

"Briefcase": Intelligent Court Transcript Management: Timeline and Budget

Lucille Njoo
Honors Summer Research Fellowship

VII. TIMELINE

Week 1: Refactor old data structures and scripts to be more organized and scalable.

- Our old data structures primarily relied on one monolithic document class that stores all the document's pages; this is not granular, organized, scalable, or flexible enough to support complex analysis on a thousand-page transcript.
- Refactor into a hierarchy in which a top-level document contains examination sections, which each contain pages, which each contain quotes.
- Create classes for actors and other data that we will extract later.

Week 2: Clean up processed text and create an interface that just displays raw text.

- Try to correct misspelled words from processing imperfect PDFs.
- Fix lingering bugs and handle edge cases in the sorting of text fragments.
- Explore spaCy's tools for better, cleaner, more accurate alternatives to regular expressions for capturing certain parts of the text.
- Create a front-end that just displays text from mock data.

Week 3: Begin constructing custom knowledge bases; make key words in the text interface interactable.

- Dive into spaCy documentation and experiment with making knowledge bases.
- Separate out key words in the text to make them color-coded and interactable.
- Create mouse-over popups that display information on a given actor.

Week 4: Work on entity linking; implement a toolbar with a document outline and search/filter options.

- Use custom spaCy knowledge bases to analyze the document text and extract names more accurately.
- Use spaCy's entity linking tools to create a single object for each actor.
- Connect all mentions of a given actor to the single actor entity object; do this for all actors.

Week 5: Implement annotation features and build the minimap.

- Use React libraries and frameworks to implement features for highlighting, bookmarking, and commenting in the interface.
- Programmatically build the minimap by generating colored blocks whose sizes and locations correspond to the length and location of the section of the document they each represent.

Week 6: Host the web app and send it to the Project for the Innocent.

- Find a web hosting platform.

- Send the MVP to our liaison at the Project for the Innocent.
- Communicate with the Project through email, phone, or Zoom to gather feedback.

VIII. BUDGET

The SURP stipend covers 10 hours of work per week; I plan to spend roughly 20 hours per week on research.

- $6 \text{ weeks} \times (20 \text{ hours per week} - 10 \text{ hours per week}) = 60 \text{ hours}$
- $60 \text{ hours} \times \$13.25 \text{ per hour} = \795.00

Web Hosting: We do not yet know what kind of web hosting service would best suit this project, so the numbers below are estimates/averages.

- Purchasing a domain: \$10.00
- Web hosting: \$50.00 per year
- $\$10.00 + \$50.00 = \$60.00$

TOTAL: $\$795 + \$60 = \$855$