Grenze International Journal of Engineering and Technology, Jan Issue

Resume Parsing using Natural Language Processing

Dipti Suhas Chavare[1] and Mrs. Archana Bhaskar Patil[2]
[1-2]Department of Computer Engineering and IT, College Of Engineering, Pune (COEP)
Pune, India
Email: diptisc20.comp@coep.ac.in, abp.comp@coep.ac.in

***Abstract*—Screening resumes out of bulk is a challenging task, and recruiters or hiring managers waste a lot of their valuable time searching through each resume. Job seekers should have access to the best tools to find the perfect match for their profile without wasting time on irrelevant recommendations and manual searches. Often resumes are populated with irrelevant and unnecessary information. Therefore, parsing thousands of resumes manually consumes a lot of time and energy; thereby, it makes the hiring process expensive. In general, however, traditional job recommendation systems are based on simple keyword and semantic similarities that are usually not well suited to providing good job recommendations since they don't consider the interlinks between entities. In this paper, screening resumes is automated by using advanced Natural Language. Our model helps the recruiters to screen the resumes based on job descriptions within no time. It makes the hiring process easy and efficient by extracting the required entities automatically by using the Spacy NER model from the resumes and Joint NER and relation extraction will offer an entirely new approach to retrieving information using knowledge graphs, where you can explore different nodes to find hidden relationships.**

***Index Terms*— Machine Learning, Natural Language Processing, Named Entity Recognition (NER), Spacy3, Knowledge Graph.**

I. INTRODUCTION

In every discipline, technology has significantly evolved. Its importance can be seen in all businesses and organizations, whether large or small. Artificial Intelligence and its components enable us to create models with superior prediction and decision-making abilities, reducing the amount of manual labor required. Computers may act like people and make decisions based on how the model is educated with the help of Artificial Intelligence; all the above introduces Natural Language Processing, a branch of artificial intelligence that deals with human languages.

It analyses behaviour, patterns, and semantics to connect emotions (in human language) to computers to achieve personal goals. Recruiters used to have to sift through many resumes for a single job application in the traditional hiring approach, resulting in the rejection of some eligible candidates and, on rare occasions, the acceptance of some unworthy candidates [4]. Today, the industry's primary issue is how to find the proper personnel with low resources, over the internet, and in a short Amount of time. By automating the process, the system intends to give a solution to the challenges mentioned above.

While the field of natural language processing has been rapidly expanding in recent years due to the development of transfer-based models, its applications in the job search industry have been limited.

In a mad rush to get jobs, a fresher may apply for an irrelevant job; in such scenarios, one needs to spend more time analysing the resumes according to the company's requirements. So, to place a "Right Person in the Right

Grenze ID: 01.GIJET.9.1.587

Job," an intelligent resume ranking system is required to shortlist the candidates based on job descriptions.

II. Preliminaries

A. Natural Langugage Processing

Language acquisition is complicated by nature. A technology must grasp not just grammatical rules, meaning, and context, but also colloquialisms, slang, and acronyms used in a language to interpret human speech. Natural language processing techniques help computers understand language by replicating the human ability to do so.

B. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a new technique of utilizing the transformer design. BERT, for example, makes a prediction by analyzing both sides of a sentence with a randomly masked word. In addition to predicting the masked token, BERT predicts the sentence sequence by inserting a classification token [CLS] at the start of the first sentence and attempting to predict whether the second sentence follows the first by inserting a separation token [SEP] between the two sentences.

C. spaCy

spaCy is a Python library for Natural Language Processing (NLP) that comes with a variety of built-in features. It has become widely used in NLP for data processing and analysis. Unstructured textual data is generated on a massive scale, and it’s critical to filter and extract insights from it. To do so, you’ll need to represent the information in a computer-readable format.

