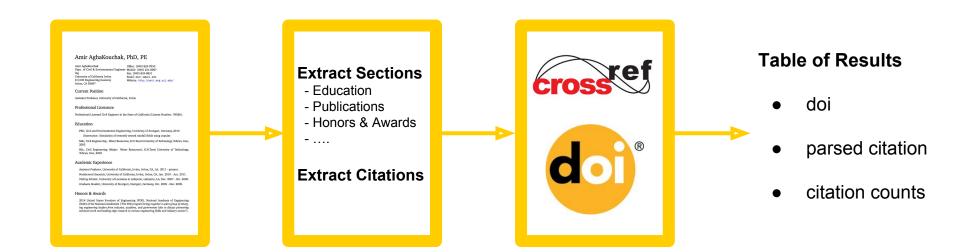
UC Recruit - CV Parser









Extract Section Details

- Group text based on features
 - UPPER CASE
 - text size
 - left indentation
 - font
 - bold / italics
 - o line space*
- Compare each group with known section names



Extract Citation Details

- Collect all text in publication section
- Parse text by identifying citation style
 - o numbered / bulleted list
 - o author's name
 - indentation style
 - etc.



Example 1 - Amir AghaKouchak

- 5 possible section groups found
 - Group 2 contains the sections
- Extracted all text from *publications* to *service committee* panel assignments
- Parsed citations using the numbered list



Example 2 - Ashley T. Rubin

- Column format resume
 - Code determines resume uses columns, then processes sections appropriately
- Extracted all text to the right of peer-reviewed to other publications

Parsed citations using the author's name



Example 3 - Owen Whooley

• Pass 1 finds 2 groups, neither contain sections.

- We then proceed to pass 2 and find the sections
 - Pass 2 allows centered section names to be identified
- Parsed citations using the author's name



Example 4 - Shana Siegel

Group 1 is identified by the uppercase text

- Citation text is extracted and citations are parsed and passed to crossref
- No results likely means citation text was not grouped correctly



Example 5 - Daniel Cohen

Sections were found and citation text extracted

 All citation parsing techniques were attempted but none were successful



Results Summary

- CrossRef returns the title, authors, journal, and a score for the match
- Check returned title against original citation title_score
- Check returned authors against CV author remove results that don't match



Training Data Summary

- 56 of 65 CVs had results returned from CrossRef 86.2%
- 514 citations parsed
- Title_scores: mean = 0.72, median = 0.9
- CrossRef_scores: mean = 2.9, median = 2.8