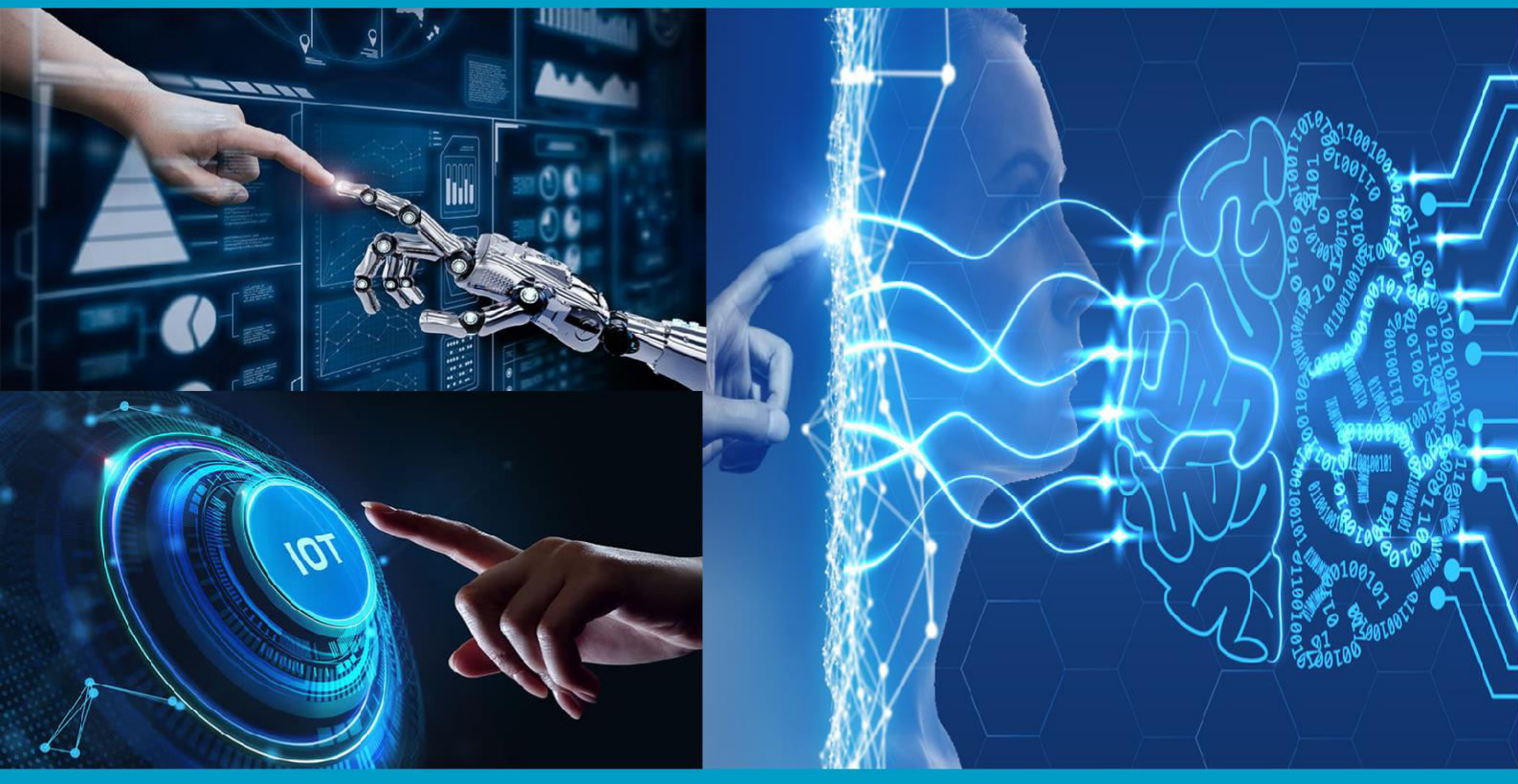
IJMRSETM

e-ISSN: 2395 - 7639



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 10, Issue 6, June 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.580

