Mining the US Technologist data

Abulitibu Tuguluke
DHI Group Inc.
6465 Greenwood Plaza Blvd
Greenwood Village, CO 80111
abtu8803@colorado.edu

1 PROBLEM STATEMENT/MOTIVATION

(like homework 1, what knowledge and how would you apply that knowledge, what is interesting that you hope to find) 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 setsa are more popular/unpupular, 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.

2 LITERATURE SURVEY

(previous work) describe and cite.

3 PROPOSED WORK

E.g., what do you need to do for data collection, preprocessing (cleaning integrating, transforming, etc.), process for derived data, design, evaluation. Describe how it is different than what has been done previously from your literature survey (or if replicating).

- Data scraping: Screaping 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

4 DATA SET

(make sure you have the data set!). Provide URL and details about the data set (similar to homework 1, chapter 2, etc.)

- H1B data scraped from I scraped from open data online.[1] The data set is mainly from the United States Department of Labor (DOL) on how many H1-B petitions were filed and approved, with detailed information such as Employer, the title of the job, city/state locations, along with base salary, and submission and acceptance dates.
- Skill set (Internal data)
- US Bureau of Labor Statistics [2]
- Data mined and stored on my machine

5 EVALUATION METHODS

E.g., metrics, existing solutions, …
here are many ways we can evaluate the results. Comparing time series analysis data within each job we can compute the increaing or decreasing sides of each job. Location analysis on which job are located where with repect to other jobs and locations.

6 TOOLS

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

7 MILESTONES

What you plan to have done by when
The Following table is my work plan for carrying out the project.

Date	Work Plan
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.
Week 1	Submit the Research Proposal.

[1] *https : //h1bdata.info/index.php*.

[2] *https : //www.bls.gov/developers/api$_p$ython.htm*

CSCI 5502, Boulder, University of Colorado
© 2022 ACM. 978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

8 UPDATED, EXTENDED VERSION OF INITIAL PROPOSAL

9 PROPOSAL REVIEW: MOTIVATION, PROPOSED WORK, TOOLS, EVALUATION, MILESTONES

10 MILESTONES COMPLETED: WHAT YOU HAVE ACHIEVED SO FAR (IN MILESTONES SECTION AS A SUBTOPIC)

11 MILESTONES TODO: WHAT REMAINS TO BE DONE (IN MILESTONES SECTION AS A SUBTOPIC)

12 RESULTS SO FAR (NEW TOPIC AT END)

Any graphs, correlations, etc. if any