

Natural Language Processing for the Job Portal Enhancement

^{*1}Mukesh Soni,
Department of Computer Engineering
Smt. S. R. Patel. Engineering College
Unjha, India
soni.mukesh15@gmail.com

²S. Gomathi
IEEE Member Madras Society
Coimbatore, India
gomathisrinivasan@gmail.com

³Yagna Bhupendra kumar Adhyaru
Department of Computer Engineering,
HCET,
Patan, India
yagna.1988@gmail.com

Abstract— The manual helps work applicants to tailor the career hunt to their needs. It increases their awareness of the mission through the appropriate amount. The key advantages are somewhat accessible and economical in utilizing electronic job platforms. Candidates can upload their CVs (Curriculum Vitae) to all areas free of charge and search jobs. However, certain services are necessary to improve the work profile, such as resume building and visibility services. It intended to provide end users/clients with a platform to obtain tremendous opportunities such as Job to Talent, Talent to Job, and Similar Talent.

Keywords— NLP, job, portal, candidates, CV, Natural Language Tool Kit (NLTK), pre-processing

I. INTRODUCTION

I.

This study begins with the brief description on job portal enhancement followed by using natural language processing and by the prominent features of the tool. This provides the guidelines to job seekers and system features to track and manage exact search results. It improves the existing job search algorithm for talent by newest, best match, location, and job title. It has new and existing job skills and enhances parsing CV, get data in a meaningful way and to extract data such as skills, experience, education, social information and certification if any through machine learning techniques Natural Language Processing (NLP) [1]. While new talent (skills/ keywords) or any talent profile updated every time. The process will set to scan the entire talent information and get all keywords related to job or matching algorithm [3]. The process is to finding the related keyword's generator-use to generate related job titles, skills and CV parsing for extracting data from resume.

As a job portal there are different types of portal available like alumni web portals [8], e-commerce web portals, sale support portals [9]. In all the portals there are different information are available and that information should be process in very efficient and secure manner. Sales or e-commerce portals need to provide security like authentication [10], integrity, confidentiality and must implemented cryptographic algorithms for maintaining information privacy, security among the users.

II. NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing is the development of human language applications and services [1]. With the NLP, developers can organize and structure their expertise to carry out tasks such as automated summary, translation, recognition of entities, relationship removal, feelings analysis [5], speech recognition, and the segmentation of the topic. NLP used for text analyses to allow computers to understand how people communicate. This interaction of human-computers enables real-world applications like automatic resuming text, sentiment analysis, extraction of subject matter, identified entities, speaking components, extraction of relationships, stemming, etc. NLP (Natural Language Processing) also used for the mining of documents, computer translations, and automatic answering queries.

A. NATURAL LANGUAGE TOOLKIT

NLTK is one of the leading forums for the development of human language data programs for Python [2]. It offers simple to use interfaces for over 50 companies and a collection of text libraries to define, tokenize, start, identify, decode, and procedural reasoning, NLP digital database wrappers, and an engaging platform for debate [2].

III. PORTAL DAY ONLINE STRAY

Where fishing, searching for work or seeking employment are activities of career quest because of homelessness, sub-employment, dissatisfaction, or the desire for a better career [4]. This is a forum which joins employers and job seekers in fulfilling their aims and demands. Recruiters are searching for the best applicant to perform the workload effectively with the required training. On the other side, work-seekers want a position with which expertise and experience will use to grow their future. The easiest way to find the right applicant is to submit work applications to job portals. Nonetheless, it is indispensable to pick the correct app.

