Progress Report: Build Experience Profiles from Resumes

High-level Tasks

Please refer to the high-level tasks identified for this project in the proposal document.

Tasks Completed

> Gather representative data set for training and evaluation

This task has been completed. The dataset gathered contains resumes from multiple departments. We are confident this corpus is large enough to demonstrate the core idea of the project.

Tasks In Progress

> Parsing engine to parse resumes and job descriptions

Two members are currently working on implementing the parsing engine module that takes

- resumes and extracts information, in a batch and online mode
- job-description and extracts information in online mode for matching/searching at runtime

> Analysis engine to analyze resumes

Two members are currently working on implementing the analysis engine, using the interface defined between the parsing and analysis engine, w.r.t the output format of the parsing engine.

Tasks Pending

The following tasks are pending start:

- Scoring engine
  - Implement the Apache Lucene's TFIDFSimilarity to score the documents.
- Basic UI
  - User specifies a job-description to get the ranked resumes as response
- Documentation

Current Challenges

- A thorough analysis on how to give more weight to a skill that occurs less number of times in a resume, but the candidate worked on those skills for multiple years.