

Mining the US Technologists Data

Abulitibu Tuguluke

DHI Group Inc., CU Boulder

June 18, 2022

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Title

Mining the US Technologists Data

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Team member

1 Team Member: Abulitibu Tuguluke

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Description

There are many reports saying US is (or will) face a technology workers (Technologists). Our study will mine through the US H1B data set along with Dice.com's job and candidate profiles (along with skill sets for each job title), to try to understand which technologist job are in increasing/decreasing demands, which skill sets are more popular/unpopular, through time and geo-location analysis. Our goal is to better understand the US technology market and what we can do to help people who try to get into tech market by providing them with skillset guidance.

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Prior Work

There is already a skillsets collection been done on tons of job descriptions for technologist through data mining. We have 1432 job titles associate with particular skill sets. Unfortunately, most of the online literatures are focused on salary analysis, which motivates job seeker, where we will do our own salary report as well.

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Datasets

- ▶ H1B data scraped from `https://h1bdata.info/index.php`
- ▶ Skill set (Internal data)
- ▶ US Bureau of Labor Statistics
`https://www.bls.gov/developers/api_python.htm`
- ▶ Data mined and stored on my machine

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Proposed work

- ▶ Data scraping: Scraeping HTML data from the web and store them in seperate json file
- ▶ Data cleaning: Probably the most important part of the project
- ▶ Data preprocessing:
- ▶ Data integration: Integrate internal data with mined H1b data
- ▶ Data mining and analysis

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

List of tool(s)sets

- ▶ BeautifulSoup for mining
- ▶ Python library for all the computing and analysis
- ▶ JSON for data store

Outline

Title

Team member

Description

Prior Work

Datasets

Proposed work

List of tool(s)sets

Evaluation

Evaluation

TBD. There are many ways we can evaluate the results. Comparing time series analysis data within each job we can compute the increasing or decreasing sides of each job. Location analysis on which job are located where with respect to other jobs and locations.