

Resume Parser

Aneesha T Ibrahim¹, Annette J K², Geethika S³, Archana Naik⁴

^{1,2,3,4}*Department of Computer Science & Engineering, Nitte Meenakshi Institute Of Technology,
Yelahanka, Bengaluru*

Abstract—Nowadays, employments through online entries has turned out to be much advantageous for the job seekers and also the recruiter. These entries help the recruiter to locate the ideal applicant and increase the chances of the job seeker's possibility to land their ideal position based on their skills. Computerized and easy to use programming techniques provides us with a solution to parse each and every resume to select the best candidate based on the requirements of the recruiter. When an organization has a vacant position, it gets a large number of resumes for a solitary position. The concerned experts need to scrutinize every parameter of the resume and then select the candidates for the recruitment process which is very time consuming. This paper provides the faster selection process, by listing the candidates suitable for a particular job position easier based on the specifications provided by the recruiter. The parsing application is produced for the resumes in .doc, .docx and .pdf format. The resumes are converted into a text format for extracting the details. Segmentation is used to select the best resume for a particular position is suggested.

Keywords—Dictionary, Resume Parser, Resume, Segmentation, Recruitment

I. INTRODUCTION

A resume contains one or more pages. It is a formal document submitted to job recruiters as means to see a list of an applicant's work experience, education and skills. The word *résumé* is of French origin and means a "summary" of relevant job experience and education. The *résumé* is usually one of the first items, along with a cover letter and sometimes an application for employment, which a potential employer sees about the job seeker. It is typically used to screen applicants from a large number of applied candidates, which is often followed by an interview.

Recruitment plays a major role in any organization. It is the process of finding and hiring the best qualified candidate for published job opening, in a reasonable and effective manner. Recruitment process includes certain steps. They are: analyzing the job requirements, attracting potential employees, screening and selecting applicants, hiring, and integrating the new employee to the organization. At the same time, it is a time consuming and laborious job for the recruiters.

When a company has a vacant position, it receives a large number of resumes for a position. The concerned Human Resource (HR)/the hiring team have to go through all the relevant details of the resume and then select the candidates, for the interview.

This paper discusses developing a parsing application, for resumes received in multiple formats which include .docx, .doc, and .pdf. This application reduces the time and manual effort of searching through the multiple resumes for choosing the suitable resumes.

The technique involved is known as resume parsing. Other names include, resume extraction, CV parsing, CV extraction, which allows the automated storage and analysis of resume information. Multiple resumes are uploaded into parsing software and the information is extracted so that it can be sorted and searched. Resume parsers first analyzes a resume, and then extracts the desired information. After the resume has been analyzed, a recruiter can specify the job skills required and get a list of relevant resumes as the output. Some parsers provide semantic search, which adds context to the search terms and tries to understand the intent in order to make the results more reliable and comprehensive.

II. RELATED WORK

The resumes are in the semi structured format in which the candidates specify their details in their own format. Selecting the suitable candidate for the selection process requires the HR team to manually read through all the resumes and select the suitable ones. Using the techniques of text extraction, the selection of the suitable resumes is made simple. The experiments over 96 Chinese resume in different forms are parsed to derive accurate results. They reported statistics on 2784 available resume fields, which showed that the overall recall rate is 93.75%, and the accuracy rate is 91.71% [1].

In [2], a matching algorithm for fitting a new skill into its corresponding class in the skill ontology is proposed. So any new skill that is inserted as a new concept becomes a child class of an existent class in the skill. The ambiguity problem caused by the polysemy of the phrases, in expressing the skills in the text of a resume is overcome by the solution provided. Since the system performs very well on a large corpus of resumes, the developed system has proved to be very efficient. A total of 13,337 skills are defined in the skill ontology, which are efficiently recognized or found in the skill ontology.

Information extraction based on regular expression and text automatic classification used in this paper provides high accuracy in basic information extraction [3].

The sentences from the resumes are classified based on the sentiment of the text as “positive”, “negative” and “neutrality” for evaluating the sentiment of the text in the resume [4].

