CSCE 5290: Natural Language Processing
Project Proposal

Project Title
Resume and Job Posting Keyword and Skill Extractor

GitHub Link
https://github.com/Zedoumous/NLP-5290---Project-Group-17

Team Members
Tyler Parks - tylerparks@my.unt.edu
George Thomas - georgethomas2@my.unt.edu
Jofiel Gomez - jofielgomez@my.unt.edu
Tyrell Richardson - TyrellRichardson@my.unt.edu

Project Description
As part of the NLP CSCE5290 course requirements, our team will work on the development, integration, testing, and deployment of a program intended to process resumes and job postings and extract the keywords and skills within those mediums. This program is intended to present to the user keywords in a manner that is efficient and accurate in providing the user useful information about skills or other conveniences that would eliminate the hassle of a user identifying them on their own. The project will feature multiple techniques and methods of information extraction and processing researched and used in the NLP 5290 course.

Goals and Objectives

1. **Motivation**
   Navigating the job market is difficult as a new graduate. There is a misalignment between academic coursework and expectations from hiring managers. This project is meant to parse through job applications to extract the key skills a student needs to be successful. This allows students to use this key information to expand their studies to include these skills and allow professors to get a deep understanding of the current job market and plan their coursework accordingly.

2. **Significance**
   There are billions, if not trillions of webpages, resumes, files, etc. with key information attached to them. Most of these files/pages, often hold information that can be considered unimportant to a user if they are looking for specific information. For this project case specifically, skill extraction from web pages like job postings or from resumes, can be useful for skipping irrelevant or overwhelming information. For example, if an employer is looking to hire about 10 employees out of 100 applicants. It would be inefficient to scroll through every file to look for the section that lists the employee's skill. The employer would have to do this 100 times. If the employer had a program that could go through an employee's resume and instantly extract information about the skills on their resume, the employer could instantly determine if they are a candidate for the job much sooner, and more

efficiently. That is the significance of the project. By creating a program that extracts information quickly, efficiently, and accurately, we can save a lot of users time and effort.

3. Objectives

a. To develop a program that will serve as an accessibility tool for users who wish to obtain and analyze information at an efficient or accelerated rate.
b. To understand how similar computer program algorithms/architecture operate in web applications for information extraction.
c. Display and implement the understanding of skills and concepts learned in the course. The completed project will be a validation of progressive mastery of NLP techniques.
d. To understand the development flow of a project within a team. Team members will experiment and standardize communication and development workflows that will better assist in understanding how to develop a project efficiently.

4. Features

a. Resume and Job Posting Identification (Required)
Based on input text or files, our program should be able to determine whether that input resembles a resume or job posting. Depending on the type of document, our parsing and extraction methods can be tailored to that type of document specifically.

b. Information Extraction (Required)

i. Skills
Our program should extract identifiable skills from the input and return those skills. For resumes, the skills our program extracts would be skills that the resume's writer possesses. For job postings, any skills extracted would be skills that companies require from a prospective employee.

1. Skill Classification
Once a skill set has been extracted from either a resume or job posting, our program should classify and/or cluster the skills into appropriate groups. These clusters can lend insights in skill distribution and prominence.

ii. Metadata
Both resume and job postings contain metadata that will be useful to collect and use for skill extraction. This includes information about a resume's owner, retrieved from the resume itself, and company/location information from a job posting.

c. Resume Matching (Optional)
i. When a resume is parsed and skills collected, that resume could be compared against similar parsed job postings that are stored in our program. If there is a job posting that matches closely enough to a user-inputted resume, our program could suggest that job posting to the user.

References

A. A. Awan, “Spacy resume analysis,” *Deepnote*, 27-Oct-2021. [Online]. Available: https://deepnote.com/@abid/spaCy-Resume-Analysis-81ba1e4b-7fa8-45fe-ac7a-0b7bf3da7826. [Accessed: 07-Oct-2022].

D. Moore, “Using NLP to improve your resume,” *Medium*, 16-Feb-2021. [Online]. Available: https://towardsdatascience.com/ai-is-working-against-your-job-application-bec65d496d22. [Accessed: 07-Oct-2022].

“Lightcast skills,” *Lightcast Skills*. [Online]. Available: https://skills.lightcast.io/extraction. [Accessed: 06-Oct-2022].