Follow-up
(CSE499 - Capstone Project)

CV Analyzer: An NLP-Driven System for HRM Optimization.

Presented By

Imran Nasir Bappy (201014041)

Rownak Tasneem Shishir (202014005)

Sadia Jaman(202014009)

Nayeema Rahman Bhuiyan Sawda(202014015)

Supervised By

Raihan Kibria
Senior Lecturer
Department of Computer Science and Engineering (CSE)

July 23, 2023

Motivation

- Streamline the recruitment process by extracting, organizing and ranking CVs efficiently.
- Reducing manual effort and ensuring fair evaluation the system will save time, improve accuracy
- Help organizations to identify the best candidates based on their specific requirements.

Objectives

- Develop a system that automatically extracts information from CVs in various formats (PDF, Word).
- Rank CVs based on criteria such as experience, skills, CGPA and completed projects.
- Enable HR professionals to search and filter CVs according to specific requirements (Years of experience, skills, CGPA).
- provide focused views of CVs, showing only the information relevant to the search criteria.
- Save time, ensure better hiring decisions and reduce costs in the recruitment process.

Problem Statement

- Recruitment processes involve receiving hundreds or thousands of CVs which makes manual review impractical and inefficient.
- Manual review leads to delays, increased costs and the risk of overlooking talented candidates.
- Existing manual and outsourced solutions are often inaccurate and not cost-effective.

There is a need for an automated NLP-driven system to:

- Process, rank and organize CVs efficiently.
- Allow HR professionals to search and extract relevant information quickly.

ULAB
UNIVERSITY OF LIBERAL ARTS BANGLADESH

Literature Review

Title	Problem identification	Methodology	Research Gap	Results and Findings
1)An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems	The paper highlights the need for efficient stemming algorithms and the challenges of accuracy, performance, and language-dependency across different approaches.	The paper reviews and compares stemming techniques, analyzing their strengths, weaknesses, and performance, while discussing design and evaluation challenges.	The paper focuses on stemming for basic information retrieval but does not cover advanced extraction and ranking from CVs based on criteria like experience, skills, and CGPA. Your project requires techniques like named entity recognition and machine learning-based ranking.	The paper concludes that no stemming algorithm works best for all cases. Effectiveness depends on language and task, and hybrid approaches may help. More research is needed for better solutions.
2)Study on Information Extraction Methods from Text Mining and Natural Language Processing Perspectives.	The paper highlights the growing need for Information Extraction (IE) to make sense of unstructured text data and enable knowledge discovery.	The paper reviews TM and NLP techniques in Information Extraction and briefly mentions QA, without proposing new methods or results.	The paper focuses on general Information Extraction, but your project addresses specialized CV parsing, multi-criteria ranking, and targeted information retrieval within resumes.	The paper concludes that TM and NLP are useful for extracting information from unstructured data and suggests further research to improve IE for complex tasks.

3) Natural Language Processing (NLP) based Text Summarization - A Survey.	The difficulty of automatically summarizing large amounts of text. Different types of summarizations (extractive vs. abstractive) and learning approaches (supervised, unsupervised, reinforcement) are explored	Includes graph-based methods, latent variable models, and deep learning approaches like recurrent neural networks and transformers. Different learning strategies: supervised, unsupervised and reinforcement learning.	primarily focuses on summarizing textual content, ranking sentences based on relevance, query-based summarization our project needs ranking entire CVs based on multiple user-defined criteria, accurately extract and categorize this structured data for efficient searching and ranking.	The findings are that different methods have strengths and weaknesses. The findings are that different methods have strengths and weaknesses. Extractive methods are simpler but can produce disjointed summaries. Abstractive methods can generate more fluent summaries but are more complex and prone to errors.
4) Text Summarizing Using NLP	the problem of information overload, specifically the difficulty in manually summarizing large amounts of textual data, proposes automatic text summarization as a solution, preserve the core meaning and content of the original text.	uses an extractive text summarization approach based on the Text Rank algorithm. uses an extractive text summarization approach based on the Text Rank algorithm, iteratively calculates a score for each sentence	The paper focuses on single-document text summarization using extractive methods, specifically the Text Rank algorithm, explores different graph representations (weighted, unweighted, directed,	Demonstrates the Text Rank algorithm, implemented with the Genism library in Python, can effectively generate extractive summaries of text. They show example outputs where the algorithm correctly identifies and extracts the most important sentences from a sample paragraph.

