

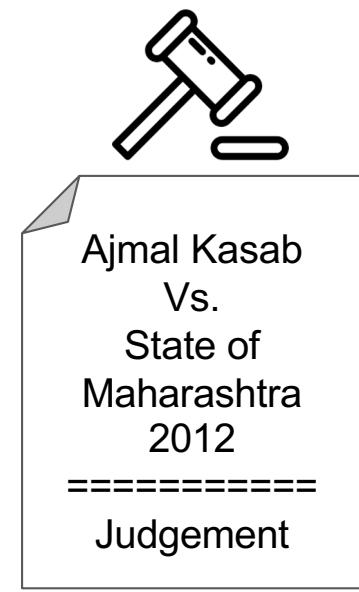
# Part 2

## State-of-the-art AI & Law research, datasets, benchmarks and tools

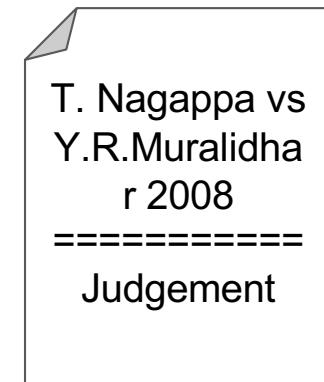
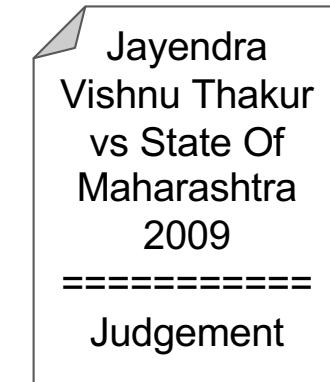
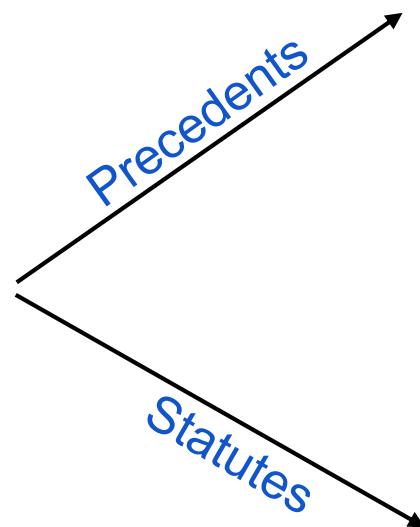
# India: a Common Law system

- Two broad types of legal systems
  - Civil Law
  - Common Law
- About 1/3<sup>rd</sup> of the world's population lives in **common law** jurisdictions or in systems mixed with civil law → India, USA, UK, Canada, Australia ...
- Common Law (principle of Stare decisis)
  - Similar facts or similar situations should yield similar results
  - Two sources of Law in such a system

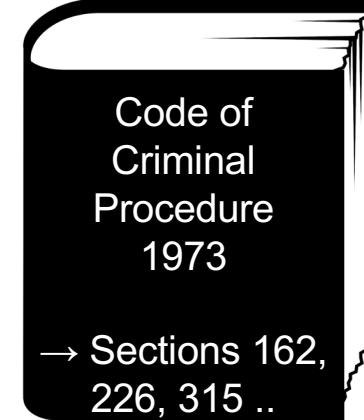
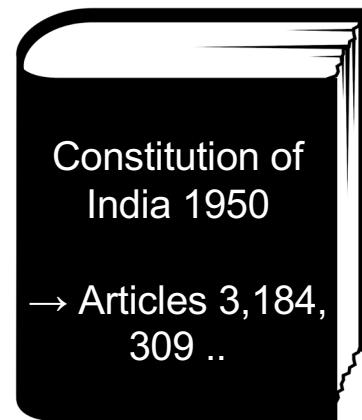
# Two sources of law in a Common Law system



Verdict of a case depends on



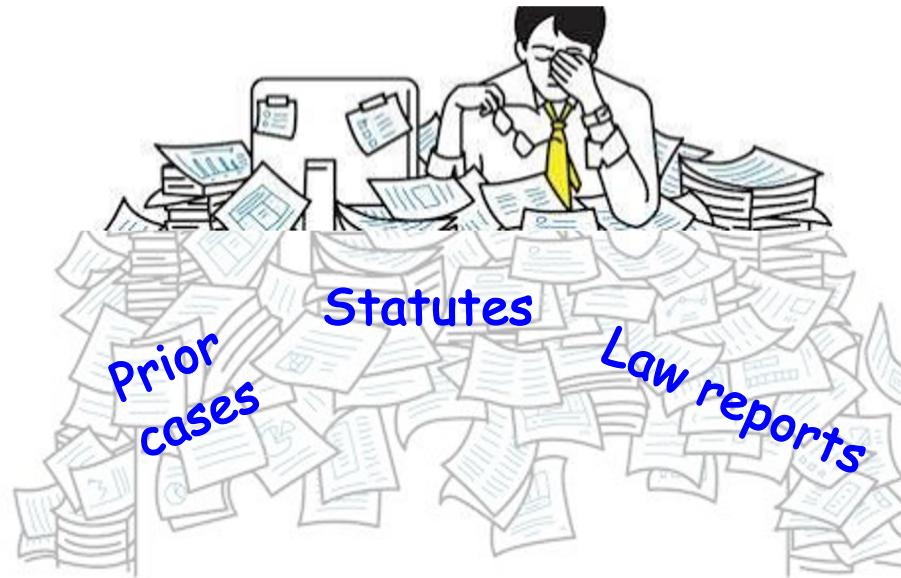
Cases decided in the past by important courts



Written laws

# Implications of the Common Law System

Law practitioners have to look up a huge number of documents (prior cases, statutes, law reports, etc.) that match a given situation or an ongoing case



# Many types of legal documents

- Statutes, e.g., Acts, Sections
- First Information Reports (FIR)
- Docs generated during a case, intermediate orders, ...
- Case judgements (generated at the end of a case)
- Contracts, MoUs
- Patents
- ... any many more

# Many types of legal documents

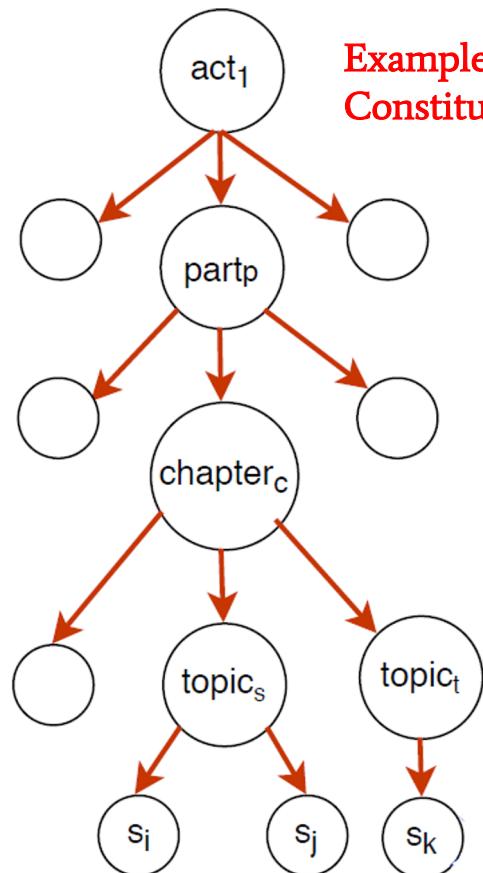
- **Statutes, e.g., Acts, Sections**
- First Information Reports (FIR)
- Docs generated during a case, intermediate orders, ...
- **Case judgements** (generated at the end of a case)
  