The objective of the study conducted in [5] is the identification of how e-recruiting affects the overall process of recruitment and whether it causes changes in the nature and sequence of tasks associated with the traditional recruitment of candidates. The research conducted is relevant for both academia and practitioners.

In [6], an attempt is made to develop a system which will eliminate the difficulties arising in existing system. When the resume is uploaded, it will be parsed and then transforms into a standardized format. The candidate can upload his resume in multiple formats like .doc, .pdf etc. A user interface screen will be given where the authority concerned will store the job profile for the particular post. The system will go through all the resumes and suggest the eligible candidates who are nearly falling under those job profiles and display it to the concerned authority.

The rest of the paper is organized as follows. Architecture of the system is explained in section III. Implementation of the system and experimental results are presented in section IV. Concluding remarks and future scope are given in section V.

III. ARCHITECTURE OF THE SYSTEM

The various dictionary frequencies are first loaded into the memory upon starting of the software. Resumes need to be uploaded. Selection of programming languages, software tools and packages, hobbies need to be done as per the requirements of the job role. The resumes are parsed and compared with the skills provided through the selection. The best candidate resume is suggested after applying the algorithm. Separate bar graphs for programming languages, tools and packages containing the frequency of occurrences of the skillset are specified. This will be displayed through the interface in each of the resumes.

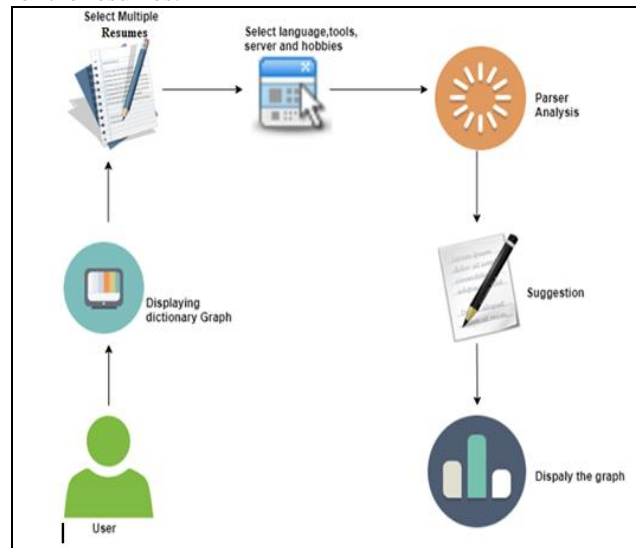


Figure 1: System Architecture

In the system architecture as shown in Figure 1, user represents the HR person. The dictionary graph provides the frequency of words present in the dictionary files. For the skill set selected by the HR, the system will take all the resumes and performs the parsing. Once the parsing is completed the graph will be displayed only for the resumes which are matching with the requirement as selected by the HR.

Implementation

To look for the suitable resume for a given requirement the HR person has to search the available stack of the resumes. There will be new resumes which have to be added into the stack, such that the parser can search in all. The parsing begins with adding new resumes into the system, providing the required skill set for the selection. Using these details the parser will provide the suggestion and display the graph.

The process is been given considering following scenario.

Resumes selected: JyothiPHPprofile_Copy1.docx, Kalyan Chakravarthy.docx.

Skills selected: Programming languages: Java, PHP. Packages: spring. Tools: WAMP, MySQL.

IV. HOBBIES: BADMINTON.

4.1. Uploading resumes

The new resumes are added into the system. After uploading multiple resumes all the data present in these resumes is converted to text format.

External packages are used for the purpose of conversion. For conversion of .doc files to text, the org.apache.poi.hwpf.HWPFDocument and org.apache.poi.hwpf.extractor.WordExtractor packages are used.

For conversion of .docx files to text, org.apache.poi.xwpf.usermodel.XWPFDocument and org.apache.poi.xwpf.usermodel.XWPFParagraph packages are used. For conversion of .pdf files to text, org.pdfbox is used. The bar graph of word vs frequency of the different keywords that are loaded in dictionaries is displayed.