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 6, June 2023 |

Resume Parser Using Natural Language Processing

Dr. T.Geetha, Jayashree.J

Head, Department of MCA, Gnanamani College of Technology, Namakkal, India

Student, Department of MCA, Gnanamani College of Technology, Namakkal, India

ABSTRACT: Agencies and various high-level firms must deal with a large number of new jobs seeking people with various resumes. However, managing large amounts of text data and selecting best fit candidate is more difficult and time-consuming. This paper provide is an overviewof an ongoing Information Extraction System project that helps recruiters in identifying the best candidate by extracting relevant information from the resume.This project presents a system that uses Natural Language Processing (NLP) techniques to extract minute data from a resume, such as education, experience, skills, and experience. The recruiting process is made easier and more efficient by parsing the resume. The proposed system is made up of three modules: an administration management system, File upload and parser system, and an information extraction system. The administrator will upload the applicant's resume into the system, and the relevant information will be extracted in a structured format. Using the parsed information from the Resume, HR can select the best candidate for the job based on the company's needs.

KEYWORDS: Resume parser, resume analyzer, text mining, natural language processing, resume JSON, semantic analysis

I.INTRODUCTION

Daily, corporate firms and recruiting agencies have to process a large number of resumes. Working with a large volume of text data is usually time consuming and stressful Data gathered from different resumes can be in a various form, including.pdf, And these formats might not be suitable for the particular application. So, questions may arise in our mind that, what is resume parsing? The process of converting the unstructured form (.pdf) of resume data into a structured format is known as resume parsing. Subsequently, converting a resume into prepared text or structured information makes studying, analyzing, and comprehending easier. As a result, many organizations and institutions depend on Information Extraction, where unstructured data and vital information are extracted and converted to make information more readable and organized data forms. The completion of this task takes a long time for humans. So, it is necessary to develop an automated intelligent system that can extract all relevant information to determine whether an applicant is suitable for a particular job profile. The foundation of this project is a resume automation system.

II. LITERATURE SURVEY

A. NLP Based Extraction of Relevant Resume using Machine Learning: This technique stated parsing of the resumes with least limit and the parser works the utilization of two or three rules which train the call and address. Scout bundles use the CV parser system for the determination of resumes. As resumes are in amazing arrangements and it has different sorts of real factors like set up and unstructured estimations, meta experiences, etc. The proposed CV parser approach gives the component extraction method from the moved CV's.

B. E-Recruitment System Through Resume Parsing, Psychometric Test and Social Media Analysis: It follows an approach of 4 stages, the first stage was to get the data (resume) and convert them into structured format and then perform the analysis using deep learning techniques. Second step includes the psychometric test where the text mining is used to generate scores for each candidate. In the third step they perform web scraping on various social media sites to get the additional information about the candidates and recommend suitable jobs to them. In the fourth step, the system will recommend the skills and requirements in which the students are lacking and also help them to get recruited in the desired company.

III.EXISTING SYSTEM

In the existing system of resume checking by HR, the process was typically manual and time-consuming. HR professionals would physically review each resume that was submitted for a job opening and evaluate them based on

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 6, June 2023 |

specific criteria and qualifications. The process involved reading through the resumes, highlighting relevant information, and making subjective judgments on the suitability of candidates. During this stage, HR professionals would primarily focus on the following aspects of a resume:

Formatting and presentation: They would check if the resume followed a clear and organized format, had proper section headings, and was visually appealing. Attention would be paid to the effective use of fonts, bullet points, and overall readability.

Contact information: HR would ensure that the resume contained accurate contact details, suchas the candidate's name, address, phone number, and email address, to facilitate communication.

Objective or summary statement: They would review the objective or summary statement at the beginning of the resume to gain an initial understanding of the candidate's career goals and aspirations.

Education: HR would assess the candidate's educational background, paying attention to the institutions attended, degrees earned, and any relevant certifications or honors.

