

Yeounoh Chung

yeounoh_chung@brown.edu

Scott Houde

scott_houde@brown.edu

Johannes Novotny

johannes_novotny@brown.edu

Erfan Zamanian

erfanz@cs.brown.edu

Goal: To provide an intuitive web site merging three main data sets together that allows potential applicants to find the right Computer Science doctoral program for them.

Citation Network Dataset¹

Extracted from DBLP, ACM and other citation sources, contains more than 4,354,534 citations. We reduced this down by faculty and transformed paper abstracts into a TF/IDF format.



Entity resolution was tricky and not all professors were in the dataset!

Faculty Professor Dataset

Generated as part of a class assignment contains 2,195 faculty from 55 different universities.² We enhanced this by adding a generated faculty ID and existing Institution ID.³

Unique Faculty ID's

IPEDS Institution ID's

Integrated Postsecondary Education Data System (IPEDS)⁴

Every US Institution is required to submit this data every year to the government. Highly accurate but required filtering and interpretation. We used Python scripts to reduce the data down to institution ID's and award levels we are interested in.

Final Dataset

TF/IDF Data

For every unigram and bigram in every abstract for our faculty we store the term, the faculty ID and the frequency, sorted from most frequent to least.

Faculty Name,
Institution, Sub-Field,
Join Year, PhD earned
Inst, etc...

All Tables by Institution ID

General Institution fields

Current Enrollment by student level, gender and race

Computer Science doctoral completions, for multiple years

Visualization and Interactive Search

Main Page



Search Results

Faculty Detail



1) Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008)*. pp.990-998.

3) Postsecondary Institutional Codes crosswalk table http://www.pesc.org/interior.php?page_id=145

2) Analysis of Over 2,000 Computer Science Professors at Top Universities , http://jeffhuang.com/computer_science_professors.html

4) National Center of Education Statistics, <http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>