Resume Parsing and Standardization

Pooja Shivratri[1*], Preeti Kshirsagar[2], Rashmi Mishra[3] Ronit Damania[4] and Nandana Prabhu[5]

[1,2,3,4,5]*Deptof Information Technology,*
K.J.Somaiya College of Engineering Mumbai University,India
www.ijcseonline.org

Abstract— Internet plays a crucial role in our daily life, from shopping to banking one can do anything just by a single click. Similarly, seeking jobs through online portals has become much convenient for the candidates as well as the recruiters. These portals help recruiter to find the perfect candidate and the candidate to get their perfect job. Automated and user-friendly software is being created which endows with a solution to parse all the resumes and provides a quality candidates required for the job. This paper deals with the parsing application developed for the resumes received in multiple formats like doc, docx, pdf, txt. When a company has a vacant position, it receives thousands of resumes for a single position. The concerned authorities have to go through each parameter of the resume and then select the candidates, for the interview. However, this system transforms original resumes into a standard format which contains only desired details of the candidate.

Keywords—Natural Language Processing, Metadata, Parsing, Standardization, Recruiter, Candidate.

I. INTRODUCTION

Recruitment is an important activity in any organization or a company. It is the process of finding and hiring the best-qualified candidate for a job opening, in a timely and cost effective manner. The recruitment process includes analyzing the job requirements, attracting employees, screening and selecting applicants, hiring, and integrating the new employee to the organization. At the same time, it is a time consuming and tedious job for the recruiters.

Resume plays an important role in recruitment process. A resume is a document used by persons to present their backgrounds and skills. It is a semi-structured document without any standard format [5][10]. The "Resume Parser" is a Web-based application which helps people like HR(Human Resource)manager, recruiters to parse resume and convert them to standard format. The standardized format contains desired details of candidate like Name, contact no, email, technical skills and Experience. The application should be easy to access and available 24 x 7. Given below are the sample formats of resume:

Fig. 1 Sample Resume 1

Fig.2 Sample Resume 2

Job seekers can upload their resumes using the web-portal after registration. HR manager, job recruiters/agency can access the uploaded resumes and convert them into standard format for better analysis. This standard format will act as the bases for the search results displayed to the HR manager, job recruiters/agency. The HR manager/ recruiters can search for the relevant candidate for available job position. The application will also suggest the eligible candidate for

corresponding job position depending on the resume uploaded.

In this paper we describe a system, which is capable of processing resumes in multiple formats and building an electronic database from the resume. The system aims to aid a large enterprise by removing the manual effort in screening resumes received by them to ascertain the suitability of candidate. The organization of the paper is as follows: In Section II., we have discussed the system which already exists and how it is different form our system. In Section III., we describe the system that is capable of automatically extracting relevant information from a resume and pushing the information into a database. In Section IV, the different module for the proposed system has been explained in detail. The extraction of relevant information is based on a set of natural language processing and metadata is also described in Section IV. The complete system is web enabled to make it reachable to a large number of people within the company.We conclude in Section V.

II.   EXISTING SYSTEM

There are various job portals that are available. In these sites the candidate must Sign Up first, after signing the candidate has to fill another form regarding his educational qualification, experience and various other details [8].
Various drawbacks of the existing portals are as follows:-

* In existing job portals, a candidate has to manually fill the job profile form which is a lengthy form and time consuming.
* The standardized resume is not provided by the system because of which recruiter has to go through every parameter of resume which is a tedious task.
* Resumes are written in different format that make it difficult for online recruitment companies to keep these data in their relational databases.

The point here is, if a candidate has a resume which has all the necessary details then why should he fill that lengthy form. In our system the candidate only has to upload his resume and our system will automatically find the details from the resume. Also, uploading the updated resume will update the candidate's details, but this is not the case in the existing portals. In the existing sites only uploading the updated resume will not do, the candidate must fill the lengthy form again to update his credentials.