Work experience: They would evaluate the candidate's work history, looking for relevant job positions, company names, employment dates, and key responsibilities. Attention would be given to the alignment of the candidate's experience with the requirements of the job opening.

Skills: HR would review the skills section of the resume, focusing on the candidate's abilities, competencies, and proficiencies that were relevant to the job.

IV. PROPOSED SYSTEM

Input interface: Design a user-friendly interface to allow users to upload resumes in various formats, such as PDF.

Resume parsing engine: Develop a robust parsing engine that can accurately extract information from resumes. Implement algorithms and techniques to handle different resumeformats,layouts,and languages. The parser should be capable of extracting key details like personal information,work experience, skills.

Data extraction and structuring: Extract the relevant information from the resumes and structure it into a standardized format or schema. This ensures consistency and enables easy retrieval and analysis of the parsed data.

Entity recognition and normalization: Implement algorithms to recognize and normalize entities like names, addresses, phone numbers, and email addresses. This helps ensure the accuracy and consistency of the extracted information

Skill and keyword extraction:

Develop algorithms to identify and extract skills, keywords, andindustry-specific terms mentioned in the resumes. This enables efficient filtering and matching of candidates based on desired skills or qualifications.

Error handling and validation: Implement error handling mechanisms to address parsing errors and inconsistencies. Perform data validation checks to ensure the accuracy and completeness of the extracted information.

Integration capabilities: Provide options for integration with other systems, such as applicant tracking systems (ATS), HR databases, or job portals. This allows seamless transfer of parsed resume data and facilitates the overall recruitment workflow.

Security and data privacy: Implement robust security measures to protect the confidentiality and integrity of the uploaded resumes and parsed data. Comply with data privacy regulations andensure that sensitive information is handled securely.

User management and access control: Implement user management features to control access and permissions for different users or roles within the system.

Scalability and performance: Design the system to handle a large volume of resumes efficiently,ensuring scalability and optimal performance. It's important to note that developing a comprehensive resume parser can be a complex task. It may involve leveraging existing natural language processing (NLP) libraries, machine learning techniques, or even integrating with third- party services or APIs to enhance parsing capabilities.

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 6, June 2023 |

V. SYSTEM OVERVIEW

Agencies and different high-level companies have to deal with an extreme number of new jobs seeking employees with different resumes. However, looking after those large numbers of text data and filtering out the needed candidates is a burden on the brain and moretime consuming. Therefore, the essence of this literature review is on studying resumes in different formats such as single-column resumes, double-column resumes with extension.pdf,.docx, and how the suggested Information Extraction System converts that unstructured information into structured layout through Parsing.Accordingly, this review also helps to understand and apply several in-use and well recognized algorithms currently being used in industries to reduce human labor. Depending upon the Company's preference to hire employees, the Extraction System will manage the gathered information with more readability and organized data forms. Furthermore, the analysis of various machine learning algorithms and natural language processing techniques would be equally carried out along with their proper implementation and evaluation. The reviews from multiple research publications and journals are included below.

VI. SYSTEM IMPLEMENTATION

Front End:
The front end of the college bus transport management web application is responsible for presenting the user interface to the end-user and receiving input from them. The front-end of the application is developed using the following technologies:

HTML:
HTML is the standard markup language used for creating the structure and content of web pages. It consists of a series of tags that define the elements of a webpage, such as headings, paragraphs, images, links, tables, and more. HTML provides the basic building blocks for organizing and presenting content on the web. It focuses on the structure and semantics of the webpage rather than its visual appearance.

CSS:
CSS is a stylesheet language used for describing the presentation and style of a document written in HTML. It allows you to control the layout, colors, fonts, and other visual aspects of a webpage. With CSS, you can separate the design from the content, enabling you to apply consistent styles across multiple web pages. CSS works by selecting HTML elements and applying rules to them, specifying how they should be displayed in the browser.