- Contracts, MoUs
- Patents
- ... any many more

Focus of this part

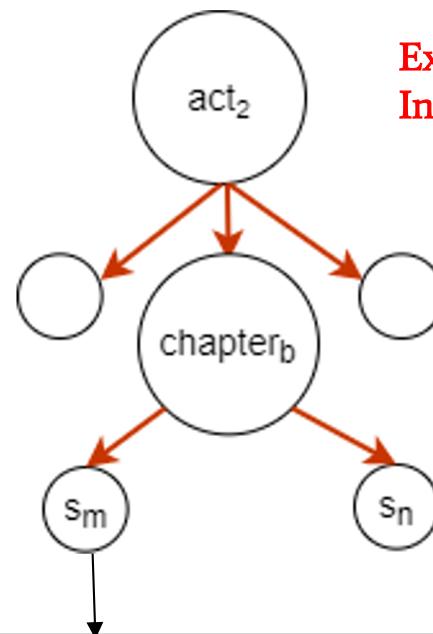
\* Most Legal NLP/IR research is  
on these type of documents

\* Easily available publicly

## Examples of Statutes in the Indian Jurisdiction



Example of an Act:  
Constitution of India, 1950



Example of an Act:  
Indian Penal Code, 1860

Section 302 of the Indian Penal Code, 1860  
Title: Punishment for murder  
Description: Whoever commits murder shall be punished with death, or imprisonment for life, and shall also be liable to fine.

# An example of a court judgement

<https://indiankanoon.org/doc/61901185/>

Decided by the Supreme Court of India on 3 May 2013

Metadata and the main judgement text

Main text usually contains

- Facts of the case
- Arguments by the two contending parties
- Citations to relevant statutes and precedents
- Reasoning by the Court
- Final judgement given by the Court

Ankush Shivaji Gaikwad vs State Of Maharashtra on 3 May, 2013

Author: .....J.

Bench: T.S. Thakur, Gyan Sudha Misra

REPORTABLE

IN THE SUPREME COURT OF INDIA  
CRIMINAL APPELLATE JURISDICTION  
CRIMINAL APPEAL NO. OF 2013  
(Arising out of S.L.P. (Crl.) No.6287 of 2011)

Ankush Shivaji Gaikwad

...Appellant

Versus

State of Maharashtra

...Respondent

JUDGMENT

T.S. THAKUR, J.

1. Leave granted.

2. This appeal arises out of a judgement and order dated 24th August, 2010 passed by the High Court of Judicature at Bombay, Aurangabad Bench, whereby Criminal Appeal No.359 of 2008 filed by the appellant and two others has been dismissed in so far as the appellant is concerned and allowed qua the remaining two, thereby upholding the appellant's conviction for the offence of murder punishable under Section 302 of the I.P.C and the sentence of imprisonment for life with a fine of Rs.2,000/- awarded to him. In default of payment of fine the appellant has been sentenced to undergo a further imprisonment for a period of three months.

3. The factual matrix in which the appellant came to be prosecuted and convicted has been set out in detail by the trial Court as also the High Court in the orders passed by them. We need not, therefore,

# An example of a court judgement

<https://indiankanoon.org/doc/61901185/>

Decided by the Supreme Court of India on 3 May 2013

Metadata and the main judgement text

Main text usually contains

- Facts of the case
- Arguments by the two contending parties
- **Citations to relevant statutes and precedents**
- Reasoning by the Court
- Final judgement given by the Court

Ankush Shivaji Gaikwad vs State Of Maharashtra on 3 May, 2013

Author: .....J.

Bench: T.S. Thakur, Gyan Sudha Misra

REPORTABLE

IN THE SUPREME COURT OF INDIA  
CRIMINAL APPELLATE JURISDICTION  
CRIMINAL APPEAL NO. OF 2013  
(Arising out of S.L.P. (Crl.) No.6287 of 2011)

Ankush Shivaji Gaikwad

...Appellant

Versus

State of Maharashtra

...Respondent

JUDGMENT

T.S. THAKUR, J.

1. Leave granted.

2. This appeal arises out of a judgement and order dated 24th August, 2010 passed by the High Court of Judicature at Bombay, Aurangabad Bench, whereby Criminal Appeal No.359 of 2008 filed by the appellant and two others has been dismissed in so far as the appellant is concerned and allowed qua the remaining two, thereby upholding the appellant's conviction for the offence of murder punishable under Section 302 of the I.P.C and the sentence of imprisonment for life with a fine of Rs.2,000/- awarded to him. In default of payment of fine the appellant has been sentenced to undergo a further imprisonment for a period of three months.

3. The factual matrix in which the appellant came to be prosecuted and convicted has been set out in detail by the trial Court as also the High Court in the orders passed by them. We need not, therefore,

# A part of the same judgement

13. Reference may also be made to the decision in **Mahesh v. State of MP (1996) 10 SCC 668**, where the appellant had assaulted the deceased in a sudden fight and after giving him one blow he had not caused any further injury to the deceased which fact situation was held by this Court to be sufficient to bring the case under Exception 4 to **Section 300 of the IPC**. This Court held:

“...Thus, placed as the appellant and the deceased were at the time of the occurrence, it appears to us that the appellant assaulted the deceased in that sudden fight and after giving him one blow took to his heels. He did not cause any other injury to the deceased and therefore it cannot be said that he acted in any cruel or unusual manner. Admittedly, he did not assault PW-2 or PW-6 who were also present also with the deceased and who had also requested the appellant not to allow his cattle to graze in the field of PW-1. This fortifies our belief that the assault on the deceased was made during a sudden quarrel without any premeditation. In this fact situation, we are of the opinion that Exception-4 to **Section 300 IPC** is clearly attracted to the case of the appellant and the offence of which the appellant can be said to be guilty would squarely fall under **Section 304 (Part- I) IPC...**”  
(emphasis supplied)

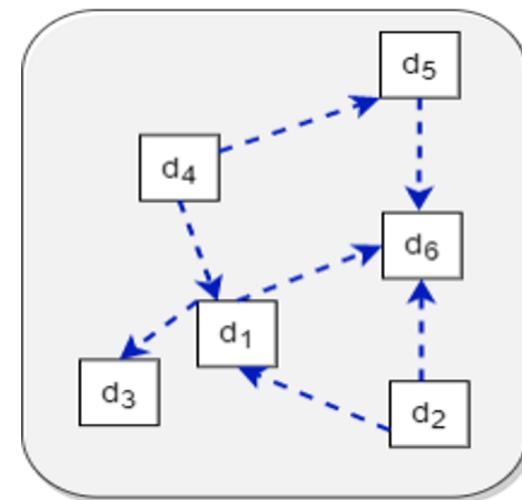
14. To the same effect are the decisions of this Court in **Vadla Chandraiah v. State of Andhra Pradesh (2006) 14 SCALE 108**, and **Shankar Diwal Wadu v. State of Maharashtra (2007) 12 SCC 518**.