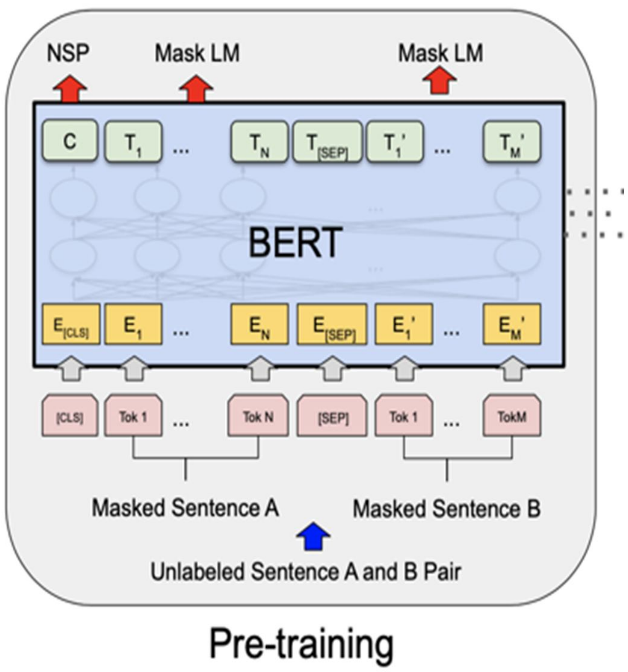

Fig.1 BERT Architecture (Pre-training)

III. Literature Review

The author proposed a model that helps the recruiters screen the resumes based on job descriptions by automatically extracting the required entities using the Spacy NER model from the resumes and then generating a graph displaying each resume's score. Based on the scores, the recruiter can choose the required candidates without rummaging through piles of resumes from unqualified candidates[1]. The author mentioned that the employment market is constantly changing and to fulfil the demand of today’s economy, occupations change, are introduced, or are removed. Due to globalization and the transition to working from home, the development speed has increased rapidly in recent years. In this paper, they have proposed a custom-build skills and occupation knowledge graph that fits the dynamic nature of the recruitment process and, by using the knowledge graph, explore various applications for skill-based matching of jobs to job seekers. Using link prediction can express the connection between skills and occupation. A method such as a node similarity and shortest path algorithm used for career path finding[2]. In this the Model is used to extract data from resumes and analyse the data to turn it into usable information for recruiters The first phase focuses on extracting tests from PDF resumes, which are then used to train the Model, and the data is then cleaned. After that, tokens are constructed for all of the dataset's skills, which are then vectorized. The Model's Machine Learning component is the second phase's focus. The first and most important step is to cluster the talents based on the vectorized dataset using the K-Means clustering algorithm. The level or rank of the resume is then determined by predicting the category of the new resume. The third phase focuses on comparing resumes to the company's needs and ultimately supplying recruiters with the best applicant[3].They designed the web application to screen resumes (Curriculum Vitae) for a specific job description is the topic of this article. Job seekers can upload their resumes and apply for any job posts they are still interested in using the interactive web application. The candidates' resumes are compared to the job profile criteria issued by the company. Using machine learning and natural language processing techniques, a firm recruiter can find a job (NLP). The resumes can then be given scores and rated in order of best match to worst match. Only the firm recruiter interested in selecting the best applicants from a vast pool of candidates will see these rankings. Recruiters will have less work because they won't have to manually go through every resume in the immense pool of candidates[4]. This Author designed a Web application in such a way that it makes the hiring process much more straightforward. This web tool assists us in screening and ranking resumes. Using Natural Language Processing technology, the submitted resumes are evaluated to the job description to find the best-profiled resumes (NLP). Finally, the resumes are graded and rated in the best match to worst match. The proposed method would allow for reading more than 30 resumes each minute, saving 85 percent of the time spent reading resumes. Recruiters will no longer have to review every resume from a big pool of prospects manually[5].Professor F. N. A. Al Omran and C. Treude suggested that researchers make an informed decision regarding which NLP (Natural Language Processing) library to use and that library changes may be required to attain good results[6]. Professor G. Prasad and K. K. Fousiya compared different techniques to Named Entity Recognition using English and Hindi corpora. This research aims to build on previous work and develop a more efficient approach for named entity recognition in different native languages[7]. They ranked resumes based on the accuracy of their matches using the KNN technique and cosine similarity. The KNN algorithm was evaluated using five metrics. Kappa Statistics, Accuracy, f1 Score, Recall, and Precision are the terms used to describe them[8]. Professor S. K. Kopparapu provided an unstructured text analytics approach for qualitative evaluation of CV/Resume documents and a system to organize and comprehend textual data.[9]. There are resume parsers that evaluate applicants based on their talents, but they are restricted to a few domains.[10]. Previously, most resumes were reviewed manually, which required time and energy. Their hiring procedure was far too drawn out. A potential applicant could lose out on the job if their application were made to fill in unnecessary information. Resumes had to follow a particular format. To forecast abilities, credentials, and experience in job descriptions for software, this paper develops a refined BERT model. The spaCy3 package was used to create a useful NER transformer model. Information retrieval will expand to a new area by discovering hidden relationships using NER and relation extraction.