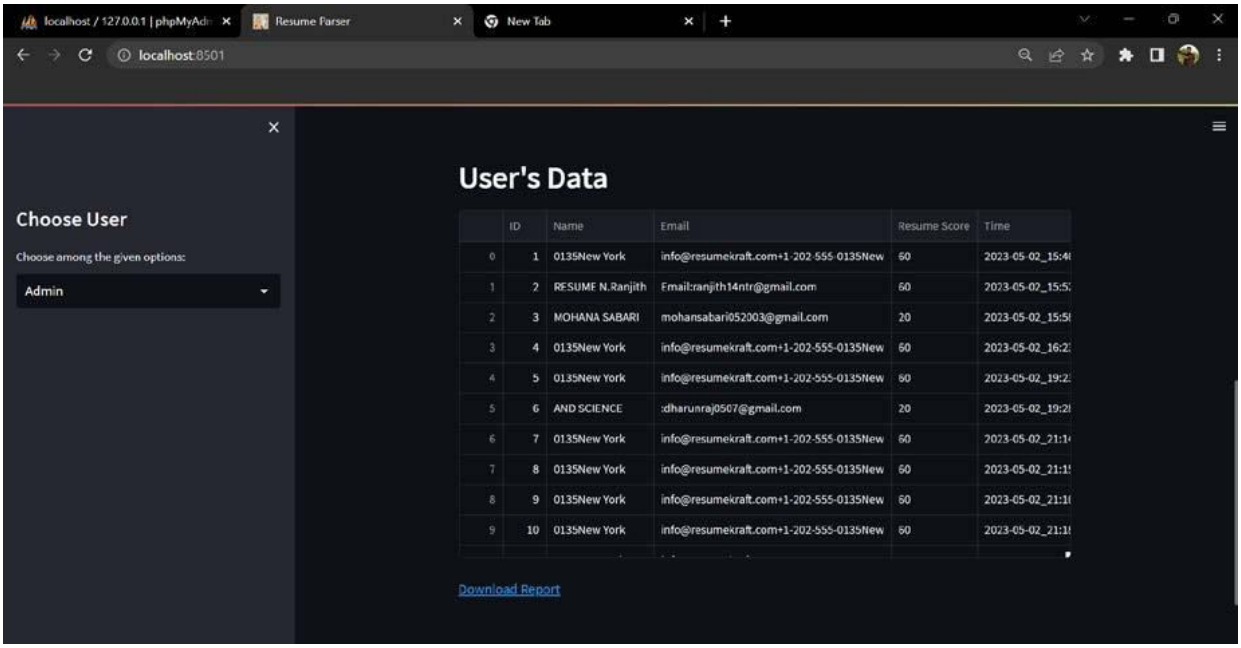
PYTHON:
Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception
When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

Back End: The back end of the college bus transport management web application is responsible for handling the business logic, database operations, and server-side processing. The back-end of the application is developed using the following technologies:

Streamlit :
It is an open-source app framework for Machine Learning and Data Science teams. You can create beautiful data apps in hours. All in pure Python. Streamlit was released in October 2019 and was recently acquired by Snowflake. There's huge excitement about it in the Data Science community. It's not just for Data Science, though. With its component extensibility architecture, you can build and integrate most kinds of web frontends into Streamlit apps. My experience with Streamlit can be verified on the official page of Streamlit Creators.Every single data analysis

