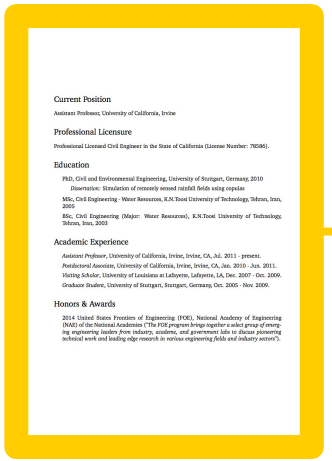
Parse CV

Pipeline



Current Position

Assistant Professor, University of California, Irvine

Professional Licensure

Professional Licensed Civil Engineer in the State of California (License Number: 78586).

Education

PhD, Civil and Environmental Engineering, University of Stuttgart, Germany, 2010
Dissertation: Simulation of remotely sensed rainfall fields using copulas

MSc, Civil Engineering - Water Resources, K.N.Toosi University of Technology, Tehran, Iran, 2005

BSc, Civil Engineering (Major: Water Resources), K.N.Toosi University of Technology, Tehran, Iran, 2003

Academic Experience

Assistant Professor, University of California, Irvine, Irvine, CA, Jul. 2011 - present.
Postdoctoral Associate, University of California, Irvine, Irvine, CA, Jan. 2010 - Jun. 2011.
Visiting Scholar, University of Louisiana at Lafayette, Lafayette, LA, Dec. 2007 - Oct. 2009.
Graduate Student, University of Stuttgart, Stuttgart, Germany, Oct. 2005 - Nov. 2009.

Honors & Awards

2014 United States Frontiers of Engineering (FOE), National Academy of Engineering (NAE) of the National Academies (*"The FOE program brings together a select group of emerging engineering leaders from industry, academe, and government labs to discuss pioneering technical work and leading edge research in various engineering fields and industry sectors"*).

Extract Sections
- Education
- Publications
- Honors & Awards
- ….

Extract Citations

crossref

doi®

Table of Results

- doi
- parsed citation
- citation counts

Extract Section Details

- Group text based on features
  - UPPER CASE
  - text size
  - left indentation
  - font
  - bold / italics
  - line space*

- Compare each group with known section names

*Note: Line spacing currently not an effective feature for identifying sections

Extract Citation Details

- Collect all text in publication section

- Parse text by identifying citation style
  - numbered / bulleted list
  - author’s name
  - indentation style
  - etc.

Example 1

- 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

- 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

- 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

- 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

- 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