Automated Resume Evaluation System using NLP

$RohiniNimbekar^1, YogeshPatil^2, RahulPrabhu^3, ShainilaMulla^4$
Department of Computer Engineering
Don Bosco Institute of Technology, Mumbai, India.
shainila.dbit@dbclmumbai.org

***Abstract*— Recruiting candidates to fit a particular job profile is a task crucial to most of the companies. Due to increasing growth in online recruitment, traditional hiring methods are becoming inefficient. The conventional techniques usually include a labor-intensive process of manually searching through the applied candidates, reviewing their resumes, and then producing a shortlist of suitable candidates to be interviewed. In this era of technology, job searching has become smarter and more accessible at the same time. The companies receive enormous numbers of resumes/CVs, which are not always structured. There have been lots of work done for the job searching process. Whereas, the process of selecting a candidate based on their resume has not been entirely automated. This research proposes a model of extracting valuable information from the resume and ranking it according to the preference and requirement of the company. To achieve the desired goal, the entire process has been divided into three segments. The first segment consists of converting the unstructured resumes in structured data using NLP, and the second segment consists of the extraction phase, where the relevant information is extracted from the resume and giving them an identifier value. Finally, based on the values assigned,the resumes are ranked accordingly in the final segment.**

I. INTRODUCTION

While searching for jobs, the most important thing to represent an applicant is the resume. All the major industries today are driven by technology. The applicant can upload their resume. Resumes received are challenging to process and store in a unified database format. It becomes very tedious to select the most appropriate ones. Since resumes are unstructured documents based on the applicant's writing skills, they can be created in a multitude of formats. Dynamic extraction techniques are used to extract the most relevant information from the resumes. The filtering techniques match the resumes from the database to a single job posting. Resumes obtained by these filters are generally similar to each other as they fulfil the same criteria posted by the employer. This system fetches resumes which satisfy the requirement of a particular job post and rank them accordingly.

II. RELATED WORK

The hiring process has been changing gradually over the amount of your time. Graduates aim for getting placed in a good company whereas the recruiters search for people who would contribute effectively for structure growth. The primary impression of each candidate is their resume. Resume structure and its content are live for the leader to form a call for holding or eliminating the candidate for future rounds of any achievement method. Statistics indicate that the recruiter takes just a minute to digest a resume completely. There are numerous sorts of CV like combination resume written account resumes, targeted resume, visual resumes, practical resumes, etc[1]

In this competitive world, a candidate ought to possess a strong resume that conveys the specified information required by a company for any specific designation. The organization databases hold lakhs of resumes that are in free vogue and unstructured. The data and also the structure contents of resumes are going to be assortment below sub topics; the classification and also the illustration of data disagree from each other.[2] Thus, collection relevant information from every resume and storing it into the database in a structured format would reduce human effort. There are some difficulties in resume service by corporations since they consume abundant of your time, money, capacity, human effort. These companies need parsed resumes for achievement.

Automated recruitment systems need that Job seekers post their resumes on numerous websites like LinkedIn, Naukri.com, Monster.com, Resume builder, etc. Certain websites could retrieve unwanted resumes, whereas some may provide a terribly minimum range of resumes. This involves an Associate in Nursing approach for qualitative analysis of resumes. Since there are numerous alternative existing websites provide advanced facilities like searching on the idea of keywords, domain, location, etc. Their search doesn't take into consideration, the skill level of a selected candidate.[3] For example, if an organization searches for a willing candidate who can work in python language, they will look for candidates who have python language mentioned in their resumes. But how will they recognize the proficiency of that individual candidate in python language?

