International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Resume Ranking based on Job Description using SpaCy NER model

Dr.K.Satheesh*(Professor), A.Jahnavi[1*], L.Iswarya[2], K.Ayesha[3], G.Bhanusekhar[4], K.Hanisha[5]

[1-5]Computer Science & Engineering, Dhanekula Institute of Engineering and Technology
Jawaharlal Nehru Technological University Kakinada, Ganguru, India

---

ABSTRACT

Screening resumes out of bulk is a challenging task and recruiters or hiring manager wastes lot of their valuable time by searching through each and every resume. Often resumes are populated with irrelevant and unnecessary information. Therefore, the process of parsing thousands of resumes manually consumes lot of time and energy there by it makes the hiring process expensive. In traditional hiring, resume screening is a manual process which consumes a lot of time and energy. In this paper the process of screening resumes is automated by using advanced Natural Language Processing which is a field in Machine Learning .Our model helps the recruiters in screening the resumes based on job description with in no time. It makes the hiring process easy and efficient by extracting the required entities automatically by using Spacy NER model from the resumes and then generates a graph displaying the score of each and every resume. Based on the scores recruiter can choose the required candidates without rummaging through piles of resumes from unqualified candidates.

Key Words: Machine Learning, Natural Language Processing, entities, Spacy, Named Entity Recognition(NER).

1. INTRODUCTION

Earlier, being under the pressure of higher authorities, Recruiters are forced to spend more time on analyzing and matching thousands of resumes with the Job descriptions in order to select required candidate .Being in a madrush of getting jobs, a fresher may apply for an irrelavent job in such scenarios one need to spend more time to analyse the resumes according to the requirements of the company .So in order to place a "Right person in a Right job". An intelligent resume ranking system is required to shortlist the candidates based on job description with in no time.

1.1 *Problem Definition:*

The problem is that the present system is not much flexible and efficient and time saving as it is not guaranteed that only eligible candidates will upload their resumes for a particular company. So out of bulk of relevant and irrelevant resumes the recruiter has to srcutinize them. Our system saves time for the recruiters to scrutinize the resumes based on job description by automatically ranking the resume .This not only helps the recruiters for srcutinization but also it helps the candidate will be get satisfied because he will get job in that company which really appreciates candidates skill and ability.

Learning to rank refers to machine learning techniques for training the model in a ranking task. Learning to rank is useful for many applications in Information Retrieval, Natural Language Processing, and Data Mining. Intensive studies have been conducted on the problem and significant progress has been made. This short paper gives an introduction to learning to rank, and it specifically explains the fundamental problems, existing approaches, and future work of learning to rank.

1.2 Objective:

The major objective of our system is to automate the hiring process in order to reduce the cost of hiring and to make the hiring process more efficient.

1) Candidates, who has been hired.

2) Client company, who is hiring the candidates.

Candidates, who has been hired:

Candidates who are searching for jobs after been graduated. Out of those, major number of candidates are so much desperate that they are ready to work on any post irrelevant to their skill set and ability. Where our algorithm will work in such a way that with the help of the previous result and previous ranking constraints, it will try to optimize the current result, which we called it Machine Learning. This will make sure that the relevant candidate is been hired for that particular vacancy. You can say best possible candidate.

Client Company, who is hiring the candidates:

Like I am the owner of a particular organization, obviously my aim would be to create such a team which is the best team in the world. It is like, if there is a vacancy of a java developer in my organization. So, I won't prefer to hire a python developer and then make him learn Java. That will be pretty useless and time consuming for both that candidate and for the organization too. Where our system help the organization to make out the best possible candidates list according to their given constraints and requirement for that particular vacancy. So there would be no regrets for both the entities, client company and that hired candidate. Hence satisfaction will be achieved.

2. LITERATURE REVIEW

1) Professor F. N. A. Al Omran and C. Treude proposed that the researches should make an informed decision about the particular NLP(Natural Language Processing)library they choose and that customizations to libraries might be necessary to achieve good results .

2) Professor D. Çelik et al proposed a system based on ontological structure model and called Ontology based Resume parser will be tested on number of Turkish and Engllish resumes.

3) Professor S. K. Kopparapu proposed a system to organize and understand the textual data and presents an unstructured text analytics approach for qualitative evaluation of CV/Resume documents.

