

Review: Resume Extraction & Candidate Recruitment System

Vipin Kumar¹, Vikram Nath², Payal Piwalatkar³, Heena Shaikh⁴, Sneha T. Shinde⁵

¹ Student, Computer Science Department, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India

² Student, Computer Science Department, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India

³ Student, Computer Science Department, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India

⁴ Student, Computer Science Department, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India

⁵ Professor, Computer Science Department, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India

ABSTRACT

Automated Resume Extraction and Candidate Selection System is a product which can be best suited for any organization's recruitment process. The system will be robust enough which will automatically extract the resume content and store it in a structure form within the Database. Classification algorithm (Decision tree) will be run on the profiles to identify profile Categories or classes. Even the employer can suppose to specify his criteria and also decide the portentous level. As the internet is growing rapidly, there is huge amount of electronic text that is increasing rapidly. This brings the favor of reaching the information sources in a cheap and quick way. Keywords are handy tools as they give the shortest summary of any document. But they are rarely composed in the texts. There are proposed methods for automated keyword extraction. This paper also specifies such a method, which identifies the keywords with their frequencies and positions in the training set. It uses decision tree algorithm with supervised learning.

Keyword : - Clustering, Classification, Data Mining, Data Preprocessing, Decision Tree.

1. INTRODUCTION

The purpose of this project was to build Resume Extractor and Candidate Recruitment System which will be built on Google's Cloud. Large enterprises and head-hunters receive several thousands of resumes from job applicants every day. HRs And Managers go through a number of resumes manually. Resumes or Profiles are unstructured documents and have typically number of different formats (eg: .doc, .txt). As a result manually reviewing multiple profiles is a very time consuming processes. How to get the Appropriate Candidate in the right jobs at the correct time. This is a major problem faced by MNC today in the market.

Now a day's too many job portals are attainable but the basic problem in available system are as it required manual endeavour for both candidates and Employers. Candidate has to provide complete information in given text field and employer also needs to apply many filters to select the candidate. Even if the Employer has applied several filters he would get as many resumes even going through it and selecting candidates is very ineffective and time-consuming task.

Some expensive extraction systems are available in the market that also do the same task of searching the keywords and has many extraction drawbacks like enforcing candidates to fill templates and keep on updating the templates as per job profiles.

Not a single intelligent and self-analyzing tool is obtainable in the market which has benefits of data mining and machine learning as well as which will take consideration of information present in social networking.

The highlights of this paper are arranged as follows. In section 2, we introduce the existing system. In Section 3, the benefits of online recruitment. In Section 4, describes proposed system. In Section 5, describes literature survey. In section 6, we will discuss the conclusions and possible future work.

2. EXISTING SYSTEM

In this section, we present an overview of the existing system architecture and discuss its main modules.

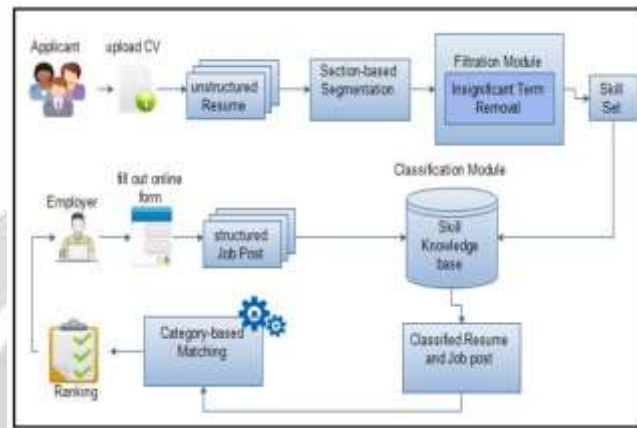


Figure -1: Architecture of the Existing System

As shown in Figure 1, the existing system comprises several modules that are organized as follows. First, a Section-based Segmentation module is used to extract a list of candidates matching concepts, in addition to information such as personal, education, experience and applicant's employment history. Second, the Filtration module refines the concept lists by removing insignificant terms that don't contribute in the matching process. Third module of the proposed system takes a set of skills extracted from both resumes and job posts as input in order to classify them under their corresponding occupational categories. At this step, we exploit an integrated occupational categories knowledge base which combines two main classification schemes: DICE and O*NET. Then, the Category-based Matching module takes the lists of skills from both resumes and job posts to construct semantic networks by deriving the semantic relatedness between their concepts. Finally, the matching algorithm takes the semantic networks as input – as long as they are in the same space - and produces the measures of semantic closeness between them as an output.