In our epitome, we will include other parameters like the projects and other skills like learning software in which the willing candidate can embrace in their resume. This information is going to be considered as input from the candidate, and by analyzing the text employed by him or her, we have a tend to will categorize the candidate into numerous expertise levels. So if a company need an associate employee who mainly incorporates a high experience level in python language, solely then his resume can be shortlisted. A knowledge domain of different

keywords can be designed, which can type the idea of categorization. This will conjointly facilitate the ranking, of multiple resumes to inform which one is better or worse than the opposite allowing the apply candidate to present themselves within the best possible method. The section identification and pdf parsing can also be improved. The standard reference documents can be augmented. The algorithmic formulation has the possibility of being used in an annotation and recommendation system.[4]

III. The existing recruitment process

Traditionally achievement and selection are two necessary methods of human resource management. The purpose of achievement is to spot appropriate manpower to fulfill the work necessities and job specifications. It is the foremost necessary perform of personnel administration on the opposite hand choice is bothered with securing the right info regarding someone. The article of choice method is to see whether or not the prospective candidate possesses the qualification for the particular job it's, an extended method. It Starts from the interview and ends with the contract of employment.[5]

A. Recruitment

HR plays essential role in achieving all the requirements of every department of a company. The strategic desires of the organization are specially designed. The activity that comes below the achievement method is to rent a worker for a corporation to confirm the continued operation of the organization.[6] This involves act with actual or potential job seekers, attract them to own a chance, and try to persuade them to figure for the organization in any means that of the fields. The target to attain the simplest quality result's to own the simplest quality within the best manner and numbers.

In the past, CVs/Resumes submitted by job seekers accustomed to be manually analyzed and judged by the employers [3]. This methodology remains followed within recent times. There is a range of various sources to recruit personnel reckoning on the sort of job vacancy. There are numerous companies that have totally different necessities for his or her organization. However, a massive-scale company receives many resumes every day, thus handling those numbers of resumes has become a crucial task and time overwhelming method. Due to these reasons, numerous companies provide specific format for job seeker. Jobseeker ought to replenish with needed info then the CV/Resume is going to be analyzed by machine, with easy pattern recognition and keywords looking. Whereas this methodology reduced the employment for the employers, it exaggerated the quantity of labor for the candidates considerably as they have to maintain different formats for every job they apply. To boot, it additionally tends to scale back the power and also the flexibility of writing the abilities at the side of the qualifications during a CV/Resume.[3]

B. Selection

Selection may be a later stage of achievement. It is the method of choosing people United Nations agency have relevant qualifications to fill jobs in a company. Choice is way quite simply selecting the simplest candidate. It's a trial to strike a balance between what the organization need and what candidate need to try. It involves selecting not solely new members of the organization however, additionally guaranteeing that the choice method will manage to attract competent and qualified candidates suited to work.
The focus in the selection process is on:
(1) The method of selection and skills in terms of contribution to the dependability of selections.
(2) The factor outlined and applied by decision-makers. And the way these reflect their comprehension of "necessary competence".
(3) However, the processes of selection embody the assumptions and commitments, the generalities, truths and confusions—of decision-makers about the imperatives of organization's culture and also the means they request to require care of and change this.

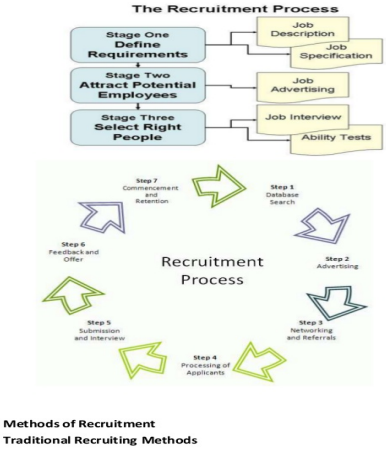

Fig. 1. Traditional recruitment process

IV. RESEARCH METHODOLOGY

Based on previous literature review form was developed that unstructured data contain virtually real-world data are difficult to classify since there's no predefined structure.

A. Conversion Phase

The first process that comes into existence the conversion of unstructured data to structured data. It has three main stages delineated in figure 2. Database consists of unstructured data, data in its negotiant state and successfully structured data. The user uploads the resumes. This can be automatically stored in a database.