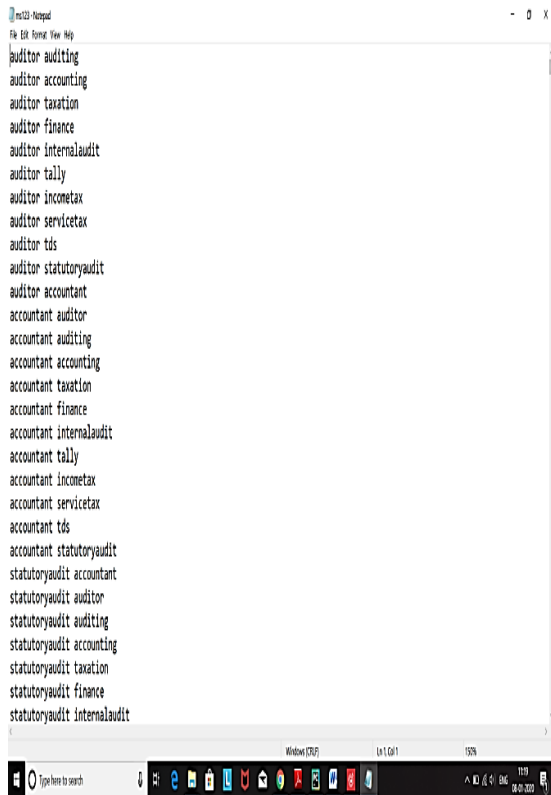


Fig. 3. Keyword Parsing

VI. DATA PROCESSING

Data pre-processing [7] applies to transformations that performed before they loaded into the algorithm. Pre-processing of data is a technique used to turn the raw data into a clean collection of data.

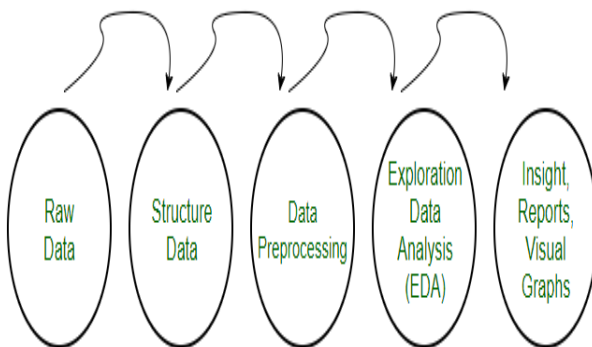


Fig. 4. Processing Resume data

A. Extracting Text From Doc And Docx:

Resumes have no set file structure and can be stored in any file format like .pdf or .doc or .docx. The most significant difficulty is to interpret and translate the curriculum vitae into plain text. Two Python modules are required to do this: pdfminer and doc2text. This software will also remove txt from .pdf and .doc, and the formats for .docx files.

B. Extracting Name

The best solution is to use regular expressions when extracting names from curricula. Nonetheless, it utilizes a more sophisticated method named SpaCy. SpaCy is a natural language processing platform with industrial strength used for document and language analysis. SpaCy has a pre-trained interface for recognition, sorting, and description of individuals. The key slogan is to remove names from the Agency Identification.

C. Extracting Phone Numbers and mail id's

It uses regular expressions to extract telephone numbers and e-mails. Pandas an open-source library to manipulate data in Python. The library of Panda built over Numpy, so Pandas has to operate Numpy. Pandas offer an easy way of creating, manipulating, and displaying data. A panda is also a stylish time series solution. It will use the panda module to read the CSV file once the file has read.

```

"C:\Users\Lenovo\PycharmProjects\NLP PROJECT\venv\Scripts\python
Shalini Gopal 12,GANDHI NAGAR 2nd CROSS,KARUR 639001 shalinigopa
Shalini Gopal 12,GANDHI NAGAR 2nd CROSS,KARUR 639001 shalinigopa
Shalini Gopal
9698934721
shalinigopal98@gmail.com
a good knowledge knowledge dobj Have
testing testing pobj in
[]
[('MSC(IT)', '2015')]
  
```

Fig. 5. Data Extraction

VII. DATA PREPROCESSING

A keyword-based CV parser uses the words, phrases, and simple patterns to identify and find understandable heuristic algorithms around these words in the text of the CV / Resume. Keyword extraction consists of extracting in a document the most relevant words and phrases. It contributes to summarize the contents of a text and to recognize the main issues under discussion.

REFERENCES

NLTK is the natural language toolkit literally. The main slogan of this article will learn to tokenize data (words and phrases). All stop-words must discard. It will load a whole list of stop words with the Natural Language Toolkit (NLTK) module and then delete them from the resume file.

