Military-to-Civilian Terminology Translation Tool: Analysis of Job Postings in Technology

Student Authors: Jairo Garciga, Ryan Smith, Aiden Ridgeway, Owen Doyle

Project Sponsor: Intevations LLC, Thomas Ritter

Course Instructors: Dr. Fernando Gonzalez and Dr. Huzefa Kagdi

FGCU Software Engineering Dept.

Introduction

- This job data project is focused on finding the current desired job skills and terminology used to describe them
- The current iteration uses Indeed job postings for Information Technology and Software Development
- Currently displaying information on an AWS hosted website
- This sub-project that is part of a veteran resume translation tool that is designed to automate the process of displaying veteran experience in civilian terminology for resume use
- The current desired skills and terminology information gained from this research project is crucial to the success of the greater MC-TTT system

Process and Structure

- For this project we used an Agile Methodology. Requirements were kept in a Jira backlog alongside constraints and other critical information.
- **Key Features**
  - Gather current vocabulary from technology job postings
  - Display the most popular words from current technology job postings
- **Architecture and Design**
  - Cloud service provider must be AWS
- **Technology Stack**

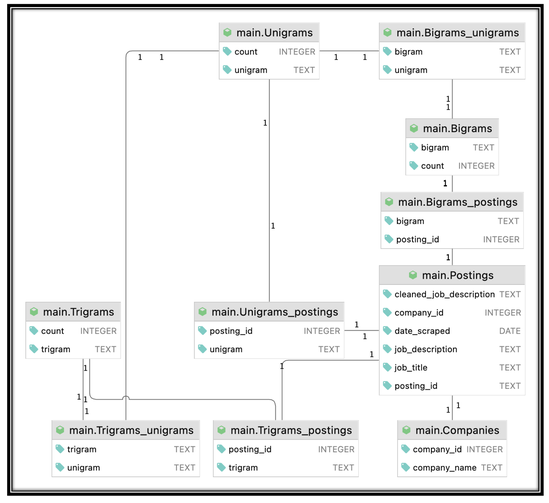
Data

- In order to gather the data from job postings, we created a web scraper for Indeed.
- The scraper grabs job postings and then inserts the posting id, the posting title, the company for the posting, and the job description.
- Once the data is in the Database, we run several scripts that process the data.
  - The data is first cleaned and removed of unnecessary characters that are not in Unicode, stop words, and most punctuation.
  - Then all unigrams from the entire set of job posting descriptions are gathered alongside their count.
  - Once this is complete, bigrams and trigrams from the entire set of job postings are also gathered and linked to the specific unigrams that make them up alongside the count.

Top 5 N-grams

Rank	Unigrams	Bigrams	Trigrams
1	Experience	bachelor degree	degree computer science
2	Or	computer science	bachelor degree computer
3	Skills	communication skills	qualifications bachelor degree
4	Ability	years experience	computer science or
5	Work	information technology	or related field

Database Outline



Lessons and Future Work

- One of the biggest issues we came across was how to access the data. In the past, Indeed had an API for accessing their job data. However, they no longer have it available. If there is a public job API that should be used instead since a scraper would have to change if the webpage changes
- The services we used are in the free-tier of AWS. This created some bottlenecks with the Database's ability to take in requests. In the future, a database with a faster CPU should be used. This would allow for more data to be processed at a faster rate.
- Our work ended at the data gathering side. Further work should look to use the data we have collected to begin the translation part of the project.

Conclusion

- The overall project was a great learning experience for us all, we had the learn and use Python, Flask, Jira, AWS, and Google Colab
- We did a week-long Agile sprint with daily Scrum meetings; this gave good insight into learning and understanding Agile workflows with practical experience that supported our previous knowledge.
- There are still many parts to complete the entire MC-TTT project, but the job data collection is complete and running strong.