3. BENEFITS OF ONLINE RECRUITMENT

HR's and Technology experts say executing computer-based systems that eradicate the need for paper can reduce the time and money. Productivity and efficiency are the key reasons for people to go to a paperless system. Such systems surely reduce the storage space required. And they are designed to contain the vast and changing array of acquiescence issues. Many HR departments have removed most of their paper from payroll administration and different functionalities but hesitated to do so with processes such as performance tuning and management. Using the paper habit has been a pain point for HR for a number of years. "Paperless is somehow more than just converting documents to digital. It's also about what you can do with the data." Implementation and execution are a challenge. It takes lot of time to explain it to users in a team meeting, through group demonstrations, via one-on-one sessions and even create a PowerPoint presentation to answer the queries. Even i take time basically to scan each and every single document. But once, HR department starts using it, it will make life of HR so much easier. Technology has changed the whole process of hiring people. It has shifted the liability and ownership.

All aspects of candidate recruitment management from searching for imminent applicants to new employee adaptation now have the ability to fit into an online recruitment system. Few employers, employment, and businesses work in recruiting job seekers for placement can make today's recruitment practices without executing some form of online recruitment system, even if only an elementary system. Many MNC, in fact, now require

candidates to complete an initial application for employment and placement online, or last but not the least submit a resume to an online database.

In the early time of online recruitment, the idea simply meant using online supplies to find candidates for different job positions. Job resume databases makes employers and employees to find and contact each other. Once contact was entrenched, the remaining tasks convoluted in the hiring process were managed via face-to-face interviews and other traditional methods. Today, online recruitment comprises of job boards, advertising, web-based applications, social recruiting, and pre-employment assessment tools. Additionally, software packages developed to enable recruiters and human resource personnel to increase the hiring process through automated database searches, auto desponds, and estimation tools integrated into web-based employment applications.

With the eruptive growth of businesses that exist entirely online, the demands on online recruitment tools have also grown. Companies have the capability to conduct all aspects of the business using freelance contractors and employees working parenthetically. As such, these businesses often do not have the proficiency or means to organize traditional in-person interviews and other human resources responsibilities. For these businesses, the use of and availability of online recruitment tools is inestimable. Applicants are directed to apply online, submit their work samples or complete assessments, and upon selection, address necessary documentation to begin work.

4. PROPOSED SYSTEM

Automated Resume Extraction and Candidate Selection System is a product which can be best suited for any organization's recruitment process.

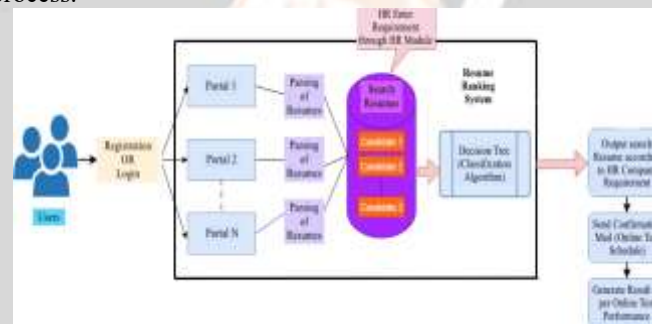


Figure -2: Architecture of the Proposed System

Users can register or login into our website. Through our website users add there different job website accounts, which is integrated by our system into single account. Which can be manage easily. The system will be robust enough which will automatically extract the resume content and store it in a structure form within the Database. The employer can also specify his criteria and also decide the importance level. Classification algorithm (Decision Tree) will be run on the profiles to identify profile categories or classes. Based on classification some result (list of users which meet the requirement of employer) is generated. Employer can view the profile of particular user if they want. It select the users & send the confirmation mail for selection process in which Online Test schedule is given. As per Online Test performance users can be selected for personal interview.

5. LITERATURE REVIEW

Resume extraction from huge number of resumes is useful in many real-world applications. Because resumes are a kind of data. The data that is stored in resumes is simple, but complex for machine to understand and there is need to store this data in such a manner that can be searched and extracted easily whenever required. Elimination of the use of paper is one of the steps to progress towards a world of electronics.