Figure 2 shows how the resumes are selected from a folder that contains the resumes in .doc, .docx and .pdf format.

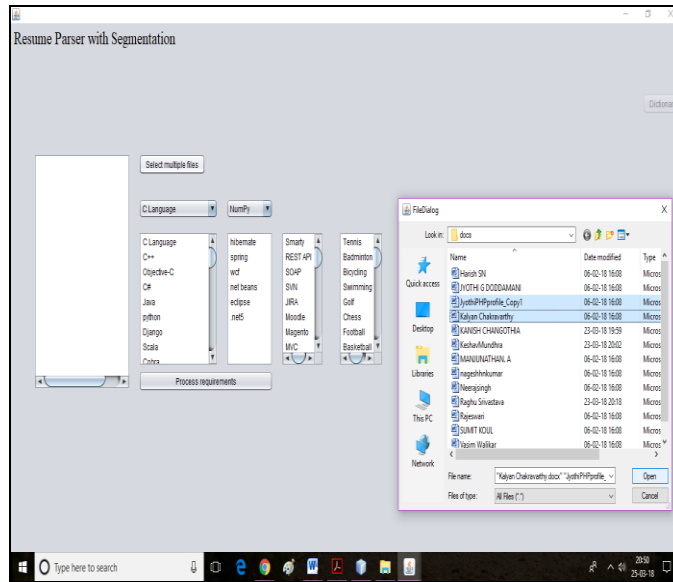


Figure 2: Selection of resumes

Two or more resumes can be selected at a time for parsing the resumes into a common text format as shown in Figure 3. The resumes which are in semi structured format, saved with any of the format are converted into plain text format.

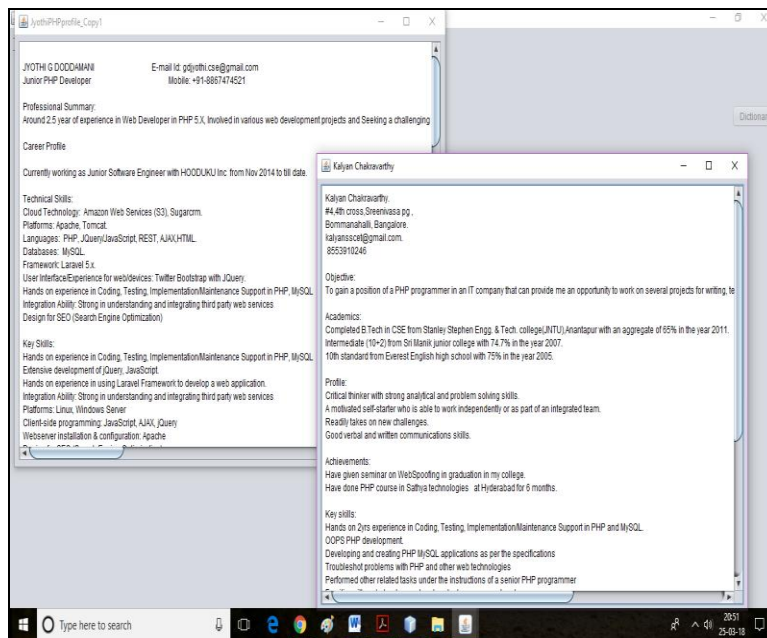


Figure 3: Conversion to text format

4.2. Selection of skills

The HR person has to enter the skill set required for the desired position in the organization. Figure 4 shows the window with option to enter skills, tools, packages and programming languages to select the best resumes.