4) Professor G. Prasad and K. K. Fousiya made a comparative study of different approaches for Named Entity Recognition applied to English and Hindi Language corpus. This study is conducted with a view to extend the work and developing an efficient method for named entity recognition in other native languages.

3 .RELATED WORK

3.1 First Generation Hiring Systems:

In this system the hiring team would publish their vacancies in newspaper, television and through mouth. The interested candidates would apply for the job by uploading their resumes. These bulk of resumes are sorted and screened by the hiring team then the shortlisted candidates were called for further rounds of interviews. This entire hiring process is time consuming and hectic to find right candidate for right job.

3.2 Second Generation Hiring Systems:

As the industries have grown, the hiring needs has rapidly grown. To serve this hiring needs certain consultancy units have came to existence. They offered a solution in which the candidate has to upload their information in a particular format and submit it to the agency. Then these agencies would scrutinize the candidates based on some constraints. These agencies were middle level organizations between the candidate and company. These systems were not flexible as the constraints are changed from system to system.

3.3 Third Generation Hiring Systems: This is our proposed system, which allow the candidates to upload their resumes in flexible format. These resumes are then analyzed by our system, indexed and stored in a specific format. This makes our search process easy. The analyzing system works on the algorithm that uses Natural Language Processing, sub domain of Artificial Intelligence. It reads the resumes and understands the natural language/format created by the candidate and transforms it into a specific format. This acquired knowledge is stored in the knowledge base and is compared against the job description in order to shortlist the candidates based on company requirements

3.4 Drawbacks:

1. Earlier bulk of resumes are screened manually which consumes lot of time and energy.

2. Expensive hiring process.

3. Potential candidate may loose the opportunity because of populating irrelevant data

4. Resumes needed to be in specific format.

3.5 How to overcome:

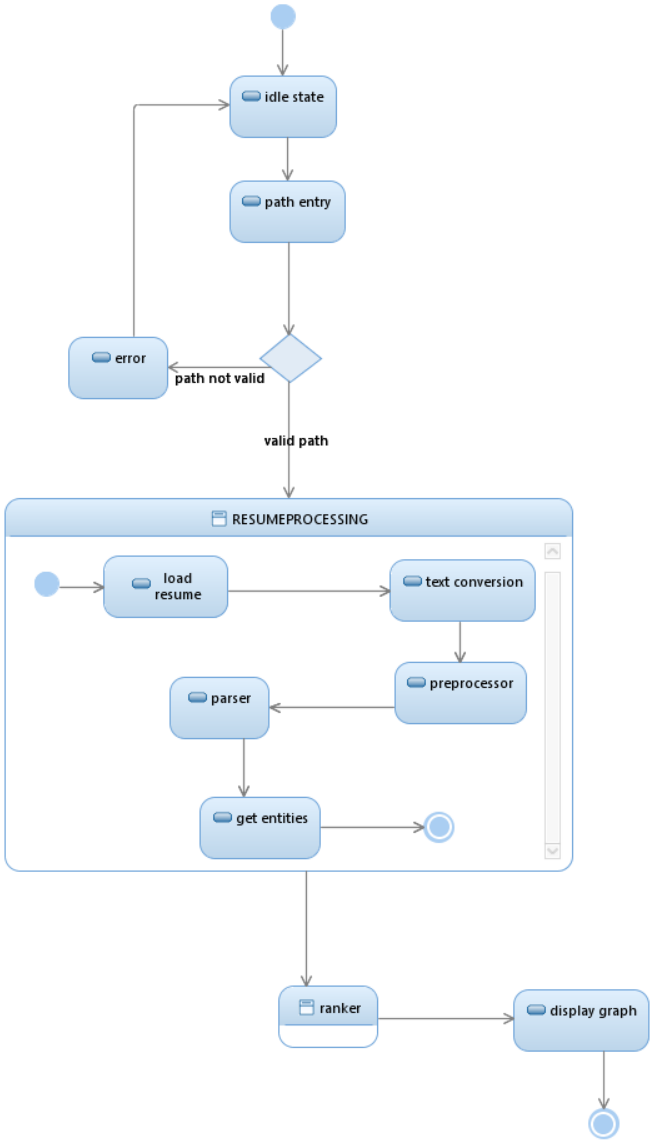
1. Use of NLP allows the candidate to upload the resume of any format because everyone will have their own style of writing.