IV. Proposed Methodology

Information extraction (IE) of named entity recognition (NER), which looks out and classifies certain entities in a body or bodies of texts, is used to automate the screening of resumes. Transformers have significantly changed the field of NLP, and they are used in information extraction.

Construct a career recommendation and skill discovery model that will take unstructured text as input and produce employment recommendations and skill suggestions based on elements like skills, years of experience, a diploma, and a major.

Use of spaCy 3 to extract entities and fine-tune the BERT Transformer:

NER is a semi of extracting information (IE) that searches out and categorizes specific items in a body or bodies of text and automates the resume review process. Recruiters can use our methods to assess applications related to job descriptions swiftly. It uses NER to automate the recruiting process by extracting the necessary entities. NER is used in various AI fields, including Natural Language Processing (NLP) and Machine Learning.

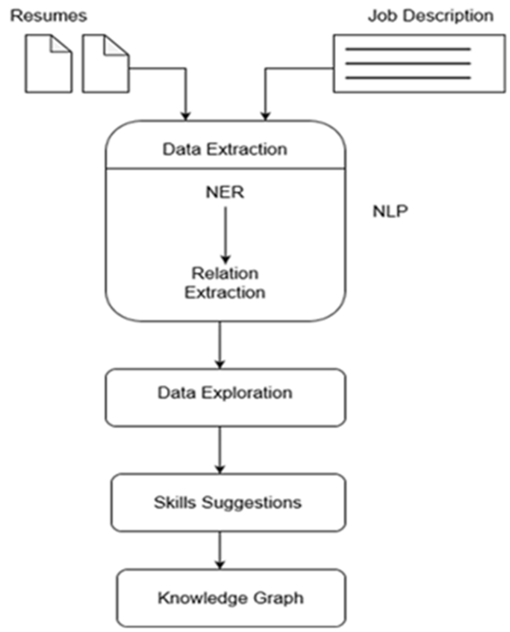

Fig.2 Job Analysis pipeline

SpaCy for NER:

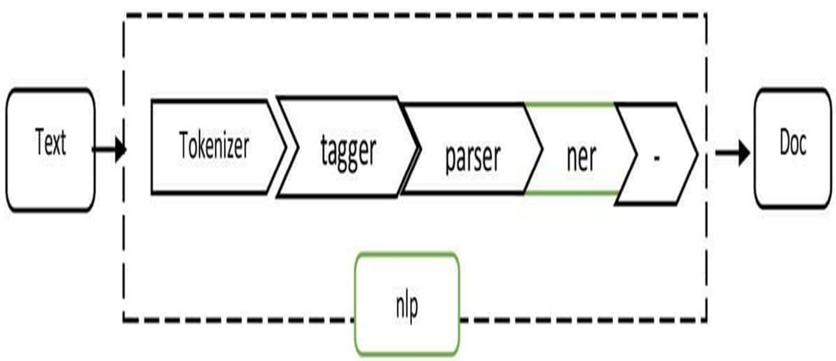

Fig.3 NER Pipeline

Preparing a customized NER model using SpaCy using BERT:
SpaCy comes with a default model that can recognize various names or numeric items, such as people, organizations, languages, and events.
First create manually labeled training data to train the model. As a result, use UBIAI, an online automation platform that parses texts automatically and allows us to write annotations for required entities.
A strong GPU featuring parallel processing is required for fine-tuning transformers. Utilize Google Colab for this because it gives free servers with GPUs.

Training the Model:
The NER model is trained using Python's spaCy package. Every "choice" made by spaCy's models — for example, which part-of-speech label to give or whether a word is a named item — is a prediction. The model's forecast is based on the instances it saw during training.

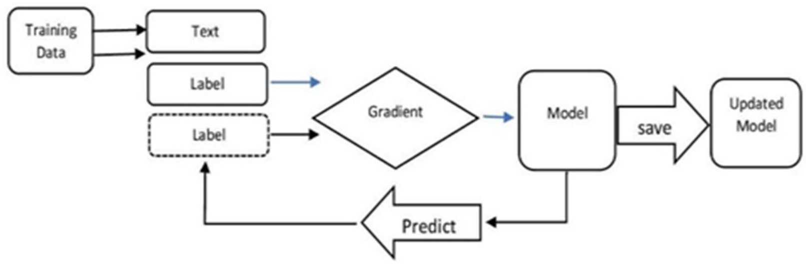

Fig.4 NER Model

Testing the Model:

The NER model is trained using Python's spaCy package. Every "choice" made by spaCy's models — for example, which part-of-speech label to give or whether a word is a named item is a prediction. The model's forecast is based on the instances it saw during training.