		based on its connections to other sentences, ultimately selecting the highest-scoring sentences to form the summary.	undirected), doesn't address the core challenge of our project: multi-document summarization and ranking of CVs based on specific criteria.	
5)COMPREHENSIVE OVERVIEW OF NAMED ENTITY RECOGNITION: MODELS, DOMAIN-SPECIFIC APPLICATIONS AND CHALLENGES	Problem Identification: The paper identifies the general problem of extracting structured information from unstructured text, which is crucial for various NLP tasks, including NER. It also highlights domain-specific challenges in areas like finance and biomedicine.	The paper provides a comprehensive overview of AI techniques for NER, including rule-based approaches, supervised learning, unsupervised learning, deep learning (RNNs, CNNs, LSTMs, Transformers like BERT), reinforcement learning, and the integration of OCR. It also discusses domain-specific adaptations and challenges.	The paper explores various NER applications but doesn't address the specific challenge of ranking and filtering CVs based on extracted entities. Besides, the paper doesn't explore how NER can be combined with ranking algorithms or faceted search interfaces for efficient CV shortlisting.	The paper highlights the effectiveness of transformer models like BERT for NER and the potential of combining different techniques like OCR and reinforcement learning. It also emphasizes the need for domain-specific adaptations and further research to address open challenges in NER.

Methodology

Data Collection:

Gather sample CVs in various formats (PDF, Word, etc.) to train and test the system.

Preprocessing:

Extract text from CV files and clean the data by removing unnecessary elements like formatting artifacts or irrelevant content.

Information Extraction:

Use NLP techniques to identify and extract key details such as name, contact information, skills, experience, education, CGPA and projects.

Ranking System:

Develop an algorithm to rank CVs based on criteria like experience, skills, CGPA and number of projects.

ULAB
UNIVERSITY OF LIBERAL ARTS BANGLADESH

Methodology

Search Functionality:

Implement a search feature that allows HR to filter and find candidates by specific requirements (Skills, years of experience or CGPA).

User Interface:

Design a user-friendly interface where HR can upload CVs, view ranked lists and access detailed or focused views of candidate information.

Testing and Evaluation:

Test the system with real-world CVs to ensure accuracy, efficiency, usability and refine based on feedback.

ULAB
UNIVERSITY OF LIBERAL ARTS BANGLADESH

Expected Results

- A fully functional CV Analyzer system that can automatically extract, organize, and rank CVs based on criteria such as experience, skills, CGPA and completed projects.
- A search feature that allows HR to filter CVs according to specific requirements such as years of experience or skill sets.
- A user-friendly interface where HR professionals can view entire CVs or specific information without reading the whole document.

Resources and References

1) A. Jabbar, S. Rehman, M. B. Taimuri, S. Ali Bana and T. Saba, "An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems," in IEEE Access, vol. 11, pp. 153501-153532, 2023, doi: 10.1109/ACCESS.2023.3326472.

2) F. Soleimanian Gharehchopogh and Z. Abbasi Khalifehlou, "Study on Information Extraction Methods from Text Mining and Natural Language Processing Perspectives," AWERProcedia Information Technology & Computer Science, vol. 1, pp. 1321-1327, 2012.

3) Awasthi, I., Gupta, K., Anand, S. S., Bhogal, P. S., & Soni, P. K. "Natural Language Processing (NLP) based Text Summarization - A Survey." 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1310-1317. DOI: 10.1109/ICICT50816.2021.9358703.

4) Kumar, G. V., Yadav, A., Vishnupriya, B., Lahari, M. N., Smriti, J., & Reddy, D. S. (2021). Text Summarizing Using NLP. In Recent Trends in Intensive Computing (pp. 60-67). IOS Press.

5) Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.