Clause Search

Legal Document clause search

Aggarwal, Rohit

**Problem Statement**

Finding the clauses that tells us power exercised by a Director of a company from various legal documents. After that summarize it to a smaller report for the user.

**Final Approach**

The problem is one of unsupervised learning. Following are the steps followed by me:

1. Converting text to image
2. Converting image to text
3. Splitting text into clauses
4. Separating clauses containing Director word
5. Performing clustering on the clauses
6. Separating the clauses with different labels
7. Summarizing the clauses

**Problems Faced and Solution**

**P: Problem S:Solution**

* Some documents are scanned and uploaded
* Reading text from pdf
* Splitting the text into segments. (still needs to refine)

Software needed:

* Tesseract-ocr
* Poppler
* Pdf2image (pip)
* Pillow (pip)
* pyTesseract (pip)
* numpy(pip)