

UC Recruit - CV Parser





Pipeline

Amir AghaKouchak, PhD, PE

Assoc. Professor
Dept. of Civil & Environmental Engineering
University of California Irvine
C/ENR Engineering Gateway
Irvine, CA 92697

Office: (949) 224-5155
Mobile: (949) 231-8987
Fax: (949) 231-8851
Email: amir.akhaki@uci.edu
WebSite: <http://amir-aghakouchak.org>

Current Position

Assistant Professor, University of California, Irvine

Professional Licensure

Professional Licensed Civil Engineer in the State of California (License Number: 70508).

Education

PhD, Civil and Environmental Engineering, University of Stuttgart, Germany, 2010

Dissertation: Simulation of transient natural rainfall hydrologic response

MSc, Civil Engineering - Water Resources, K.J. Somaiya University of Technology, Mumbai, India, 2005

BSc, Civil Engineering (Major: Water Resources), K.J. Somaiya University of Technology, Mumbai, India, 2003

Academic Experience

Assistant Professor, University of California, Irvine, CA, Jul. 2011 - present.

Postdoctoral Associate, University of California, Irvine, Irvine, CA, Jul. 2010 - Jun. 2011.

Visiting Scholar, University of Louisiana at Lafayette, Lafayette, LA, Dec. 2007 - Dec. 2008.

Graduate Researcher, University of Stuttgart, Stuttgart, Germany, Oct. 2005 - Nov. 2008.

Honors & Awards

2014 United States President of Engineering (PENG) National Academy of Engineering (NAE) of the National Academies ("The PEG program brings together a select group of emerging engineering leaders from industry, academia, and government to focus on advancing technical work and leading edge research in various engineering fields and industry sectors").

Extract Sections

- Education
- Publications
- Honors & Awards
-

Extract Citations

crossref



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 - 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