**Citations to prior cases (precedents)**

**Citations to statutes**

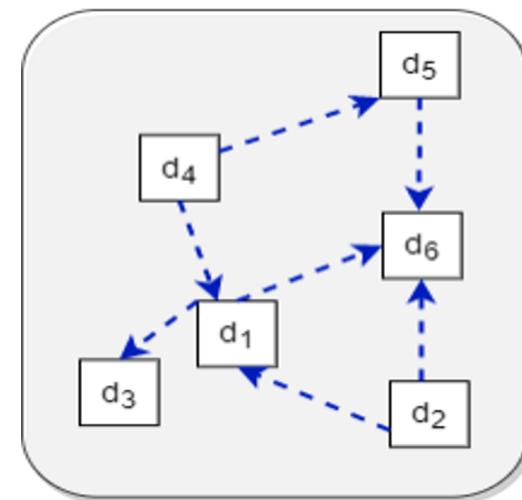
## Network representation of legal data

- Precedent Citation Network (PCNet)
  - Nodes: Case judgements
  - Edges: Citation links (case-case)
- Several prior analyses of legal PCNet

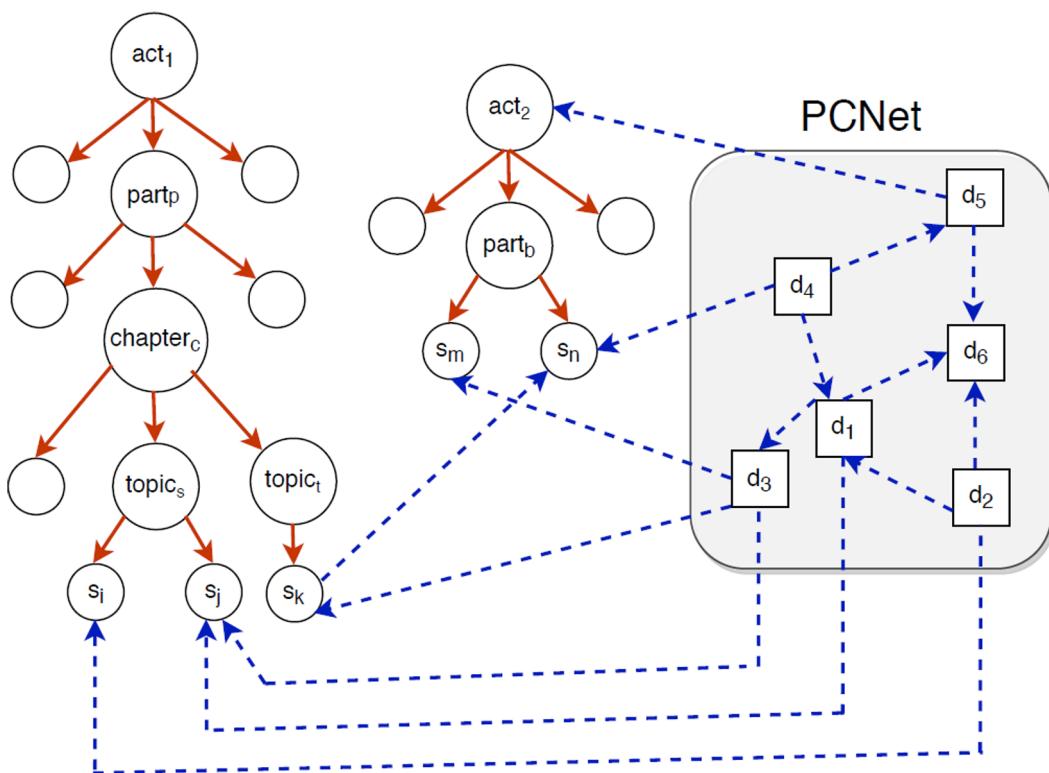


## Network representation of legal data

- Precedent Citation Network (PCNet)
  - Nodes: Case judgements
  - Edges: Citation links (case-case)
- Several prior analyses of legal PCNet
- **PCNet misses important signals coming from statute citations**



## Hier-SPCNet (Hierarchical Statute + Precedent Citation Network)

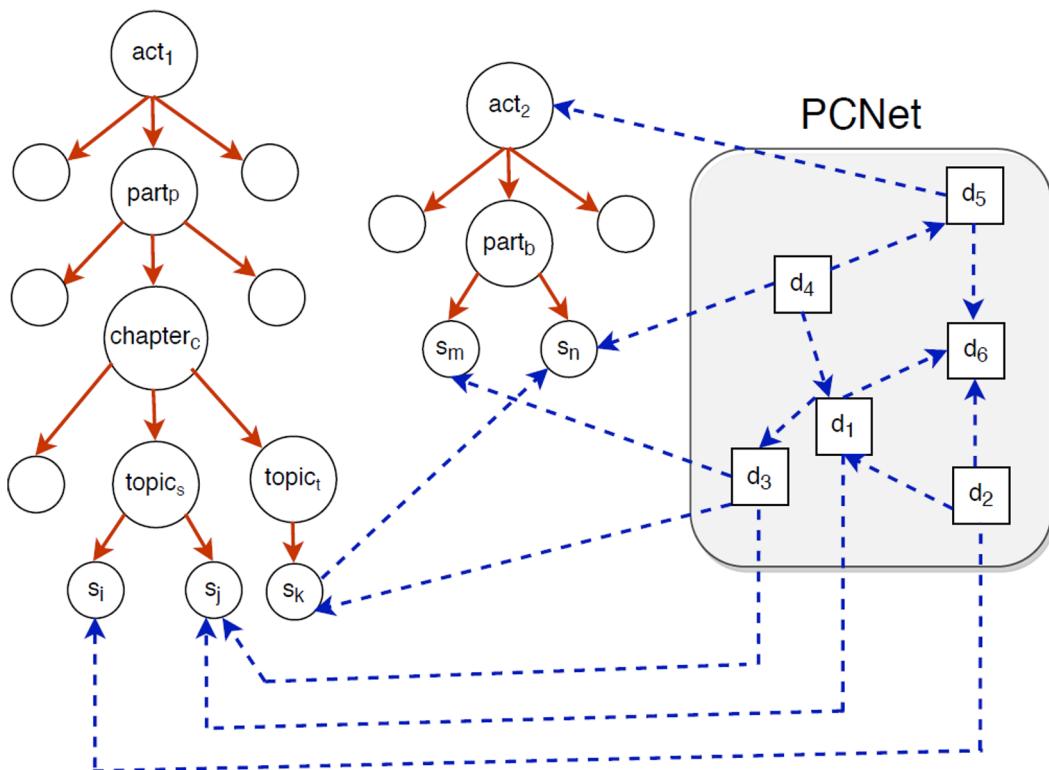


Heterogeneous network

- Nodes: Case judgements, statutes (Acts, sections, ... )
- Edges: Citation links (case-case, case-statute, statute-statute), hierarchy links between parts of statutes

—→ Hierarchy links  
- - - → Citation links

# Hier-SPCNet for estimating similarity between legal documents



Hier-SPCNet enables much better estimation of similarity between legal documents, compared to PCNet

Details:

*Hier-SPCNet: A Legal Statute Hierarchy-based Heterogeneous Network for Computing Legal Document Similarity, SIGIR 2020*

*Legal Case Document Similarity: You Need Both Network and Text. Information Processing and Management, Elsevier, 2022*