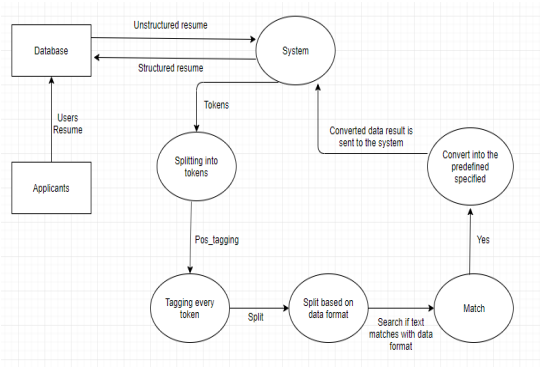

Fig. 2. Conversion of unstructured data to structured data

B. Extraction Phase

Resumes are dissimilar and have different structure. The resume consists of many information like education, skills, personal information, achievements, internships, etc. All this information in a resume varies considerably. Our proposed system will work on unstructured resume. It will extract unstructured resumes.[9]

Entity Name Extraction (NER). It helps in the transformation of unstructured text to structured text. Input resume will be in any format like web posts, PDFs, documents, spreadsheets, etc. NER is used for recognizing the entities into different parameters such as place, people, numerical expressions like phone number, time, etc. likewise temporal expression like durations, frequency, etc. It is a subtask of the information extraction method. Once a CV is structured, then extracting the relevant information becomes a straighforward task.[8]

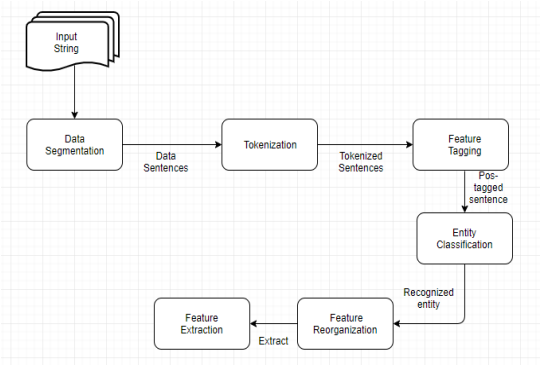

Fig. 3. Extraction Process

C. Filtration

The process of filtering resumes is predicated on comparing the job skills posted by the employer with the extracted data from the resumes. This process gives all the applicants who match the job description. To make the filtration process more efficient, a score is given to each resume to rank the applicant. Collaborative filtering is used to to predict the trend of selection [11].

D. Ranking

Ranking the resume can facilitate the organization to choose the candidate wisely. The ranking process will in shortlisting variants of CV. The increasing number of submitted CVs could overwhelm HR departments, which usually perform manual analysis of job applications. This process will improve the decision-making process, careful parameterization by the department's skilled recruiters which helps in better recruitment process in a shorter period.

Our system uses a machine learning algorithm to build the applicant learning model. It requires training data to learn the algorithm for constructing a ranking model, so it will predict the recruiter's judgment once given the candidate's resume. To combine predefined features for ranking, supervised learning algorithms are called "learning-to-rank" methods is used.[9]

In the testing phase, the candidate application is applied to learn model for sorting, and then finalized rank candidates list is generated.

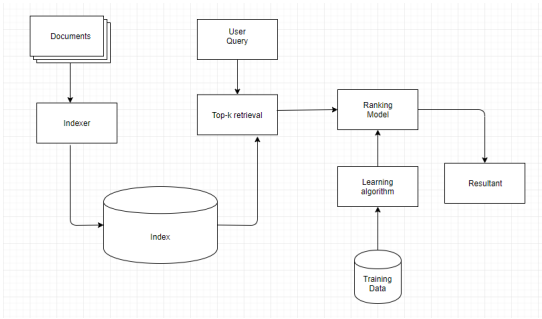

Fig. 4. Ranking Process

V. Architecture Overview

The proposed system consists of multiple modules. First the Section-Based Segmentation module which is used to extract the candidate's information like personal, education, experience, technical skills, internship(if any), hobbies, etc. The next module is the filtration module, which refines the lists by removing the insignificant terms that do not contribute to the matching process[10]. The third module takes a set of skills extracted from both resumes and job portals as input to classify them under their corresponding occupational categories. The category-based matching module takes the list of skills from both the resume and job posts to construct the semantic network. By deriving the relatedness between their concepts. Finally, the matching algorithm takes the semantic network as input and produces the measures of closeness between them as output.