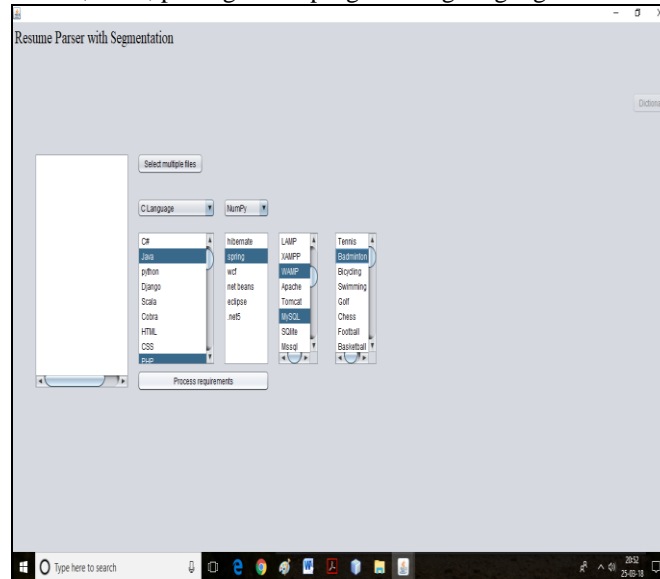


Figure 4: Selection of skills required for job role.

HR has given a provision to choose from a wide variety of resumes.

4.3. Filtration

After successful selection of all the skills, tools, programming languages and hobbies the parsing code will parse the uploaded resumes i.e. it will parse all the data present in resumes. It will compare the skills selected by the recruiter with the each of the resumes selected. A candidate resume is selected based on frequencies of the skill set given by the recruiter. The resume that has the overall highest frequency count of the skills will be chosen as the best.

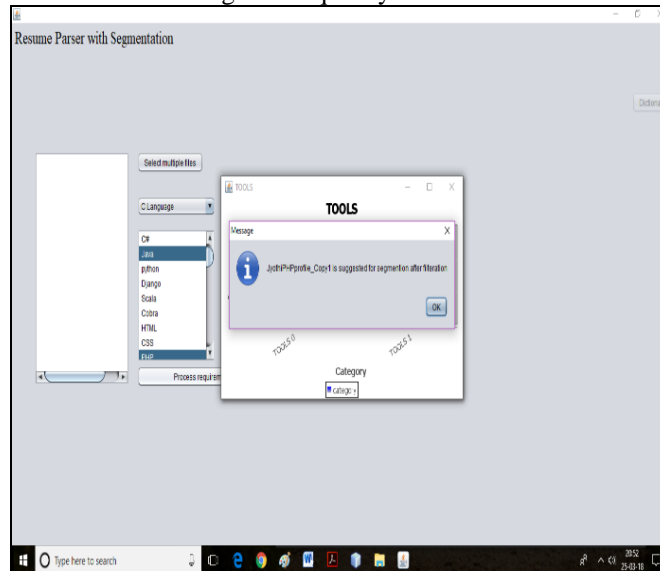


Figure 5: Suggestion after filtration

4.4. Segmentation

Segmentation is done based on the experience of each of the candidate's experience. The selection skills stage will select all the resumes with matching skill set. To filter the number of resumes, the candidate's experience is used as segmentation criteria.

The final best suitable resume is presented as the suggestion for the recruitment by the system. Figure 6 depicts the final result of the system.

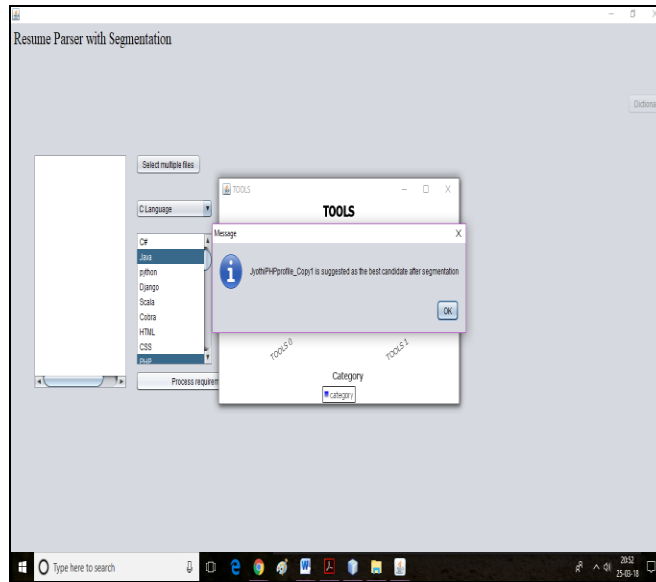


Figure 6: Suggestion after segmentation.

