Resume Text Mining And Predicting Job Scenario
Author: Samyukta Bharadwaj | Advisor: Dr. Carol Romanowski
Rochester Institute Of Technology, Computer Science

Objectives

- With the current job scenario and intense competition in the job market, applicants are standing on thin ice between getting a job or not.
- Most of the applicants have an application rejected for reasons they are not aware of. There is this one factor that is probably missing on the resume that may have changed lives for good.
- It would be nice to know what a resume should look like in order to land a job, essentially with respect to the skills that could be enhanced for a better chance.
- This project does research in the field of text mining and data mining for parsing a bunch of resume data and predicting job scenario.

Methods

- Data Collection: Resume Data is collected from job search portal which has resumes available that are legally public with permission from candidates.
- Data Extraction: Data available in PDF format is extracted in text.
- Text Mining: NLTK, Stanford NLTK and Regular Parsing
- The text mined data was then required to be converted from JSON to CSV and loaded in Weka.
- Data Mining: K means Clustering was done in Weka to determine results.
- Evaluation: To measure how accurate was the combined effort of NLTK and Stanford NLTK in creating JSON as against when manually parsing and creating JSON. Evaluation in data mining phase is done by calculating confusion matrix (if similar data is stored in same cluster or not.)

Implementation

- The implementation revolves around data cleaning, extraction and processing. The major candidate for this is unstructured raw form of data retrieved from PDF files.
- Most of this has been done using Python. This covers the Text Mining phase. Further processing of data is done in Weka.

Conclusion

There is a drastic difference and a better performance when skills are taken into consideration. Through NLP and Stanford NLP, skills were not recognized as an important element and hence, custom parsing was required to be done. The accuracy of data text mined was higher by approximately 50 percent when parsed manually and this also yielded better clustering results.

Future Work

Improve accuracy. Create user interface and perform web crawling to track job openings according to users’ profile.

References

- Raymond J Mooney and Un Yong Nahm. 2005. Text Mining With Information Extraction
- Fabio Ciravegna. Adaptive Information Extraction From Text Using Rule Induction and Generalisation