Using spacy3, train a Joint Entities and Relation Extraction Classifier with the BERT Transformer.

One of the most practical applications of NLP technology is the approach. Data from unstructured texts, contracts, economic papers, healthcare records, and other documents since it enables autonomous data queries to draw fresh insights. Named entity recognition has long been used to identify entities inside a text and save the data for better querying and filtering. NER, on the other hand, is insufficient to understand unstructured text meaningfully. Joint NER and relation extract will open a new way of collecting information from knowledge graphs, allowing you to navigate nodes to discover hidden connections. As a result, completing these chores together will be helpful

After fine-tuning a BERT model for NER with spaCy3, now add relationship extraction to the pipeline using spaCy's new Thinc library. Use an internet job description to train the relation extraction model and then put it to the test.

The connection extraction model is a classifier that predicts a relationship r between two entities e1 and e2. This classifier is applied to the output hidden states in the case of transformers. The link between the two entities Experience, Skills as Experience in and Diploma, Diploma major as Degree in will be extracted. The goal is to determine how many years of experience are required for a particular talent and the diploma major that correlates to the necessary diploma.

Using BERT Transformer to Create a Knowledge Graph for Job Search:

The extracted skills and years of experience can now be used to build a knowledge graph, with the source nodes being job description IDs, the target nodes being skills, and the strength of the relationship being the year of experience. Utilize the python tools pyvis and networkx to form our graph and use t years of experience as the weight to connect job descriptions to their extracted abilities. Following the discovery of linkages between the CV and job descriptions, the goal is to find essential talents that may not be listed on the resume but are critical to the field under consideration. Sort the job descriptions by the field for this purpose. Then, extract the relevant skills for each job found by querying all nearby jobs tied to resume skills. Query graph to find the highest job match to a target resume, find the (three) most popular skills and highest skills co-occurrence.

V. Conclusion and Future Work

While the NLP field has been growing at an exponential rate for the last two years thanks to the development of transfer-based models, their applications have been limited in scope for the job search field. LinkedIn, the leading company in job search and recruitment, is a good example. While I hold a PhD in Material Science and a master's in physics, I am receiving job recommendations such as Technical Program Manager at MongoDB and a Go Developer position at Total which are both web developing companies that are not relevant to my background. This feeling of irrelevance is shared by many users and is a cause of big frustration. The technique can be expanded to other fields such as healthcare, e-commerce, telecommunication and government jobs portals.

References

[1] Dr.K.Satheesh, A.Jahnavi , L.Iswarya, K.Ayesha, G.Bhanusekhar, K.Hanisha5, "Resume Ranking based on Job Description using SpaCy NER model" in International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 05 | May 2020.

[2] de Groot, Maurits, Jelle Schutte, and David Graus. "Job posting-enriched knowledge graph for skills-based matching." arXiv preprint arXiv:2109.02554 (2021).
[3] A. Sharma, S. Singhal and D. Ajudia, "Intelligent Recruitment System Using NLP," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-5, doi: 10.1109/AIMV53313.2021.9670958.
[4] S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika and A. Gurjar, "Web Application for Screening Resume," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 2019, pp. 1-7, doi: 10.1109/ICNTE44896.2019.8945869.
[5] M. Alamelu, D. S. Kumar, R. Sanjana, J. S. Sree, A. S. Devi and D. Kavitha, "Resume Validation and Filtration using Natural Language Processing," 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 2021, pp. 1-5, doi: 10.1109/IEMECON53809.2021.9689075.
[6] S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika and A. Gurjar, "Web Application for Screening Resume," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-7, doi: 10.1109/ICNTE44896.2019.8945869.
[7] S. K. Kopparapu, "Automatic extraction of usable information from unstructured resumes to aid search," 2010 IEEE International Conference on Progress in Informatics and Computing, Shanghai, 2010, pp. 99-103.
[8] Rajath V, Riza Tanaz Fareed , Sharadadevi Kaganurmath, “Resume Classification and Ranking using KNN and Cosine Similarity”, IJERTV10IS080057 VOLUME 10, ISSUE 08 (AUGUST 2021).
[9] D. Çelik et al., "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs," 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, Japan, 2013, pp. 333-338.
[10] Chandola,Divyanshu, et al. ‘Online resume Parsing System using text Analytics”Journal of Multi-Disciplinary engineering Technologies, 9.1(2015).