



Court Case Extraction Tool for Tracking Police Misconduct

Pragya Kallanagoudar, Chithra Anand, Rolando Garcia, Sarah Chasins, and Aditya Parameswaran

Ranking

The list of officer records is ordered according to our ranking algorithm, which is designed to reflect the priorities of public defenders when they manually make these matching decisions.

The algorithm incorporates the first, middle, and last name, and agency while “scoring” each match. Each parameter in the record is ascribed a point value in the scorecard according to this criteria:

- first name: 1 if matches document, -1 if not
- last name: 1 if matches document, -1 if not
- middle name:
 - 0 if no middle name specified in document
 - if middle name is specified in document, then 1 if it matches, -1 if not
- agency:
 - strips record and document agency down to city/county name only and takes the set intersection of these
 - 1 if set intersection is nonempty, 0 otherwise

The final score for the record is the sum of all the values in the scoreboard.

Introduction

Journalists and public defenders manually review and analyze a large volume of police misconduct case documents to source key information about these lawsuits. Given that these documents are highly varied in format, this is a laborious and time-consuming process.

Goal

We are creating software to extract data from these documents, and allow users to navigate the documents and review and evaluate the extracted information.

Parsing

Through optical character recognition, the PDFs are parsed and converted into text files and CSVs containing the coordinates of each word.

Filtering

The names of officers and their affiliated agencies are identified from the parsed results.

- Officer names: Capture strings surrounding relevant titles [“officer”, “detective”, “lieutenant”, “inspector”, etc.] then look through document for mentions of those names
- Affiliated agencies: Capture strings surrounding relevant agency indicators [“police department”, “city (of) police department”, “county (of) police department”, “PD”, etc] as well as whether city name is one/two word (Baltimore vs. New York)

Matching

- Possible matching officer records from the NACDL database are identified for each recognized officer name.
- Each officer record includes a full name, a list of affiliated agencies, and a list of positions held with start and end dates.

Future Directions

- Create buttons for user to indicate whether:
 - a recognized name is an officer or not
 - an officer record is a match, not a match, or possibly a match
- Add subscores for positions and dates to the ranking algorithm
- Modify the ranking algorithm to include weights for each parameter depending on how crucial in determining a match

Acknowledgements

Julie Ciccolini, Lisa Pickoff-White, Tristan Chambers