Classifying the structure of court judgments: A machine learned approach

Legal documents, particularly court orders and judgments, are often lengthy and complex, containing a wide range of information. A person reading a judgment may be interested in only a few specific portions of a judgment such as arguments of a particular party or the court's reasoning for its decision or how the legal sections have been interpreted. One has to read the whole judgment sequentially to come to the points of his or her interest.

The sheer size of these judgments, ranging from a 2-page bail order to behemoths like the Kesavananda Bharati case (700 pages), NJAC judgment (1042 pages), Puttaswamy case (1448 pages), and Ayodhya judgment (1045 pages) make it a time consuming effort. Further judgment writing is judge-specific and the facts, issues, argument by parties, court's reasoning etc in a judgment is spread around, making it difficult for anyone to quickly pin-point a part of interest. Therefore, there is a need to segment parts of the judgment in various classes of interest.

Our database has more than 2.8 crore orders and judgments and we add 20,000 new documents almost daily. A manual segmenting of each document is almost an impossible task. Secondly, the way judgments are written and the variety among them make any rule based system for categorizing paragraphs in a judgment is not easy.

Our approach centered around training a machine learned model to automatically classify and structure judgment paragraphs. We utilized the Virtual Legal Assistant (VLA) on the Indian Kanoon website to mark paragraphs in judgments with different classes. This meticulous process ensured no part of the judgment remained uncategorized. Then we generated over 5,000 most frequent phrases and used them as features for training the model. Beyond linguistic features, we also incorporated the starting position of each paragraph within the entire judgment as an additional feature. This allowed the model to leverage the typical flow of information within a judgment, recognizing that facts usually appear earlier (0-10%) while conclusions reside towards the end (90-100%). With the combination of linguistic and positional features, we trained a Logistic Regression model to predict the class of each paragraph. To ensure its accuracy, we meticulously tagged the paragraphs of over 1,000 court judgments, providing the model with a robust training dataset. Read our journey here.

Classifying the structure of the judgments improves readability and allows users to focus on portions of judgments that they are interested in. We classify the structure into the following classes:

1) Facts:

This class contains the description of the events that led to the issue of the case, the apathy of the parties involved and the legal course already employed till the adjudication of the case.

2) Arguments by Petitioner/Appellant:

This class contains the arguments made by the Petitioners/Appellants, the legal position of the Petitioner/Appellant and their reasoning. Everything which is extended from the petitioner side be it cases, sections, or justifications/explanations of their clients actions.

3) Arguments by Respondent/Defendant:

This class contains the primary arguments made by the Respondent/Defendant, the legal position of the Respondent/Defendant and their reasoning. Everything which is extended from the petitioner side be it cases, sections, or justifications/explanations of their clients actions.

4) Precedent Analysis:

This class contains the text of the judgment where a precedent (previous case law) is discussed/analyzed by the court with the particular case being cited. Mostly the format followed by courts is the Case being cited on the starting of the paragraph with a discussion on it followed in the subsequent paras.

5) Section Analysis:

This class contains the discussions on the Laws (sections) required for the adjudication. This contains Sectional Analysis, Listing and a legal section's trajectory/historical development.

6) Issues/legal issues:

This class contains the Legal Issue which the case is being held to decide, the questions the court has to address in order to deliver a judgment. Legal issue or issue of law is a legal question which is the foundation of a case.

7) Court's Reasoning:

This class contains the general arguments by the judges while writing the judgment. It is the discourse undertaken by the judge to reach the conclusion/decision. This class is different from Ratio as this contains the part of the judgment which is for detailed explanation and reasoning taken by the judge, while ratio is the binding precedent part of the judgment.

8) Conclusion:

This class contains the final/concluding remarks of the judgment, it also includes the decision of the court. This class also contains the final summing up of the judgment, mostly found in the ending part of the judgments text.

Overall, this feature has the potential to significantly cut down the time required to read court judgments with in certain context. It facilitates efficient information retrieval, which can lead to greater efficiency and accuracy in legal research, as well as improved understanding of legal decisions.