The result of filtration and segmentation are also depicted with graphical representation as shown in Figure 7. This Graphical representation provides the total word count tools, languages and packages selected by the recruiter present in each of the resumes.

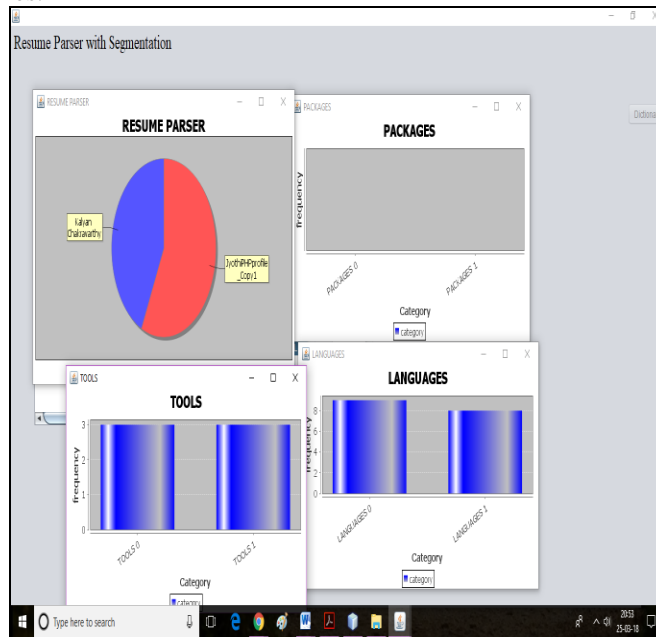


Figure 7: Graphical representation of result

V. CONCLUSION AND FUTURE SCOPE

In this project we have developed efficient resume parser software that provides high accuracy. The accuracy is dependent on the number of resumes processed. The algorithm used is efficient in calculating the frequency of occurrences of the skills required for the job role in the resumes. The speed of the system is quick. The graphical representations provide in depth information about the frequencies calculated through segmentation and analysis. Currently there is a limitation in the number of resumes that can be processed at a time. Further work involves overcoming that barrier and making it a highly efficient parser. Also the system needs to be made more large-scale.

VI. REFERENCES

- [1] YAN Wentan, QIAO Yupeng, "Chinese resume information extraction based on semi-structured text", Key Laboratory of Autonomous System and Network Control, College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China. Proceedings of the 36th Chinese Control Conference July 26-28, 2017, Dalian, China
- [2] Emil St. Chifu, VioricaRozinaChifu, Iulia Popa, IoanSalomie, "A System for Detecting Professional Skills from Resumes Written in Natural Language", Department of Computer Science, Technical University of Cluj-Napoca, Romania
- [3] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao bo, Lin Zhi-qing, "Resume Parser: Semi-structured Chinese document analysis", 2009 World Congress on Computer Science and Information Engineering.
- [4] Bin Wen, Wenhua Dai, Junzhe Zhao, "Sentence Sentimental Classification Based on Semantic Comprehension", Department of Computer Science, Hubei University of Science & Engineering, 2012 Fifth International Symposium on Computational Intelligence and Design.
- [5] Anna B. Holm, Aarhus University, Denmark, "The Effect of E-recruitment On the Recruitment Process: Evidence from Case Studies of Three Danish MNCs", Proceedings of the Third European Academic Workshop on electronic Human Resource Management, Bamberg, Germany, May 20-21, 2010, CEUR-WS.org, ISSN 1613-0073, Vol. 570, online: CEUR-WS.org/Vol-570/ , pp. 91-111.
- [6] Pooja Shivratril, Preeti Kshirsagar, Rashmi Mishra, Ronit Damania and Nandana Prabhu, "Resume Parsing and Standardization", International Journal of Computer Sciences and Engineering Vol.-3(3), PP(129-131) Mar 2015, E-ISSN: 2347-2693 © 2015.