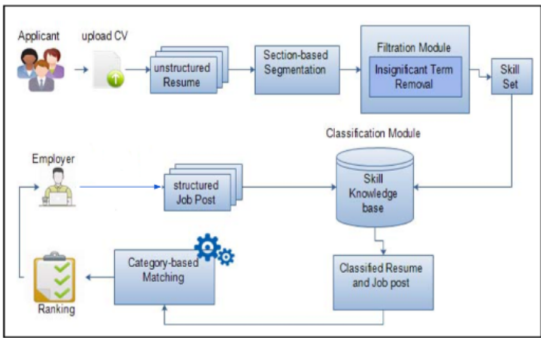

Fig. 5. Architecture Overview

VI. SURVEY REVIEW

Sr No.	Title	Work done by	Gap Identified
1.	JRC A Job Post and Resume Classification System for Online Recruitment	Abeer Zaroor, Mohammed Maree Muath Sabha	The resumes used in this research was structured
2.	Matching Jobs and Resumes: a Deep Collaborative Filtering Task	Thomas Schmitt, Philippe Caillou, and Michele Sebag	The resumes tokens were extracted in structured format
3.	Automatic Text Categorization by Unsupervised Learning	Youngjoong Ko, Jungyun Seo	The categories are predefined.
4.	Application of Machine Learning Algorithms to an online Recruitment System	Evanthia Faliagka, Kostas Ramantas, Athanasios Tsakalidis	Inconsistent CV formats,structured and contextual information.
5.	Cluster-based Ranking Index for Enhancing Recruitment Process using Text Mining and Machine Learning	Mayuri Verma	Effective ranking can improved through supervised machine learning applied in past dataset.

VII. CONCLUSION

The proposed model in this paper extracts the necessary data from a resume and segments them based on their values. However, the ranking and weights given for a segment of the resume may differ from company to company. As described, the whole process was segmented, and each segment was designed separately to perform its task. Finally, ARES gives ranks to resumes based on the necessary data, and the employers take the top few applicants into consideration.

VIII. FUTURE SCOPE

In this research, the domain was restricted to the resumes of only engineering students, and the amount of sample data versus the amount of test data was relatively small. The domain can be extended further to other domains like Telecommunication, Healthcare, E-commerce, and public sector jobs. In the future, we also plan to utilize the extracted information applicants' resumes to dynamically generate user profiles to be further used for recommending a job to job seekers. Also, this system can be extended by including aptitude tests. Therefore, future scope is vast.

REFERENCES

[1] Ryland, E. K., amp; Rosen, B. (1987). Personnel Professionals Reactions to Chronological and Functional Résumé Formats [Abstract]. The Career Development Quarterly, 35(3), 228- 238.
[2] R. Fagin, B. Kimelfeld, Y. Li, et al. Understanding queries in a search database system. In PODS, pages 273–284, 2010
[3] Malamitsa, K., Kokkotas, P., amp; Kasoctas, M. (december 2008). Graph/Chart Interpretation and Reading Comprehension as Critical Thinking Skills. Science Education International, 19(4)
[4] Di Wu, Lanlan Wu, Tieli Sun, Yingjie Jiang, "Ontology based information extraction technology", International Conference on Internet Technology and Applications (iTAP), Aug 2011.
[5] Essential of human resource management and industrial relations, subbarao Mumbai, Himalaya publishing house, 1996
[6] Abbie, R., Robert, V. (2008). Moving beyond traditional recruitment strategies. Recruiting and Hiring—Research Paper. Retrieved from http://academics.lmu.edu/media/lmuacademics/diversity/images/Moving Beyond Traditional Recruitment Strategies.pdf
[7] "Entity extraction: How does it work? - expert system," www.expertsystem.com/entity-extraction-work/.
[8] "Named entity extraction—lexalytics," https://www.lexalytics.com/technology/entity-extraction.
[9] Finn, A. and Kushmerick, N. "Multi-level boundary classification for infonnation extraction," in Proceedings of 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004.
[10] A Kmail, M Maree, M Belkhatir, and S Alhashmi "An Automatic Online Recruitment System based on Exploiting Multiple Semantic Resources and Concept-relatedness Measures," Proceedings of the EEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 620-627, 2015
[11] Chanawee Chanavaltada, Panpaporn Likitphanitkul, Manop Phankokkraud, "An Improvement of Recommender System to Find Appropriate Candidate for Recruitment with Collaborative Filtering", published in 2015 ICCSS