→ Hierarchy links  
→ Citation links

# Challenges in processing legal documents

# Challenges in processing legal documents

- Very long documents
  - Most NLP methods trained over news articles, research papers, ...
  - Legal documents often longer than all of these (often tens to hundreds of pages)
  - Most recent neural models cannot take a full legal document as input

# Challenges in processing legal documents

- Lengthy and complex sentences
- Non-standard ways of writing, frequent use of abbreviations
  - Even identifying the end of sentences is challenging [Sentence boundary detection in legal text, Workshop in Natural Legal Language Processing (with NAACL 2019)]

# Challenges in processing legal documents

- Lengthy and complex sentences
- Non-standard ways of writing, frequent use of abbreviations
  - Even identifying the end of sentences is challenging [Sentence boundary detection in legal text, Workshop in Natural Legal Language Processing (with NAACL 2019)]

This Court in Indian Petrochemicals Corporation **Ltd. & Anr. vs.** Shramik Sena & **Ors.**, 1999(6), referred to the decisions in Parimal Chandra Rahas case, Reserve Bank of India **vs.** Workmen, 1996, and **M.M.R. Khan vs.** Union of India, 1990, and held that the workmen of a statutory canteen, as in the present case, would be workmen of an establishment for the purposes of the Act only and not for other purposes.

[71 words]

# Challenges in processing legal documents

- As a result, most standard NLP tools do not work well on legal text
- E.g., standard Named Entity Recognizers (Stanford / Spacy) often incorrect

“Life Insurance Corporation, India”

Life : PER, Insurance Corporation : ORG, India : LOC

Example using  
Stanford NER

On the basis of the said sale certificate the mutation Entry No.1836 was effected in the village record in favour of Gulabai Desai PERSON , and thereby her name was entered in Survey Nos.118/1B and 328 of village Lonavala PERSON to the extent of 29.30 Ares and 70 Ares respectively. Thereafter, Gulabai ORG sold CTS ORG No.133 admeasuring 33 Gunthas on 24.4.1977 DATE to Respondent No.3 Genu Kadu PERSON . The said Gulabai ORG also gifted her remaining area from this Survey numbers to her grandson Anil Gajanan Desai PERSON on 15.1.1979 DATE , who in turn has sold his properties to Respondent no.2 Prem Hasmatraj Lalwani in the year 1980 DATE .

Example using  
Spacy NER

missed

## Challenges in processing legal documents

- Writing style and document structure varies widely from one country to another
- Let us see some examples of case judgements from various countries

# Example Court Case Document

United Kingdom Supreme Court

The Manchester Ship Canal Company Ltd and another (Respondents) v United Utilities Water Plc (Appellant)

2014

## LORD SUMPTION (with whom Lord Clarke and Lord Hughes agree)

### *Introduction*

1. The question at issue on this appeal is whether a sewerage undertaker under the Water Industry Act 1991 has a statutory right to discharge surface water and treated effluent into private watercourses such as the Respondents' canals without the consent of their owners.
2. Discharge into a private watercourse is an entry on the owner's land, and as such is an unlawful trespass unless it is authorised by statute. It is common ground that no express statutory right is conferred by the Water Industry Act. The question is therefore whether it should be implied. A statutory right to commit what would otherwise be a tort may of course be implied. But since this necessarily involves an interference with the rights of others, the test has always been restrictive. The implication must be more than convenient or reasonable. It must be necessary. As a general rule, this will involve showing either that the existence of the power is necessarily implicit in the express terms of the statute, or else that the statutory purpose cannot be effectually achieved without the implication. In particular a right to commit what would otherwise be a tort may be implied if a statutory power is incapable of being exercised or a statutory duty is incapable of being performed without doing the act in question: *Manchester Corporation v Farnworth* [1930] AC 171, 183 (Viscount Dunedin), *Allen v Gulf Oil* [1981] AC 1001, 1013 (Lord Wilberforce).

### *The law before 1991*

3. It has been said that a court "should not routinely investigate the statutory predecessors of provisions in a consolidation statute": *R v Secretary of State for the Environment, Transport and the Regions, Ex p Spath Holme Ltd* [2001] 2 AC 349, 388 (Lord Bingham of Cornhill). This is not so much a rule of construction as a valuable warning against the over-ready assumption that a consolidating Act means exactly the same as the enactments which it replaces. There are, however, cases where a consolidating Act cannot be understood without reference to the state of the law as it was when it was enacted. This is one of them.

At para 201, Lord Rodger suggested that the test could well be expressed thus:

"would the consequences of applying the statutory provision retroactively, or so as to affect vested rights or pending proceedings, be 'so unfair' that Parliament could not have intended it to be applied in these ways? In answering that question, a court would rightly have regard to the way the courts have applied the criterion of fairness when embodied in the various presumptions."

### *The first question*

57. So far as the first question is concerned, Mr Karas QC, on behalf of United Utilities, a sewerage undertaking, relied primarily to support his case for a positive answer, on the provisions of section 159 of the 1991 Act. I would reject that case and there is nothing I can usefully add to what Lord Sumption and Lord Toulson say in paras 13-15 and 26-28 of their respective judgments. At least in relation to sewers laid after the 1991 Act came into force, United Utilities' argument is not supported by the language of section 159 or any other provision of the 1991 Act. It is inconsistent with some other provisions of the 1991 Act, and it is not supported by any practical considerations (although it is fair to add that it is not undermined by any practical considerations either). The reasoning of all three members of the Court of Appeal in *British Waterways Board v Severn Trent Water Ltd* [2002] Ch 25, summarised in para 14 above, appears to me to be unanswerable.

### *The second question*

58. The question whether sewerage undertakers can claim any rights in respect of any outfalls must ultimately turn on the 1991 Act, but in my view, the issue should be addressed by first identifying the water authorities' rights in respect of outfalls from public sewers immediately before the 1989 Act came into force. Mr Karas's argument is that it is a necessary inference

# Example Court Case Document

Federal Court of Canada

Parkdale Hotel Ltd. v. Canada (Attorney General), [1986] 2 FC 514, 1986 CanLII 3940 (FC)

1986

## **Background**

[2] At the trial, the Court was informed that the action instituted by the plaintiff company was the result of a long-standing grievance of the company's owner who had publicly expressed his strong disapproval with s. 67 of the *Canada Elections Act*. Section 67 reads as follows:

67. Everyone is guilty of an offence against this Act who at any time during the hours that the polls are open on the ordinary polling day sells, gives, offers or provides any fermented or spirituous liquor at any hotel, tavern, shop or other public place within an electoral district where a poll is being held.

[3] The company's owner, in pursuing his grievance, had finally instituted action before this Court for declaratory relief. It is unfortunate that by reason of his untimely demise in December of 1984, the owner was unable to have his day in court when the trial date was reached. Nevertheless, his surviving widow, Marjorie Frimeth, took up her late husband's cudgels. As the beneficiary of her husband's estate and as a director and shareholder of the plaintiff company, she instructed her counsel to continue with the action. Such is, in my view, a commendable decision.

## **Preliminary ruling**

[4] At the opening of the trial of the action, counsel for the plaintiff filed a motion to have Marjorie Frimeth added as a party-plaintiff and to have the pleadings amended accordingly. I should grant the motion. It enables all the issues to be traversed without being bogged down by questions of status under either the *Canadian Bill of Rights* or the *Canadian Charter of Rights and Freedoms*.

