resourcefulness of PAT,** which might be capturing certain information (mainly keywords) about the situations or circumstances in which J_1 is typically cited
  - **This result concretely exhibits that PAT of a judgment can be used as a resource to represent the document for better precedent retrieval**



(a) k vs. MAP (b) k vs. P@10 (c) k vs. MRR

(d) k vs. R@100 (e) α vs. MAP

Figure 2. IRLeD-based Dataset: Performance of paragraph-level approaches

- PATU exhibits reasonable performance
- JP-PATU exhibits significantly better performance than JP.
- Fig. 2(e) shows the results of α versus MAP for the Hybrid approach.
  - We also indicated the corresponding MAP for JP with k = 4 and PATU with k = 2.
  - Highest MAP is obtained at α = 0.7

Table 4. IRLeD based dataset: Performance of paragraph-level approaches. The best performance is in bold. Asterisk (*) indicates a statistically significant improvement w.r.t. JP (t-test, p-value < 0.05)

Doc. Representation	Performance Parameter	MAP	P@10	MRR	R@100
JP (Baseline)	$k = 4$	0.4687	0.2815	0.8054	0.8230
PATU	$k = 2$	0.4099	0.2485	0.7520	0.7080
Hybrid	$\alpha = 0.7$	0.5312*	0.3115*	**0.8321**	0.8470*
JP-PATU	$k = 6$	**0.5358***	**0.3135***	0.8313	**0.8510***

- Paragraph-level approaches exhibit better performance compared to document-level approaches

Experiments on AILA-based dataset

Table 5. AILA based dataset: Performance of document level approaches. The best performance is in bold. Asterisk (*) indicates a statistically significant improvement w.r.t. baseline JT (t-test, p-value < 0.05)

Document Representation	MAP	P@10	MRR	BPREF
JT (Baseline)	0.1473	0.0675	0.2301	0.1205
CPATU	0.1096	**0.0725**	**0.2381**	0.0897
JT-CPATU	**0.1566**	**0.0725**	0.2253	**0.1280**

- The performance trend is similar to the results obtained on IRLeD.
- CPATU captures certain information that can help establish relevance, although its performance is not as good as JT.
- The performance of JT-CPATU is superior to the other two, although the improvement in MAP is not statistically significant
- A possible explanation for this is that we could successfully map only 74.3% of the prior cases and had no means to determine PAT for the rest of the prior cases

Table 6. AILA based dataset: Performance of paragraph-level approaches. The best performance is in bold. Asterisk (*) indicates a statistically significant improvement w.r.t. JP (t-test, p-value < 0.05)

Doc. Representation	Performance Parameter	MAP	P@10	MRR	BPREF
JP (Baseline)	$k = 1$	0.1536	0.0525	0.2891	0.1279
PATU	$k = 3$	0.1207	**0.0775**	0.2988	0.1040
JP-PATU	$k = 1$	0.1555	0.0550	0.2924	0.1315
Hybrid	$\alpha = 0.8$	**0.1644***	0.0600	**0.3275**	**0.1351**

- Similar to the results obtained for IRLeD, improvement in MAP obtained using Hybrid is statistically significant (t-test, p-value< 0.05) over JP

Comparison with Other Approaches

- The proposed approaches exhibit better MAP than the existing approaches
- Note: Experiments with the proposed approach have complimented PAT to IRLeD and AILA datasets
-

Table 7. Comparison with other approaches on IRLeD dataset.

Method	MAP	P@10	MRR	R@100
WordNet + RNN[35]	0.4770	0.2600	0.8010	0.7890
flt_ielab_idf [32]	0.390	0.236	0.719	0.781
PL-F [4]	0.4686	0.2820	0.8054	0.8230
Our Approach (JP-PATU)	0.5358	0.3135	0.8313	0.8510

Table 8. Comparison with other approaches on AILA dataset.

Method	MAP	P@10	MRR	BPREF
HLJIT2019-AILA [36]	0.1492	0.07	0.288	0.1286
PL-F [4]	0.1195	0.05	0.2165	0.0964
Our Approach (Hybrid)	0.1644	0.06	0.3275	0.1351

[4] [3] Sisodiya BS, Unnam NB, Reddy PK, Das A, Santhy KVK, Reddy VB. Analysing the Resourcefulness of the Paragraph for Precedence Retrieval. In: ICAIL '23. NY, USA: ACM; 2023.
[32] Mandal A, Ghosh K, Bhattacharya A, Pal A, Ghosh S. Overview of the FIRE 2017 IRLeD Track: Information Retrieval from Legal Documents. In: FIRE (Working Notes); 2017. p. 63-8.
[35] Padigi SV, Mayank M, Natarajan S. Precedent case retrieval using wordnet and deep recurrent neural networks. In: CS & IT Conference Proceedings. vol. 9, no. 16). CS & IT Conference Proceedings; 2019.
[36] Zhao Z, Ning H, Liu L, Huang C, Kong L, Han Y, et al. FIRE2019@AILA: Legal Information RetrievalUsing Improved BM25. In: FIRE (Working Notes); 2019. p. 40-5.

Discussion

- Dataset size:
  - **Experiments on large datasets would be more reliable**
- Issue of mapping:
  - We could map a limited number of prior cases to their counterparts in Indian Kanoon for extraction of PAT. This may have negatively impacted the results on the AILA dataset
- Sparsity:
  - One may argue that since citation networks are sparse, extracting PAT could impact the representation of a limited number of judgments
  - Judgements are cited by other documents
    - ❑ High Court/District Court decisions
    - ❑ Compendiums
    - ❑ Books
    - ❑ Inquiry reports
    - ❑ Council arguments
  - Indicates the presence of additional resources to extract Preceding Anchor Text.

Summary and Future Work

- Analyzed the resourcefulness of the Preceding citation Anchor Text (PAT)
- The results show that the PAT-based approaches performed better than the baseline and related approaches
- We conclude that PAT is resourceful and could be exploited to improve precedent retrieval
- **Future Work: We plan to leverage PAT to assist models in summarizing judgments**

About Indian Situation

- **“‘A Lifelong Nightmare’: Seeking Justice in India’s Overwhelmed Courts with 50 million criminal and civil cases pending, it would take 300 years to clear the country’s judicial backlog.”**
- New York Times, Jan. 13, 2024
- https://www.nytimes.com/2024/01/13/world/asia/india-judicial-backlog.html

Thank You!