Fast Text is a library created by the Face book AI Research Laboratory to learn word embedding and text classification. This model establishes a learning algorithm for vector representations for terms that is unattained or supervised.

```
from gensim.models.fasttext import FastText
mod = FastText.load_fasttext_format('/content/fil9.bin')
mod.most_similar('design', topn=15)

/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:
This is separate from the ipykernel package so we can avoid
/usr/local/lib/python3.6/dist-packages/gensim/matutils.py:737
if np.issubdtype(vec.dtype, np.int):
[('indesign', 0.9930545687675476),
 ('designer', 0.9890392422676086),
 ('bodydesign', 0.9823551177978516),
 ('3ddesign', 0.9799078106880188),
 ('logodesign', 0.9788124561309814),
 ('shipdesign', 0.9785060286521912),
 ('mouldesign', 0.9773669242858887),
 ('labeldesign', 0.9748393297195435),
 ('storedesign', 0.9746311902999878),
 ('visualdesign', 0.9708488583564758),
 ('printdesigner', 0.9706470370292664),
 ('fashiondesigner', 0.9701331257820129),
 ('plumbingdesigner', 0.967864990234375),
 ('emojisdesigner', 0.9675911664962769),
 ('interiordesigner', 0.9649461507797241)]
```

Fig. 6. Pre-processed the skills

VIII. CONCLUSION

The Career creation platform for the Natural language processing project is to monitor the work for the sectors that will help identify the best job seeker and the best job applicant. Below is a comprehensive description of the work of NLTK. The online recruitment project's main goal is to keep records of jobs. The recruiter of the organization must submit the vacancies with full area particulars, like technology, skills [6], date of interview, and other relevant details.

When a recruiter applied jobs to the database. A customer (work-seeker) can view the work as a front-end user can apply for the job and mail a recruiter or company. The company will now also pick the correct candidate according to the tests, sending the employee an interview mail. An admin is part of the job forum for recruiters and users

- [1] Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization Peter Jackson, Isabelle Moulinier, John Benjamins Publishing, 2007
- [2] NLTK: The Natural Language Toolkit, Edward Loper, Steven Bird, Submitted on 17 May 2002.
- [3] S. Bsiri, M. Geierhos, and C. Ringlstetter, "Structuring job search via local grammars," Advances in Natural Language Processing and Applications, pp. 201, 2008.
- [4] Ildis Bojars John G Breslin "ResumeRDF: Expressing Skill Information on the Semantic Web".
- [5] Gomathi, S., Ragavi, R. P., & Monika, S. (2020). Implementation of Statistical Data Analytics in Data Science Life Cycle. In Intelligent Computing and Innovation on Data Science (pp. 305-313). Springer, Singapore.
- [6] Gomathi, S., and V. Narayani. "Systemic Lupus Erythematosus Prediction Tool Using Optimal Cluster Based Classification (OCBC) Algorithm", International Journal of Engineering & Technology 7.4 (2018): 2806-2809.
- [7] Gomathi. S, Narayani. V, "Preprocessing Systemic Lupus Erythematosus (SLE) data set with Pentaho Data Integration (PDI)", International Journal of Recent Innovation in Engineering & Research, - ISSN: 2456 – 2084, Volume 2 issue 3, March 2017.
- [8] Patel M., Rami D., Soni M. (2020) Next Generation Web for Alumni Web Portal. In: Balaji S., Rocha Á., Chung YN. (eds) Intelligent Communication Technologies and Virtual Mobile Networks. ICICV 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 33. Springer, Cham.
- [9] Janki Patel, Priyal Patel, Bansari Patel, Nainil Patel, Mukesh Soni, "Sales Support for Interactive Marketing", International Journal of Trend in Research and Development. Volume 4, Issue 3(2017).
- [10] Soni M., Patel T., Jain A. (2020) Security Analysis on Remote User Authentication Methods. In: Pandian A., Senjyu T., Islam S., Wang H. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2018). ICCBI 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 31. Springer, Cham