



Law Implementation Tracker | POCSO | Project Report

CivicDataLab | HAQ

3/9/2020

A few updates (Jan/Feb, 2020)

Overview

- Developing the annotation platform for tagging judgements
- Finalise the list of variables to be identified from Judgements
- Fetching data for POCSO cases from eCourts
- Building the frames for data analysis
- Partner Collaborations

Judgement Annotation

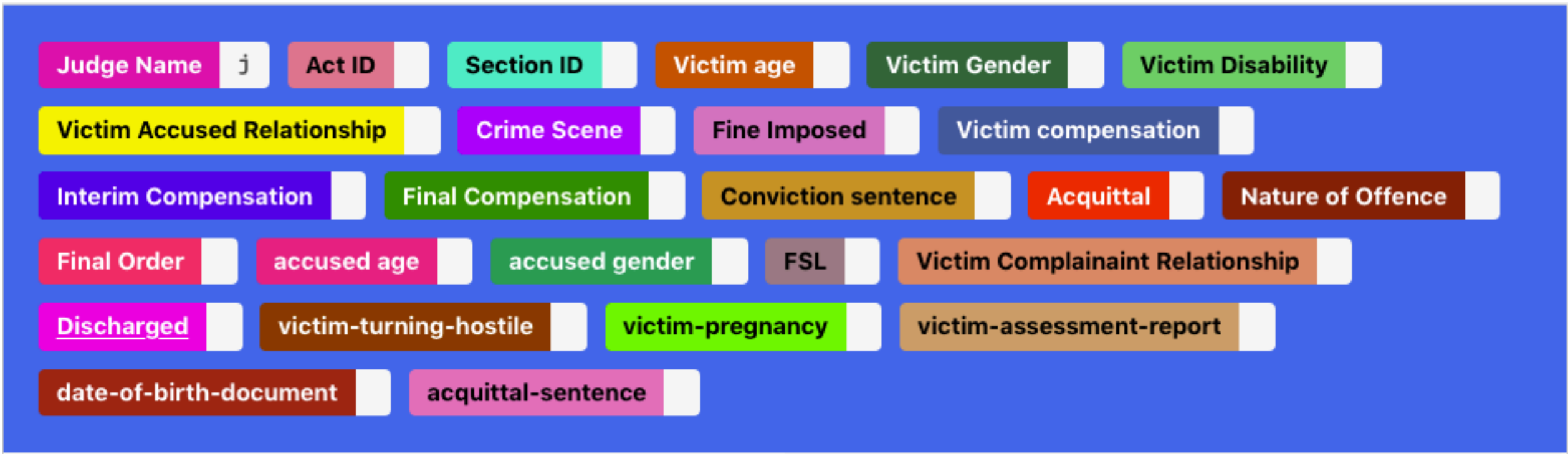
Process, Progress and Challenges

Processes

- The process was initiated by annotating judgements from the district courts of Delhi
- The first batch included 63 judgements.
- This was a *random sample* of judgements that included cases of both acquittal conviction, the former being the larger chunk

Tags

- A total of 26 variables were identified to be tagged in judgements



- [Judgement Annotation Platform](#)
 - Developed using an open source tool - Doccano [\[Link\]](#) | [\[Project on Github\]](#)

Progress

- The annotation process for the first batch was completed over a period of 5 weeks
- Overall, we observed, some variables like Judge Name, Act ID, Section ID, etc.. are easier to identify, but variables like Crime Scene, Victim Age, Victim Gender are very diverse in terms of reporting, and we shall require more judgements from different courts to identify patterns, if any.
- Victim Disability will be identified through specific words. Details [here](#).

Contd ..

- Few rule based variables:
 - `Victim Gender` - Identify `prosecutrix` - these are cases with a female victim
 - `Fine Imposed` - Presence of certain sections related to the POCSO act
- We could not find enough judgements to identify reporting patterns for variables such as `FSL`, `Discharged`, `victim-pregnancy`, `Victim Accused Relationship`

List of variables

Current Status and Next Steps

- [Annotation Report](#)

Challenges

- Finding a good sample for judgement annotation across districts
- We have a few variables such as `Fine Imposed`, `Victim Compensation`, etc that are mostly present in cases of conviction, hence a good diverse sample is required
- Victim `Gender` and `Age` variables are hard to extract.
- Some variables like `Disability`, `Pregnancy` and `Crime Scene` are a too generic and we may need to reduce their scope a bit.
 - For Eg: instead of talking about crime scene, we can find cases that happened in a school, college, neighbourhood, etc.

Next Steps on Judgement Annotation

- Annotating judgements for remaining states
- Scaling initial research to all district courts from Delhi
- Setting up processes to validate the variables identified through the annotation process

Data Mining

Process, Progress and Challenges

Process- Fetching data from eCourts

- Metadata (All variables for a case - as is) [List](#)
- All orders and judgements
- [Project Repository](#)

Progress

- [Data Report](#)

The frames of analysis

- [Link to document](#)

Partner Collaborations

Swagata Raha

- Has agreed to offer her time and expertise for legal research and data analysis.
- She led a few studies in Delhi, Assam and Maharashtra on the Implementation of the POCSO Act by Special Courts as part of the Centre for Child and the Law, NLSIU Bangalore.
- All reports for the study are shared on the [drive](#)

The Engine Room

- They're supporting us on dealing with reporting of sensitive datasets and finding ways to safely navigate using and sharing data on child sexual abuse.
- Will also help us develop a responsible data handling and sharing policy for the project.
- [Responsible Data forum](#)

Next Steps (March/April, 2020)

- Start working towards the research and analysis part
- Develop a robust data processing and validation process and
- Building a data platform to host datasets and analysis.

Project Health

Project Status (As per the project plan)

Plan till March, 2020

Task	Status
Finalize the scope (data mining+platform development) of the project	Done
Onboard the team at HAQ to internal project management tools	Done
Development of the language platform for annotating orders and judgements	Done
Setting up an Advisory Committee for the project	Done
Advisory Board Meeting 1	Done
Build data mining infrastructure and develop process to mine data from Ecourts	Started
Develop data models to store and process data mined from the ecourts platform	Started
Using the language platform to annotate judgements	Done
Develop the Privacy, Ethics and Security framework for sharing and working with child-rights cases	Pending
Develop an editorial process to publish and share research and other development related stories as part of the tracker	Pending
Scale the initial research to all identified geographies	Started
Develop a data validation mechanism to verify tags, and other extracted entities from unstructured data	Started
Identify collaborations, partnerships and research grants to fund the research and development for the next phase of the project	Started

- [Statement of Work](#)

Funding Status

- On discussion with Nilekani Philanthrophies
- Other grants