[5] Status, in my view, has two elements. One deals with the interest of any particular person, corporate or physical, in the statutory provision impugned. The other, equally important, is whether a corporate person, as against a physical person, is entitled to the protection of some or any of the rights secured under the *Canadian Bill of Rights* or the *Canadian Charter of Rights and Freedoms*.

## **Conclusion**

[69] Counsel before me have both contributed to an interesting and stimulating debate. The plaintiffs, in spite of a valiant and indeed spirited attempt have not succeeded before this court in knocking down s. 67 of the *Canada Elections Act*. Notwithstanding what appears to them and to many others as an obsolete statutory provision, I have been unable to find any judicial grounds on which to interfere.

[70] Perhaps s. 67 is deadwood. I should venture to suggest, however, that it is neither the duty nor the privilege of courts to remove deadwood from Canadian statutes. The limits to judicial review are to inquire into the constitutional validity of legislative enactments pursuant to constitutional restraints set out in the Canadian Constitution. If any enactment should successfully resist any challenges under the Constitution, it is to the Legislature, and not to the courts, that the community should find relief.

[71] The action is dismissed with costs if demanded.

[72] Action dismissed.

# Example Court Case Document

Supreme Court of India

Subimal Sarkar v Sachindra Nath Mondal and others, 2003

1. Original complainant in Sessions Case No.127 of 1989 before the Sessions Judge, West Dinajpur, at Balurghat is in appeal before us in the above Criminal Appeal, against the judgment of acquittal made by the High Court Calcutta in Criminal Death Reference No.4/1990 and Criminal Appeal No.327 of 1990. The State has not preferred any appeal against the judgment of the High Court, but is a party respondent before us.

2. Prosecution case briefly stated is that one Suchitra, the daughter of the appellant herein was married to Nakul Chandra son of Sarat Chandra Mandal, accused No.1 before the Sessions Court. At the time of the marriage, there was an agreement to pay dowry in cash of Rs. 500/- out of which, Rs. 300 was paid to the accused A-1 by the appellant but he could not pay the balance amount of dowry due to poverty. A-1 was living with his wife A-4 and five of his children of which A-3 Sachindra Nath Mandal is one of them who was living with his wife Gauri Mandal, who was A-4 before the Sessions Court. The husband of deceased Suchitra was Nakul Mandal, who was also staying with his father. It is the prosecution case that because of the non-payment of balance of dowry, there was constant torture and ill-treatment of the deceased by the accused persons. This was made known to the appellant (PW-11) as also his wife PW-1 Arati Sarkar.

3. On 16.8.1986, PW-1 had come to know that her daughter had consumed poison in the house of accused No.1 (father in-law) and in view of the fact PW-11, husband of PW-1 was along with PW-2 away in the field, had proceeded towards the village of her sambandhi and on the way they met one Nakul Bhunia, who took them on his cycle towards Durlavpur. Further on the way, they found the body of deceased Suchitra being carried on Plank tied to the same by A-3 Sachinder Mandal with the help of some people and on seeing them A-3 and others left the body there and went to some distance. When PW-1 and PW-2 went near the body, they found bleeding from the mouth and nostril of the deceased, therefore they took the deceased across the river Pagliganj to Balurghat Hospital, where PW-15 Dr. Nath examined the deceased and declared her as brought dead. He also opined that, the death was caused by throttling. On the basis of the information, received from the Dr. PW-15, the police of the Balurghat Police Station registered a case under Section 302 [IPC](#) and started the investigation. The body was sent for post-mortem examination, which was conducted by Dr. D. Shah PW-14 on 17.8.1986. He opined that cause of death of the deceased was due to manual strangulation and was homicidal in nature. During the post-mortem, he noticed the bruise mark on the neck and both on right and left side of the wind pipe. On dissection, he found the hyoid bone fractured. After completion of the investigation, the police filed charge-sheet against the above-mentioned four accused persons.

4. The prosecution has primarily relied upon the evidence of PW-1, the mother, PW-2 aunt, PW-4 brother, PW-11, the father and another relative PW-2, Govind Sarkar to establish the fact that there was a dowry demand, pursuant to which the deceased was being tortured and on the date of incident the body of the deceased was being carried on a plank tied to the same. The prosecution also relied upon the medical evidence to establish the fact that there was a homicidal death by throttling.

5. A very peculiar defence was taken by the accused persons who contended that on the fateful day, the deceased had consumed poison and when it was noticed by her husband, Nukul, he tried to prevent the deceased from swallowing the poison by pressing the neck at which time A-3, brother in law of the deceased allegedly came and put his finger into the throat of the deceased to make her vomit the poison. The further contention was that at that point of time, deceased was alive, therefore, they decided to take her to the hospital and it was during that time that the mother of the deceased met on the way when the body was being taken to the hospital. The Trial Court on consideration of the evidence on record came to the conclusion that the prosecution by circumstantial evidence has established the case against the accused, hence held the accused persons guilty and awarded capital punishment to A-3, the brother in law of the deceased, Suchindra Nath Mandal, while the other three accused persons were convicted under Section 302 read with 34 and were awarded life imprisonment.

1. Original complainant in Sessions Case No.127 of 1989 before the Sessions Judge, West Dinajpur, at Balurghat is in appeal before us in the above Criminal Appeal, against the judgment of acquittal made by the High Court Calcutta in Criminal Death Reference No.4/1990 and Criminal Appeal No.327 of 1990. The State has not preferred any appeal against the judgment of the High Court, but is a party respondent before us.

## No paragraph headings

2. Prosecution case briefly stated is that one Suchitra, the daughter of the appellant herein was married to Nakul Chandra son of Sarat Chandra Mandal, accused No.1 before the Sessions Court. At the time of the marriage, there was an agreement to pay dowry in cash of Rs. 500/- out of which, Rs. 300/- was paid to the accused A-1 by the appellant but he could not pay the balance amount of dowry due to poverty. A-1 was living with his wife A-4 and five of his children of which A-3 Sachindra Nath Mandal is one of them who was living with his wife Gauri Mandal, who was A-4 before the Sessions Court. This marriage 17 December 1981 was Nakul 11/11/81, who was 18 years old in his father's home. In the execution of law because of the non-payment of balance of dowry, there was constant torture and ill-treatment of the deceased by the accused persons. This was made known to the appellant (PW-1) as also his wife PW-1 Arati Sarkar.

3. On 16.8.1986, PW-1 had come to know that her daughter had consumed poison in the house of accused No.1 (father in-law) and in view of the fact PW-11, husband of PW-1 was along with PW-2 away in the field, had proceeded towards the village of her sambandhi and on the way they met one Nakul Bhunia, who took them on his cycle towards Durlavpur. Further on the way, they found the body of deceased Suchitra being carried on Plank tied to the same by A-3 Sachinder Mandal with the help of some people and on seeing them A-3 and others left the body there and went to some distance. When PW-1 and PW-2 went near the body, they found bleeding from the mouth and nostril of the deceased, therefore they took the deceased across the river Pagliganj to Balurghat Hospital, where PW-15 Dr. Nath examined the deceased and declared her as brought dead. He also opined that the death was caused by strangulation. On the basis of the information, received from the Dr. PW-15, the police of the Balurghat Police Station registered a case under Section 302 IPC and started the investigation. The body was sent for post-mortem examination, which was conducted by Dr. D. Shah PW-14 on 17.8.1986. He opined that cause of death of the deceased was due to manual strangulation and was homicidal in nature. During the post-mortem, he noticed the bridle marks on the neck and both the right and left side of the wind pipe. On dissection, he found the hyoid bone fractured. After completion of the investigation, the police filed charge-sheet against the above-mentioned four accused persons.