2. Use of Machine Learning to rank candidates is efficient as it reduces the time and effort of parsing through each and every resume manually.

3. Use of NLP can be used to avoid unnecessary data which is populated in most of the irrelevant resumes.

4. METHODOLOGY

The process of screening resumes is automated by using Named entity recognition (NER) is a sub-task of information extraction (IE) that seeks out and categorizes specified entities in a body or bodies of texts .Our model helps the recruiters in screening the resumes based on job description with in no time . It makes the hiring process easy and efficient by extracting the required entities automatically by using NER which is used in many fields in Artificial Intelligence (AI) including Natural Language Processing (NLP) and Machine Learning.



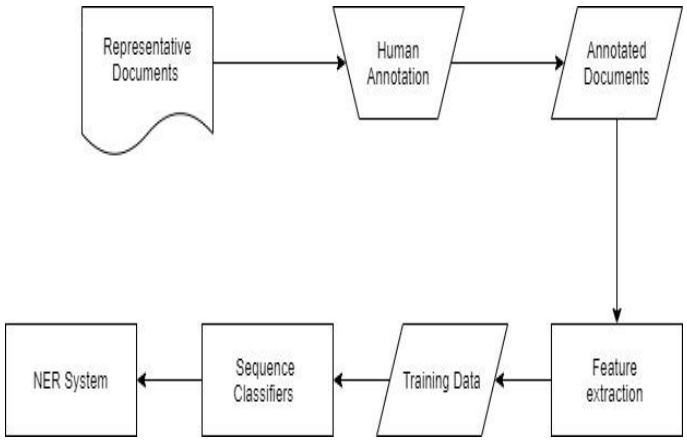
SpaCy for NER

SpaCy is an open-source library for advanced Natural Language Processing in Python.

Some of the features provided by SpaCy are Tokenization, Parts of speech tagging, Text classification, Named Entity Recognition etc.

Preparing a customized NER model using SpaCy:

SpaCy provides a default model which can recognize a wide range of named or numerical entities, which include person, organization, language, event etc. Apart from these default entities.



Resume summarization using NER:

Data Preparation:

Our first task is to create manually annotated training data to train the model. So we are using an online automation tool called Dataturks which automatically parses the documents and allows us to create annotations of required entities.

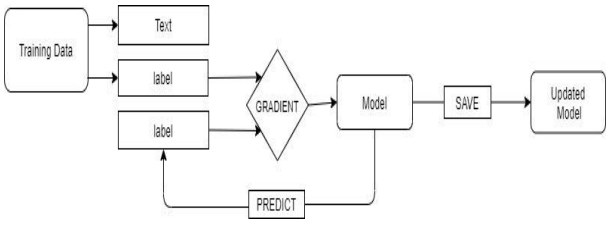
Uploading data: You can upload zip file of all documents at a time or each file individually. Each document becomes an independent data item to be tagged

Downloading results: The results are downloaded in the form of text file where each line is in JSON format containing label for each entity and start and endpoints of associated text as shown below

{"label":["Companies worked at"],"points":[{"start":2034,"end":2048,"text":"Microsoft India"}]}

Training the Model:

We use python's spaCy module for training the NER model.spaCy's models are statistical and every "decision" they make — for example, which part-of-speech tag to assign, or whether a word is a named entity — is a prediction. This prediction is based on the examples the model has seen during training.



Testing the model:

The model is tested on 20 resumes and the predicted summarized resumes are stored as separate .txt files for each resume. For each resume on which the model is tested, we calculate the accuracy score, precision, recall

and f-score for each entity that the model recognizes. The entity wise evaluation results can be observed. If it is observed high then the results obtained have been predicted with a commendable accuracy.