III.   PROPOSED SYSTEM

In this paper, 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 flowchart for the proposed system is as shown below:
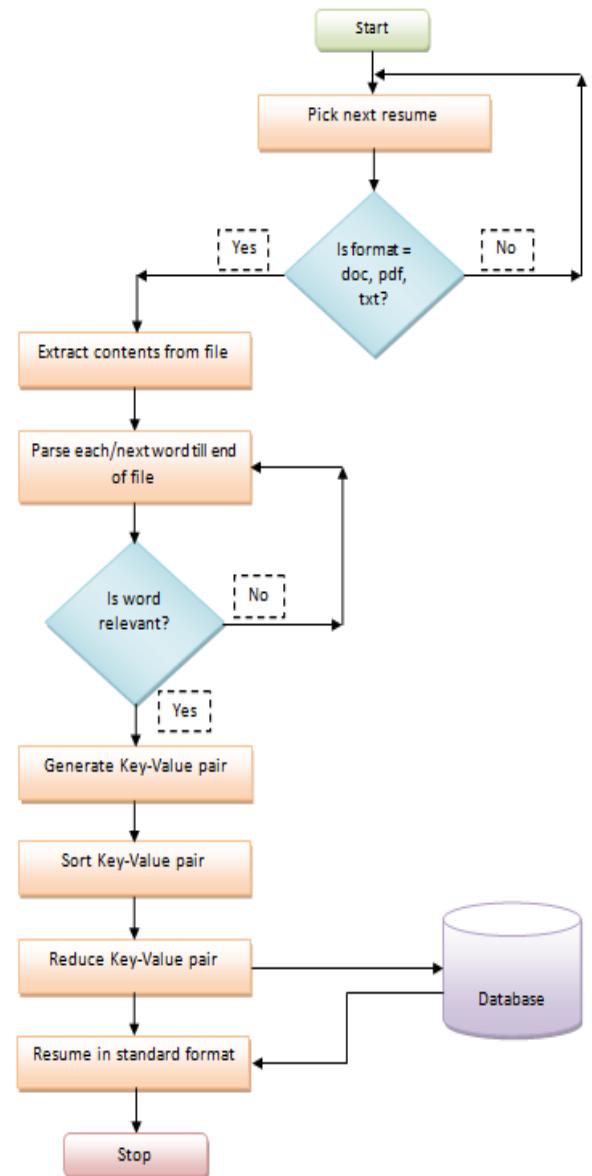
Fig.3 Flow Chart of Proposed System

As shown in Fig. 3, candidate has to upload the resume in formats like .doc, .docx, .pdf, .txt. The system will first check for the appropriate format and store it. It extracts the contents from the uploaded file and parses each word till the end of the file. While parsing it extracts only desired information from the resume and generates attribute-value pair for the purpose of standardization [1]. The attribute-value pair will be represented as NAME-XYZ, TECHNICAL SKILLS-JAVA, etc. During parsing, many duplicate key-value pairs are formed which are first sorted and then removed by selecting only one pair [9]. The key-

value pairs are then stored into the database and converted into a word document which can be viewed by candidate as well as the recruiters. The system keeps the copies of both original and the standardized resume.

IV. MODULES

This system will consist of three modules: Candidate (Job Seeker) Module, Job Recruiter Module and Resume Parser and Standardized Module.

A. The Candidate module will allow candidates to upload their resume, view and apply for appropriate jobs available. With Candidate module, any candidate will be able to register himself to portal, upload his resume and view available relevant jobs [2][6][7].

B. The Job recruiter module will allow recruiters to view standardized resume which was passed by the application. With this module, recruiter can add job details, view resumes uploaded by candidates, view standardized resume, view eligible candidates for vacant job positions [2][6][7].

C. The Resume Parser and Standardized module will parse the resumes uploaded by the candidates and convert them into standard pre-defined format. This module will find the relevant details like name, email, contact no, experience, technical skills required to fill the available job position.