4. The prosecution has primarily relied upon the evidence of PW-1, the mother, PW-2 aunt, PW-4 brother, PW-11, the father and another relative PW-2, Govind Sarkar to establish the fact that there was a dowry demand, pursuant to which the deceased was being tortured and on the date of incident the body of the deceased was being carried on a plank tied to the same. The prosecution also relied upon the medical evidence to establish the fact that there was a homicidal death by throttling.

5. A very peculiar defence was taken by the accused persons who contended that on the fateful day, the deceased had consumed poison and when it was noticed by her husband, Nukul, he tried to prevent the deceased from swallowing the poison by pressing the neck at which time A-3 (brother-in-law) of the deceased, who was riding a cycle, put his finger into the throat of the deceased to make her vomit the poison. The further contention was that at that point of time, deceased was alive, therefore, they decided to take her to the hospital and it was during that time that the mother of the deceased met on the way when the body was being taken to the hospital. The Trial Court on consideration of the evidence on record came to the conclusion that the prosecution by circumstantial evidence has established the case against the accused, hence held the accused persons guilty and awarded capital punishment to A-3, the brother in law of the deceased, Suchindra Nath Mandal, while the other three accused persons were convicted under Section 302 read with 34 and were awarded life imprisonment.

## Page after page of unstructured text

## Legal Issues?

## Arguments?

## Final Judgement?

# Challenges in processing legal documents

- Several India-specific challenges
- Indian legal documents mostly unstructured and disorganized
- Lack of standard notations for writing legal documents
- Difficult to create large datasets required for supervised NLP/IR methods
  - Requires annotators have expertise in Law → expensive, time-consuming

## Challenges → Opportunities in AI & Law

- Huge amounts of raw legal data made available digitally in the last 5-6 years
  - Law practitioners need automated systems to make sense of all this data
- Severe lack of knowledge of Law among the common masses
- Expensive to consult Law experts
- Economic benefits of developing automated systems that are **easy to use for Law practitioners and the common man** and provide usable results → startups

# Research problems in AI & Law

## Popular research problems in AI & Law

- Retrieval / recommendation of precedents
- Rhetorical labelling / semantic segmentation of case judgements
- Summarization of case judgements and other legal documents
- Legal judgement prediction from given facts
- Legal Statute Identification: Identifying relevant statutes from given facts
- Legal citation network analysis
- Explainability of AI-Law models
- Training language models specific to the Law domain
- ... and many more

## Popular research problems in AI & Law

- Retrieval / recommendation of precedents
- Rhetorical labelling / semantic segmentation of case judgements
- Summarization of case judgements and other legal documents
- Legal judgement prediction from given facts
- Legal Statute Identification: Identifying relevant statutes from given facts
- Legal citation network analysis
- Explainability of AI-Law models
- Training language models specific to the Law domain
- ... and many more

# Case judgement summarization

# Why summarize legal case judgements?

- Law practitioners have to go through hundreds of legal case judgements
  - Common Law system
  - Important to identify precedents - prior cases that are relevant to an ongoing court case
- Lengthy, unstructured, dense legal text →Reading and comprehending the full documents is time-consuming even for Law practitioners

## Challenges in summarizing legal case judgements

- Very few datasets available for training legal summarization models
- Expensive to get reference summaries written by Law experts
- Documents are lengthy (thousands of words) → cannot be easily handled by most abstractive summarization models

# Datasets for case judgement summarization

- 7,130 Indian Supreme Court case judgements and **abstractive** summaries
  - Average #words in documents: 4,378
  - Average compression ratio: 0.24
- 793 UK Supreme Court case judgements and **abstractive** summaries
  - Average #words in documents: 14,296
  - Average compression ratio: 0.11
- 50 Indian Supreme Court case judgements and **extractive** summaries
  - Average #words in documents: 4,378
  - Average compression ratio: 0.31

IN-Abs

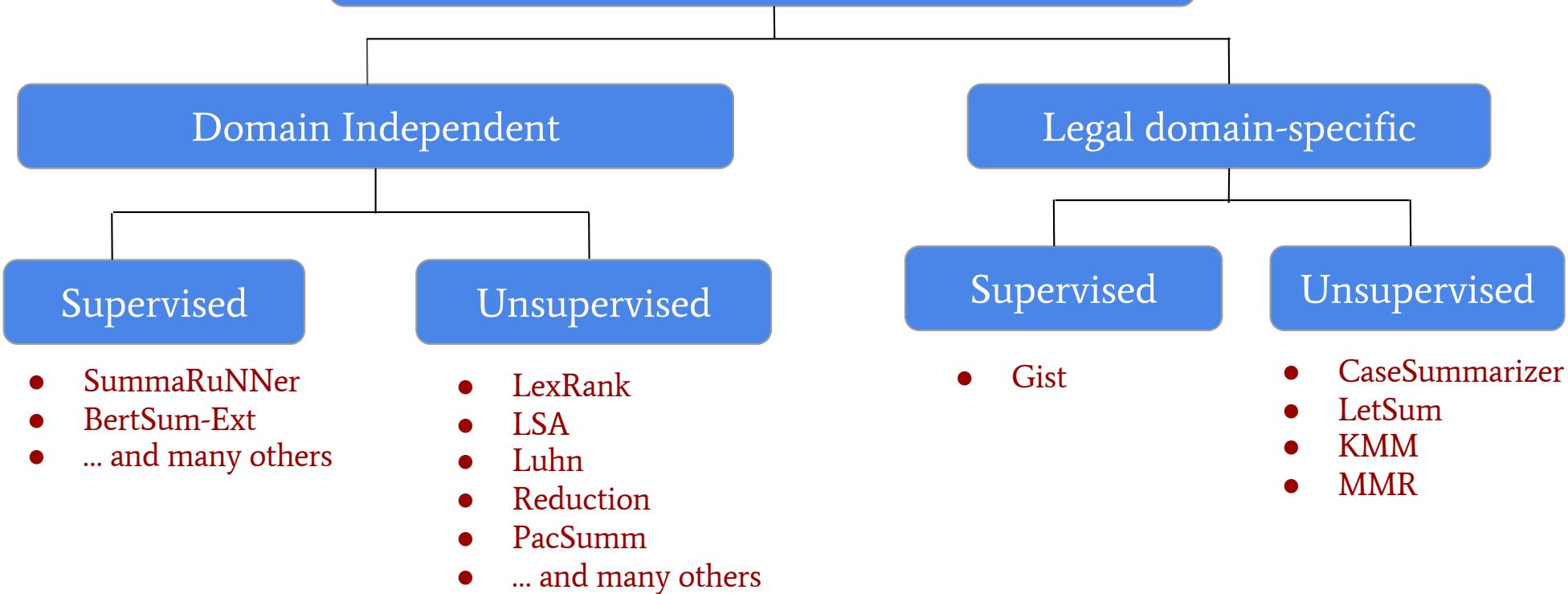
UK-Abs

IN-Ext

Dataset	# Train docs	# Test docs
IN-Abs	7,030	100
UK-Abs	693	100

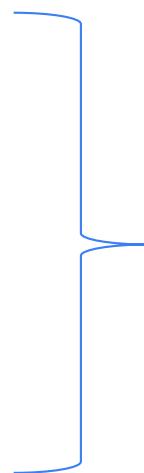
All datasets available at  
<https://github.com/Law-AI/summarization>