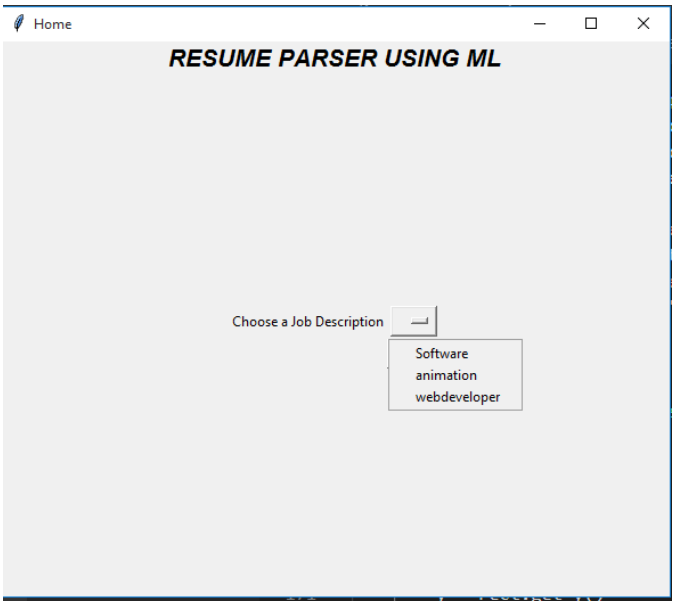
```
Deserializing model: /home/ljzhao/dilated-cnn-ner-master/models/dilated-cnn.tf
Segment evaluation (train):
              F1      Prec    Recall
Micro (Seg)   99.37   99.56   99.18
Macro (Seg)   99.33   99.54   99.12
.......
       LOC    99.56   99.68   99.44
      MISC    99.08   99.41   98.75
       PER    99.49   99.58   99.41
       ORG    99.18   99.47   98.89
Processed 203621 tokens with 23499 phrases; found: 23411 phrases; correct: 23307.

Segment evaluation (test):
              F1      Prec    Recall
Micro (Seg)   89.23   90.09   88.39
Macro (Seg)   88.40   89.46   87.36
.......
       LOC    91.35   90.14   92.60
      MISC    84.84   86.02   83.86
       PER    92.26   93.98   90.61
       ORG    85.05   86.80   83.37
Processed 51362 tokens with 5042 phrases; found: 5030 phrases; correct: 5252.
Testing time: 23 seconds
[ljzhao@localhost dilated-cnn-ner-master]$
```

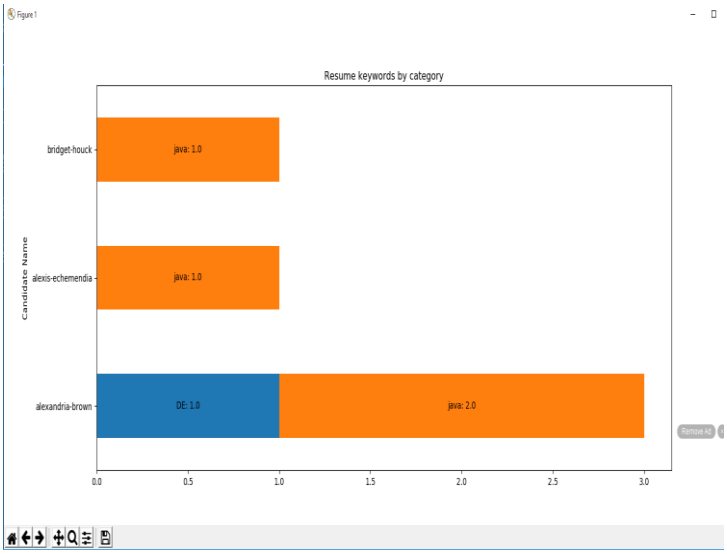
Output:

Comparison of obtained knowledge against Job Description:

Now we compare the acquired skill set against the job description and display the graph based on match score using matplotlib as follows.



Matplotlib is used to plot a graph as follows



5. CONCLUSION

Our application helps the recruiters to screen the resumes more efficiently there by reducing the cost of hiring. This will provide potential candidate to the organization and the candidate will be successfully be placed in an organization which appreciate his/her skill set and ability.

FUTURE SCOPE

The application can be extended further to other domains like Telecom, Healthcare, E-commerce and public sector jobs.

REFERENCES

1) F. N. A. Al Omran and C. Treude, "Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments," 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR), Buenos Aires, 2017, pp. 187-197.
2) "NER' 19," in IEEE Pulse, vol. 9, no. 5, pp. C4-C4, Sept.-Oct. 2018.
3) 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.
4) 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.
5) G. Prasad and K. K. Fousiya, "Named entity recognition approaches: A study applied to English and Hindi language," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-4.