INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY BANGALORE

NATURAL LANGUAGE PROCESSING

AI 829

Resume Analysis

Kartik S Gawande

MT2023045

kartik.gawande@iiitb.ac.in

PROBLEM STATEMENT

Developing an NLP-driven tool for analyzing resumes, extracting relevant skills, experiences, and qualifications, and matching candidates to job descriptions.

PROBLEM IMPORTANCE

Streamlines recruitment by quickly identifying suitable candidates, reducing manual review time, and enhancing the matching accuracy between job requirements and applicant skills.

TECHNOLOGIES USED

spaCy library for entity recognition to identify and categorize skills and job roles within resumes.

Gensim library for topic modeling and the NLTK library for text preprocessing, including lemmatization and stopword removal.

Visualization tools such as PyLDAvis and Plotly to present findings and insights derived from the data.

PROCESS FLOW

The process begins with data collection, where resumes and skill patterns are gathered. Next, the spaCy library is used to load the resumes and apply an entity ruler for skills and job categories recognition. Text data is then cleaned and preprocessed using regular expressions and NLTK, followed by the extraction of unique skills from each resume. The extracted data is visualized to identify trends and distributions in job categories and skills. Finally, the resumes are analyzed for skill matches and topic modeling is performed to uncover underlying themes and patterns.

i.e.

Resume ingestion → Text preprocessing → Skill and entity extraction → Matching against job descriptions → Candidate ranking

MANDATE 2 (Lexical Processing)

- Implement text preprocessing techniques, including stopword removal, lemmatization, and the use of regular expressions to clean resume texts.
- Utilize keyphrase extraction and phrase identification models to highlight important skills and qualifications.
- Apply semantic and phonetic hashing to deal with spelling variants and errors in resume texts.
- Explore distributional semantics to understand the context and meaning of words within resumes.

MANDATE 3 (Syntactic Processing)

- Implement shallow parsing and POS tagging to analyze the grammatical structure of resume texts.
- Use dependency parsing to understand the relationships between words, aiding in the accurate extraction of skills and experiences.
- Explore probabilistic parsing and HMMs for more accurate syntactic analysis

MANDATE 4 (Semantic Processing)

- Apply named entity recognition to identify and categorize entities such as skills, job titles, and educational qualifications.
- Utilize topic modeling techniques (e.g., LDA) to discover latent topics within resumes, helping to understand common skill sets and experiences.
- Implement word sense disambiguation and masked language models to improve the accuracy of entity recognition and the relevance of skill matches.

REFERENCES:

https://deepnote.com/@abid/spaCy-Resume-Analysis-81ba1e4b-7fa8-45fe-ac7a-0b7bf3da7826

https://www.analyticsvidhya.com/blog/2021/06/resume-screening-with-natural-language-processing-in-python/