The resumes will be uploaded to web server on one common location. The third module will scan, parse and standardize the resumes on FIFO basis. The original copy will be maintained and a new standardized copy of resume will be created and stored on server. Recruiter will be provided an option to view both the resume (original, standardized). Apart from this, the third module will also suggest the recruiters, the eligible candidates available for respective job position. This will ease the job of recruiters, otherwise they have to manually scan and read through entire resume and find out the best eligible candidate for available job position. For resume parsing, there are three approaches:

A. Metadata - In this approach, data is stored in database tables and then each word in the resume is matched with the data stored in the tables. If the match is found then that word is tagged and a Key-value pair is generated [3][4]. The limitation of this approach is that it is limited to the data stored in the tables. This approach is applicable where we have fixed set of domain values like technical skills, marital status, gender etc.

B. Natural Language Processing (NLP) - In this approach, meaning of word is derived from each word parsed and if the meaning corresponds to one in the list the application wants than the word is tagged and key-value pair is generated [3][4]. This approach is applicable where we don't have fixed set of values but have fixed format such as regular expressions [9]. Examples include email id, mobile no, date etc can be identified using this approach.

C. Hybrid Approach - This approach is combination of the above two approaches where certain details are extracted by using metadata approach while other details extracted through NLP approach.

V. CONCLUSION AND FUTURE WORK

With increasing use of electronic media to seek jobs and fill vacancies, there is a need for a tool which is capable of extracting useful information from a free form resumes. In this paper, we have proposed a functional and automatic resume parser for both structured and unstructured resumes to aid electronic search. A mix of natural language processing techniques and metadata were used to build information extraction modules to aid extraction of useful information from resumes. This system is really a boon to HR professionals as it is known as an unparallel one, which will be the revolutionary parsing system. It resolves into its components or parts of the resumes and rearranges in the categories. The resume parser thus, reorganizes the available database in a streamlined and accessible way. It helps in computing thousands of resumes within seconds. This project concentrates on the IT industry, and classifies the candidates on the basis of the technical skills; this project can be extended in other sectors of industries as well. Also, this project is limited to the parsing of the resumes in most widely used formats like .doc, .docx, .pdf and .txt but various other formats can also be added to the project in near future. The main aim of this project was to find the appropriate candidate for the particular job purely on the basis of the resume he has uploaded.

REFERENCES

[1] C# coding, http://www.c-sharpcorner.com/beginners/ last visited :February 2015.

[2] Christian Nagel, Jay Glynn, "**C# 5.0 and .NET 4.5.1**", Wiley, Page No: 24-63, 210-228, 415-457.

[3] Sovren Résumé / CV Parser, http://www.sovren.com/ last visited: November 2014.

[4] Daxtra CVX: ,http://www.daxtra.com/, last visited: November 12, 2014.

[5] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao bo, Lin Zhiqing, "**Resume Parser: Semi-structured Chinese document analysis**", Computer Science and Information Engineering, 2009 WRI World Congress on, Volume-5, Page No(12-16), 2009.

[6] Creating Web forms in visual studio using asp.net, https://www.youtube.com/watch?v=wU6roQF9uUE , last visited: January 2015.

[7] Learning the basics of asp.net, http://www.w3schools.com/aspnet/ , last visited : October 2014.

[8] Existing Portals , Naukri.com, http://www.naukri.com, last visited: February 2015.

[9] Jigyasa Nigam and Sandeep Sahu, "Fast and Effective System for Name Entity Recognition on Big Data", Int. Journal of Computer Sciences and Engineering, Volume-3, Issue-2, Page No(31-35), Jan 2014

[10] Sneha Lohbare and Ashwini Meshram, "A Review on Document RetrivalRetrieval From Unstructured Text", Int. Journal of Computer Sciences and Engineering, Volume-2, Issue-11, Page No(76-80), Nov 2014