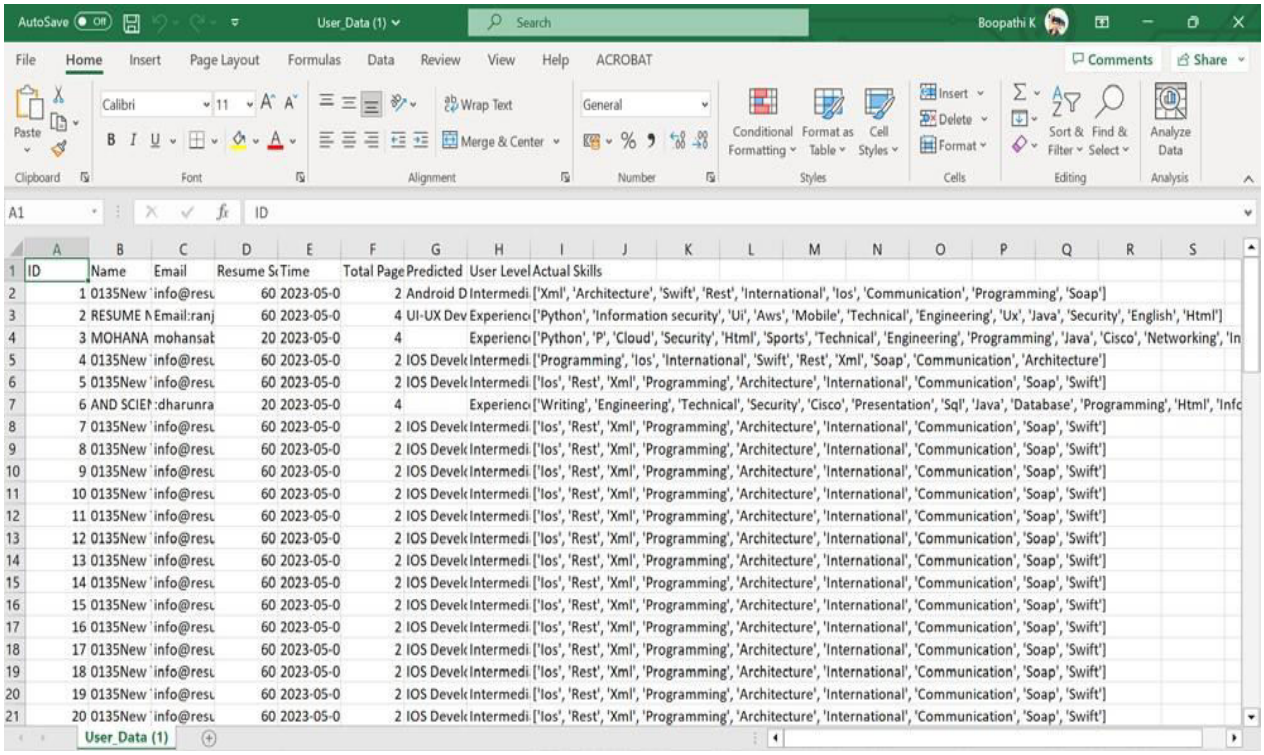
team needs to create apps. They're a focal point - like the bonfire of the team. It's where team members get together and communicate.Apps are a crucial part of the ML (data analysis) workflow, especially in a non-trivial project. This applies not only to internal apps. Machine learning researchers and data scientists need to build apps for external consumption too. Other teams need to consume models in various different ways, and it ought to be much easier to build the required but different application layers to dothat.

MySQL: MySQL is a fast, easy-to-use RDBMS being used for many small and big businesses. MySQL is developed, marketed and supported by MySQL AB, which is a Swedish company. MySQL is becoming so popular because of many good reasons. MySQL is released under an open-source license. So you have nothing to pay to use it.MySQL is a very powerful program in its own right. It handles a large subset of the functionality of the most expensive and powerful database packages.MySQL uses a standard form of the well-known SQL data language.MySQL works on many operating systems and with many languages including PHP, PERL, C, C++, JAVA, etc.MySQL works very quickly and works well even with large data sets.MySQL is very friendly to PHP, the most appreciated language for web development.MySQL supports large databases, up to 50 million rows or more in a table. The default file size limit for a table is 4GB, but you can increase this (if your operating system can handle it) to a theoretical limit of 8 million terabytes (TB).MySQL is customizable. The open-source GPL license allows programmers to modify the MySQL software to fit their own specific environments.MySQL is a highly scalable product and that scalability can come from several different performance tuning techniques. For starters, you can tune MySQL from the application level. Using a product like Redis which is also supported by OpenLogic, you can cache database queries in an in-memory database. This technique works well with databases containing a high read level and a low write level. An example would be queries for static content on your site.



| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 7.580| A Monthly Double-Blind Peer Reviewed Journal |

| Volume 10, Issue 6, June 2023 |



ID	Name	Email	Resume Sc	Time	Total Page	Predicted	User Level	Actual Skills
1	0135New	info@resu	60	2023-05-0	2	Android D	Intermedi	['Xml', 'Architecture', 'Swift', 'Rest', 'International', 'Ios', 'Communication', 'Programming', 'Soap']
2	RESUME N	Email:ranj	60	2023-05-0	4	UI-UX Dev	Experienc	['Python', 'Information security', 'Ui', 'Aws', 'Mobile', 'Technical', 'Engineering', 'Ux', 'Java', 'Security', 'English', 'Html']
3	MOHANA	mohansat	20	2023-05-0	4		Experienc	['Python', 'P', 'Cloud', 'Security', 'Html', 'Sports', 'Technical', 'Engineering', 'Programming', 'Java', 'Cisco', 'Networking', 'In
4	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Programming', 'Ios', 'International', 'Swift', 'Rest', 'Xml', 'Soap', 'Communication', 'Architecture']
5	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
6	AND SCIEN	dharunra	20	2023-05-0	4		Experienc	['Writing', 'Engineering', 'Technical', 'Security', 'Cisco', 'Presentation', 'Sql', 'Java', 'Database', 'Programming', 'Html', 'Infc
7	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
8	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
9	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
10	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
11	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
12	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
13	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
14	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
15	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
16	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
17	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
18	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
19	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
20	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']
21	0135New	info@resu	60	2023-05-0	2	IOS Develc	Intermedi	['Ios', 'Rest', 'Xml', 'Programming', 'Architecture', 'International', 'Communication', 'Soap', 'Swift']

VII. FUTURE ENHANCE

In the future, resume parsers are likely to continue evolving and incorporating new features and technologies to improve their functionality and accuracy. Here are some potential future developments for resume parsers:Semantic understanding: Future resume parsers may employ advanced natural language processing (NLP) techniques to better understand the semantics and context of the information provided in a resume. This could involve capturing the nuances of job descriptions, skills, and achievements more accurately.Contextual relevance: Resume parsers could become more adept at assessing the contextual relevance of the information provided. They might consider the specific requirements of the job being applied for and analyze the candidate's qualifications accordingly, rather than relying solely on keyword matching.

VIII. CONCLUSION

In conclusion, a resume parser is an invaluable tool in the recruitment process that helps streamline and automate the initial screening and evaluation of job applicants. It is a software program or algorithm designed to extract and analyze relevant information from resumes, such as contact details, educational qualifications, work experience, skills, and other pertinent data.Resume parsers use natural language processing (NLP) techniques to parse resumes in various or output. This structured data can be further utilized for applicant tracking, ranking, filtering, and matching against job requirements.

REFERENCES

1. Ganapathi, A., &Rambow, O. (2018). ResumeRanker: Automatic Skill Extraction forTechnical Resumes. In Proceedings of the 2018 Conference of the North AmericanChapteroftheAssociationforComputationalLinguistics:HumanLanguageTechnologies(pp.1519-1529).Retrievedfromhttps://www.aclweb.org/anthology/N18-1153.pdf

2. Zhao, W., Wang, X., & Zhang, B. (2020). A Lightweight and Extensible ResumeParsing Framework. In Proceedings of the 59th Annual Meeting of the Association forComputationalLinguistics:SystemDemonstrations(pp.231-237).Retrieved from https://www.aclweb.org/anthology/2021.acl-demo.28.pdf

3. Shekar,R.,&Pimpalkhute,P.(2019).ResumeParsingandAnalytics:AComprehensiveReview. International Journal of Engineering Research & Technology, 8(7), 317-323.Retrievedfromhttps://www.ijert.org/research/resume-parsing-and-analytics-a-comprehensive-review-IJERTV8IS070294.pdf

Impact Factor: 7.580

INTERNATIONAL JOURNAL
OF MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com