Job recommendation system based on LinkedIn API

Yutong Chen, Yuqin Zhao, Shuai Ren

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Challenges

- DATA COLLECTION: Spark Streaming on LinkedIn API
- DATA PREPOSSESING: Extract relevant data from LinkedIn profiles
- CLASSIFICATION: Using Spark ML to do classification
- DATA VISUALIZATION

- CODING!!!

Data? LinkedIn API + Spark !

1. Create my LinkedIn App on the LinkedIn Developers platform.
2. Get the Access Token I need to incorporate in my Python app.
3. Develop the script like we did in twitter API

According to LinkedIn, the dataset will include a description and category of the users' classification result.

Classification

Data Processing: split, remove stop words...

Personal Description  Job Domain

logistic regression, KNN,

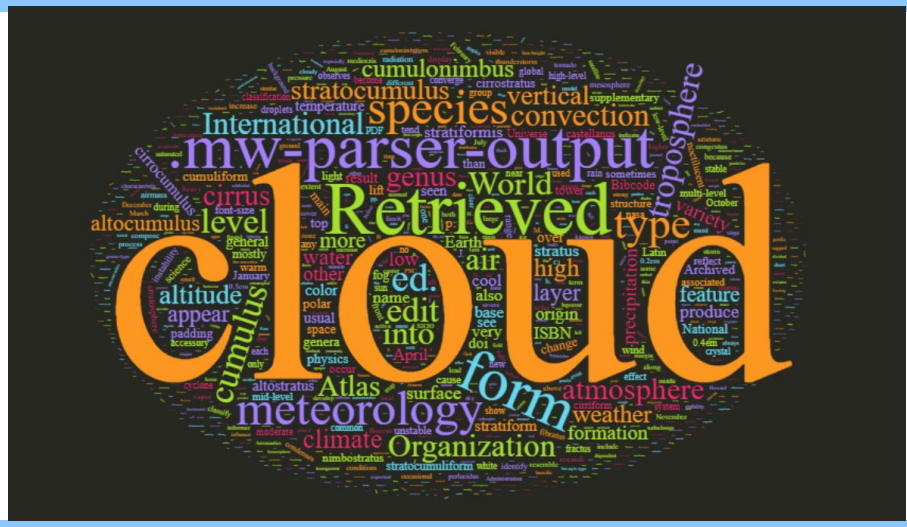
Random Forest, Naive Bayes, Gradient Boosting Trees,

Multi-layer perceptron, Linear Support Vector Machine,

One-vs-Rest

Visualization

Count the frequency of different skills appears in every domain and visualize



System Overview

Offer job recommendations to different users

Offer recommended skills to users in different domains

Schedule

- 11.05 Final Project Proposal Presentation
- 11.06-11.11 Data Collection by Yutong Chen
- 11.12-11.18 Data Prossesing by Yutong Chen
- 11.19 Final Project Progress Presentation
- 11.20-11.03 Classification by Yuqin Zhao
- 12.04-12.16 Data VisualizationI by Shuai Ren
- 12.17 Big Data Analytics Workshop