## Extractive Text Summarization Algorithms



## Several abstractive summarization models

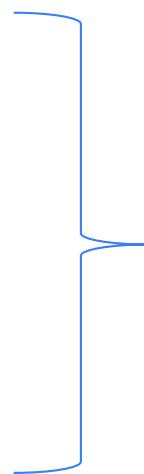
- Pointer Generator
- BERTSum-Abs
- BART
- Pegasus
- BigBird
- Longformer
- ... and others



- Pre-trained versions – usually trained over news summarization datasets
- Fine-tuned versions – fine-tuning over legal (doc, summary) pairs likely to yield better results

# Several abstractive summarization models

- Pointer Generator
- BERTSum-Abs
- BART
- **Pegasus**
- BigBird
- **Longformer**
- ... and others



- Pre-trained versions – usually trained over news summarization datasets
- Fine-tuned versions – fine-tuning over legal (doc, summary) pairs likely to yield better results

Legal-Pegasus – <https://huggingface.co/nsi319/legal-pegasus>

Legal-Longformer - <https://huggingface.co/nsi319/legal-led-base-16384>

Already fine-tuned over legal summarization data (litigation releases and complaints from US courts)

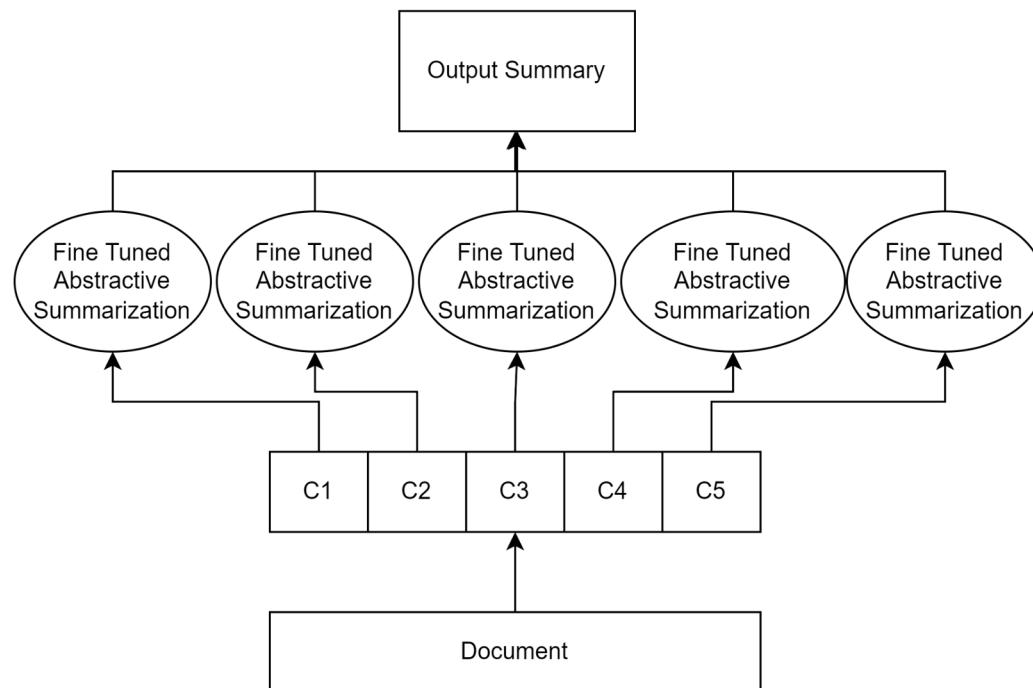
# Main challenge in applying abstractive models on legal docs

- Limit on the number of input tokens
- Many case docs do not fit in any of these models except Longformer
  - Several docs in UK-Abs do not fit fully even in Longformer

Dataset	Avg #words per doc
IN-Abs	4,378
UK-Abs	14,296

Model	Input Capacity
Pointer Generator	400
BERTSum-Abs	512
Legal-Pegasus	1024
BART	1024
BigBird	4*1024
Longformer	16*1024

## Solution: chunking with fine-tuning over legal data



Challenges:

How to get chunk-summary pairs  
for fine-tuning such models?

Loss of information across chunks

# Summarization performance

- Extractive models trained over legal data  
vs.
- Pre-trained abstractive models fine-tuned over legal data

Results on IN-Abs dataset

	ROUGE-1 F	ROUGE-2 F	ROUGE-L F
Gist (Extractive)	0.471	0.238	0.308
SummaRuNNer (Extractive)	0.493	0.255	0.274
<b>BART (Abstractive)</b>	<b>0.495</b>	<b>0.249</b>	<b>0.330</b>
<b>Legal-Pegasus (Abstractive)</b>	<b>0.488</b>	<b>0.252</b>	<b>0.341</b>

# Summarization performance

- Extractive models trained over legal data
- vs.
- Pre-trained abstractive models fine-tuned over legal data

Results on IN-Abs dataset			
	ROUGE-1 F	ROUGE-2 F	ROUGE-L F
Gist (Extractive)	0.471	0.238	0.308
SummaRuNNer (Extractive)	0.493	0.255	0.274
<b>BART (Abstractive)</b>	<b>0.495</b>	<b>0.249</b>	<b>0.330</b>
<b>Legal-Pegasus (Abstractive)</b>	<b>0.488</b>	<b>0.252</b>	<b>0.341</b>

We also conducted evaluation of summaries by Law experts

Law experts often prefer summaries by simple extractive models

Lot of scope for improvement of legal summarization models

## More details in our papers

- *A comparative study of summarization algorithms applied to legal case judgements*, European Conference on Information Retrieval (ECIR) 2019
- *Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents*, International Conference on Artificial Intelligence and Law (ICAIL) 2021 [Best Student Paper Award]
- *Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation*, AACL-IJCNLP 2022

Datasets and codes: <https://github.com/Law-AI/summarization>

Legal judgement prediction (LJP)  
Legal Statute Identification (LSI)

# LJP and LSI

- LJP: Automatically predicting the outcome of a case, given the facts of the case
- Variations of LJP – predicting one of the different aspects of a judgment
  - Predicting the outcome of an appeal (e.g., bail appeal accepted / rejected) [Malik21]
  - Identifying relevant law articles/statutes [Chalkidis19, Paul22]
  - Predicting the term (e.g., no. of months/years) of penalty if convicted [Xiao18]

## Legal Statute Identification (LSI)

[Malik21] ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation, <https://aclanthology.org/2021.acl-long.313/>

[Chalkidis19] Neural Legal Judgment Prediction in English, <https://arxiv.org/abs/1906.02059>

[Paul22] LeSICiN: A Heterogeneous Graph-based Approach for Automatic Legal Statute Identification from Indian Legal Documents, <https://arxiv.org/abs/2112.14731>

[Xiao 2018] CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction, <https://arxiv.org/abs/1807.02478>

