Project Outline & Abstract

Project Title: Automated Keyword Extraction from Job Descriptions

Team Members:
1. Naveen Kumar Kundla - 016021941
2. Sivakrishna Yaganti - 017428853
3. Vinit Kanani - 016651323

Abstract

Applicant tracking systems (ATS) are commonly used by employers to automatically scan resumes for relevant keywords from job descriptions. However, many qualified candidates may be overlooked if their resumes lack explicit matches for keywords. This project aims to automate the extraction of key skills and qualifications from job descriptions using natural language processing. The extracted keywords can then be used to optimize resumes so they contain relevant terms that will enable candidates to successfully pass initial ATS screening. By improving keyword matching between applicant resumes and job descriptions, this system can help job seekers better tailor their resumes while enabling employers to more accurately identify qualified applicants that may have been previously overlooked. Overall, automated keyword extraction will make the initial resume screening process more efficient for both applicants and employers.

Project Outline

1. Introduction

    a. Problem Statement

    b. Motivation

    c. Related Work

    d. Proposed Solution

2. Data Collection

    a. Scrape job descriptions from greenhouse.io

    b. Kaggle dataset Link

3. Data Preprocessing & Cleaning for scraped data

    Extract plain text from HTML pages through parsing and stripping HTML tags, scripts, styling, and other non-text content.

4. Model Training

      a.   Pre-trained model:
- Spacy Models
- Standford NLP Model

      b.   Fine-tuned model:
- Fine-tune annotations for scraped data
- Train model on fine-tuned data
- Evaluate model performance

5. Results

      a.   Compare performance of pre-trained and fine-tuned models

      b.   Compare performance of different models

6. Conclusion

      a.   Summary of results

      b.   Future work

7. References

[1] Lison, J. Barnes, and A. Hubin, "skweak: Weak Supervision Made Easy for NLP," arXiv (Cornell University), Jan. 2021, doi: https://doi.org/10.18653/v1/2021.acl-demo.40.

[2] "Application of Neural Network Keyword Extraction Methods for Student's CV Compilation from Discipline Work Programs | IEEE Conference Publication | IEEE Xplore," ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/10159061 (accessed Nov. 22, 2023).

[3] "Training Pipelines & Models · spaCy Usage Documentation," Training Pipelines & Models. https://spacy.io/usage/training

[4] "The Stanford Natural Language Processing Group," nlp.stanford.edu. https://nlp.stanford.edu/software/CRF-NER.html

[5] M. Ghadge, "Building Your Own Custom Named Entity Recognition (NER) Model with spaCy V3: A Step-by-Step Guide," Medium, Sep. 06, 2023. https://medium.com/@mjghadge9007/building-your-own-custom-named-entity-recognition-ner-model-with-spacy-v3-a-step-by-step-guide-15c7dcb1c416 (accessed Nov. 22, 2023).

Dataset

The data for this project will consist of 10000+ job descriptions scraped from greenhouse.io, a popular recruiting software platform. Additionally, a dataset of over

160,000 cleaned annotated job descriptions from Kaggle might be utilized to evaluate the models.

Ethical considerations

Global Impact

The project would help job-seekers to better tailor their resumes resulting in more efficient job search. This would also benefit employers by enabling them to more accurately identify qualified applicants that may have been previously overlooked. But, the project can also be misused to bypass the initial screening process and get unqualified candidates through the door.

Economic Impact

This project would benefit both the job-seekers and employers by making the initial resume screening process more efficient. However, if the candidates misuse the project to bypass the initial screening process, it would result in higher costs for employers.

Environmental Impact

The project would not have any significant environmental impact. The training of models would require computing resources, but the impact would be negligible.

Societal Impact

The project can have a significant positive impact to the well qualified candidates. However, the project can also be misused to deceive the employers.