# Datasets for LJP and LSI

- ECtHR-A and ECtHR-B
  - Cases from the European Court of Human Rights; labels are 10 articles from the European Convention on Human Rights
  - Part of LexGLUE benchmark: <https://github.com/coastalcph/lex-glue>
- CAIL
  - Cases from China Judgments Online; labels based on 183 criminal laws; also contains labels for charge and term of penalty prediction
  - <https://arxiv.org/abs/1807.02478>
- ILDC (Indian Legal Data Corpus)
  - Cases from the Indian Supreme Court; final “accept” / “reject” decision to the appeal(s) made by the petitioner (binary classification setup)
  - <https://github.com/Exploration-Lab/CJPE>
- ILSI (Indian Legal Statute Identification)
  - Cases from the Indian Supreme Court and High Courts; labels are 100 sections from the IPC
  - <https://github.com/Law-AI/LeSICiN>

# The ILSI dataset: labels

- The IPC is divided into Chapters, Topics, Sections
- The most frequently cited 100 Sections are the labels

Chapter	Topic	Section
Offences affecting Human Body	Offences affecting Life	<b>299:</b> Culpable homicide
		<b>307:</b> Attempt to murder
	Hurt	<b>321:</b> Voluntarily causing hurt
		<b>334:</b> Voluntarily causing hurt on provocation
Offences against Property	Robbery and Dacoity	<b>390:</b> Robbery
		<b>396:</b> Dacoity with murder
	Criminal Trespass	<b>441:</b> Criminal trespass <b>446:</b> House breaking by night

Details in our AAAI2022 paper:  
<https://arxiv.org/abs/2112.14731>

## The ILSI dataset: an example

On the fateful day at about 9.30 a.m. deceased accompanied by Mansingh (PW 4) and Gulabsingh (PW 7) was going from his village Talod to Alore. The accused persons were hiding behind bushes on the road near village Gharola. They were armed with lathies and farsies. When the deceased and the aforesaid two persons reached near the Khakhra, the respondents surrounded them and started attacking the deceased with weapons with which they were armed. His nose was cut. PWs. 4 and 7 tried to intervene, but they were also attacked by the accused persons as a result of which they also received injuries. The two witness rushed to the police station where PW 4 lodged the FIR (Exhibit P-10). The deceased in injured condition was taken to the hospital, and later he succumbed to the injuries. Post-mortem was conducted and large number of injuries were found on his body. During investigation the alleged weapons of the assailants were seized. After investigation charge sheet was placed.

Section 324 of IPC: Voluntarily causing hurt by dangerous weapons or means

## The ILSI dataset: an example

On the fateful day at about 9.30 a.m. deceased accompanied by Mansingh (PW 4) and Gulabsingh (PW 7) was going from his village Talod to Alore. The accused persons were hiding behind bushes on the road near village Gharola. They were armed with lathies and farsies. When the deceased and the aforesaid two persons reached near the Khakhra, the respondents surrounded them and started attacking the deceased with weapons with which they were armed. His nose was cut. PWs. 4 and 7 tried to intervene, but they were also attacked by the accused persons as a result of which they also received injuries. The two witness rushed to the police station where PW 4 lodged the FIR (Exhibit P-10). The deceased in injured condition was taken to the hospital, and later he succumbed to the injuries. Post-mortem was conducted and large number of injuries were found on his body. During investigation the alleged weapons of the assailants were seized. After investigation charge sheet was placed.

Section 302 of IPC: Punishment for Murder

# Methods for LSI mostly use the text of facts / statutes

- Encode the fact descriptions and then classify directly
  - E.g., a hierarchical encoder to capture the textual structure - HierBERT [Chalkidis19]
- Some methods also encode the textual description of the statutes
  - Efforts toward better identification of rare statutes (which mostly co-occur with frequent ones) [Wang18]
  - Learning the hierarchy between statutes themselves [Wang19]

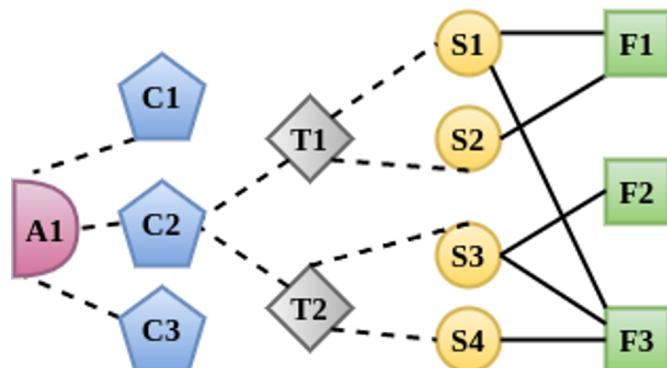
[Chalkidis19] Neural Legal Judgment Prediction in English, <https://arxiv.org/abs/1906.02059>

[Wang18] Modeling Dynamic Pairwise Attention for Crime Classification over Legal Articles,  
<https://dl.acm.org/doi/10.1145/3209978.3210057>

[Wang19] Hierarchical Matching Network for Crime Classification, <https://dl.acm.org/doi/10.1145/3331184.3331223>

# Utilizing text and network for LSI

Chapter	Topic	Section
Offences affecting Human Body	Offences affecting Life	<b>299:</b> Culpable homicide
		<b>307:</b> Attempt to murder
	Hurt	<b>321:</b> Voluntarily causing hurt
		<b>334:</b> Voluntarily causing hurt on provocation
Offences against Property	Robbery and Dacoity	<b>390:</b> Robbery
		<b>396:</b> Dacoity with murder
	Criminal Trespass	<b>441:</b> Criminal trespass
		<b>446:</b> House breaking by night



- Facts cite the sections of IPC
- We consider a heterogeneous legal network
- Nodes: Facts, Sections and categories like Topics, Chapters
- Edges: ‘part of’ and ‘cites’ relationships
- **LSI → link prediction from fact nodes to statute nodes over this heterogeneous network**

Details in our AAAI2022 paper:  
<https://arxiv.org/abs/2112.14731>

# Pretrained language models for Law

# Pre-Trained Language Models (PLMs)

Transformer-based PLMs have become popular in recent years

- E.g., BERT, RoBERTa, ...
- Can be further trained for many NLP tasks (classification, summarization, ...)

PLMs learn to understand natural language by

- Seeing a large amount on text, and
- Trying to “predict the next word/sentence” or “fill in the blank” in a phrase →  
Masked Language Modeling, Next Sentence Prediction

# PLMs for the Legal Domain

- **LegalBERT:** Chalkidis et al. (2020) pre-trained BERT-base on EU, US, UK case documents; <https://huggingface.co/nlpaaueb/legal-bert-base-uncased>
- **CaseLawBERT:** Zheng et al. (2021) pre-trained BERT-base on US case law documents; <https://huggingface.co/zlucia/legalbert>
- **Lawformer:** Xiao et al. (2021) pre-trained Longformer on Chinese legal docs
- **InLegalBERT:** Paul et al. (2022) further pre-trained LegalBERT on a large collection of Indian legal text; <https://huggingface.co/law-ai/InLegalBERT>
  - Out-performs LegalBERT and CaseLawBERT on several tasks including LJP and LSI, both on Indian datasets and ECHR dataset

# Concluding Part 2

## Resources page

- All datasets / models / papers discussed are collected here:

<https://github.com/Law-AI/codscomad2023tutorial>

- Also contains links to additional resources