We referred different base papers for implementation and development of our application. We took some of the ideas and information from those papers which helped us for designing this application. Some papers that we referred are as follows,

In [2014] Gauri R Rao, Renuka S Anami published a paper Automated Profile Extraction and Classification. It stated that we are investigating the issues related to the development of approaches to improve the performance of resume selection process. We have elongated the consideration of special features and contemplated an approach to

recognize resumes with special skill information. In this literature, the notions of special features have been applied to elevate the process of product selection in Ecommerce environment. In this system, we have proposed an approach by considering only the skills related information of the resumes. The experimental results on the real world data-set of resumes show that the proposed approach has the possibility to improve the process of resume selection.

In [2015] Mrunmayee Hatiskar, Arati Tayade, Rajashree Garud, Sayali Gardi published a paper Resume Parsing and Standardization Using Semi-Structure Algorithm. It stated that resume can be used for many reasons, but the main reason is used to secure employment. The resume is personal and academic information of employee, which a suitable employer sees related the job seeker and used to screen applicants often followed by an interview. This project is deals with the parsing application developed for the resumes received through emails in multiple formats like Docx, Document, text etc.

In [2015] Satish L. Thombare, Shweta V. Raja published a paper Automated Resume Extraction and Candidate Selection System. It stated that we collected the resumes from the students of various courses applying for a job. Students and freshers applying for the job will send the resumes in multiple formats (docx, pdf, text, etc.) and the information from all the resumes will be extracted in the database by using the decision tree classification and regression techniques in data mining, and accordingly the students will be classified according to their qualification. This project is a proposed model to find an appropriate evaluation method for the classification of students and predicting the placement opportunity in an enterprise or a firm. This model will enhance the firm to select the students in a beneficial way to their performance in the academics.

A study proposed by Connerley, Carlson, & Mecham (2003) on the indication of differences in applicant pool quality addressed the research need by examining the allure outcome of firms facing head-to-head for recruits for similar positions. Results of this analysis suggest that applicant quality can vary essentially within and across job families.

A proposed study by Chapman and Webster (2003) in their research on the execution of technologies in recruiting, and selection processes for job candidates conducted in USA and many other sections found that most organizations implemented technology-based recruitment and selection tools to improve and increases the efficiency, enable new assessment tools, reduce costs, standardize systems and expand the applicant pool.

According to the Pew Internet Research reported by NAS insights (2006), about 72 percent of American adults were online. That translated to over 145 million people. This was a large amount of audience, and the Internet hence proved to be an core part of employee recruitment because there was no faster, simpler, or more cost-effective way to reach thousands of qualified candidates. In fact, 44 percent of online Americans were looking for information about a job. The Internet allowed HR's and Managers to reach these candidates 24 hours a day, 7 days a week.

6. CONCLUSIONS AND FUTURE SCOPE

In this age of technology, there is a huge amount of data and it keeps on increasing day by day. Now a days all recruiter and job seekers are moving toward digitization, so extraction of resumes become difficult. This system can be used for resume extraction from huge number of resumes so that proper candidate should be selected. System is helpful to the recruiter as well as job seekers and time saver. Online recruitment is easy and simple way of recruiting the employees in the organization compare to traditional method. Internet has made an awful impact on the functioning of human resource department. If online HRM, is been implemented in the organization, the working of the HR department will be at ease. There will be minimum utilization of paper, less storage required, less wastage of time in documentation, reduction in manpower utilization, time saving, and data can be used as and when it is required and in multi-ways.

In future, we can expand the system where it will combine/connect all the websites which will provide jobs to the job seekers. Providing online tests so there will no wastage of time and also to extract resume content in different formats.

7. REFERENCES

[1]. Jongwoo Kim, Daniel X. Le, and George R. "Naïve Bayes Classifier for Extracting Bibliographic Information from Biomedical Online Articles", National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA.

- [2]. Ajay S. Patil, B.V. Pawar “Automated Classification of Web Sites using Naive Bayesian Algorithm”, Proceedings of the International Multi-Conference of Engineers and Computer Scientists 2012 Vol I,IMECS 2012, March 14-16, 2012, Hong Kong.
- [3]. Md. Faisal Kabir “Enhanced Classification Accuracy on Naive Bayes Data Mining Models”, International Journal of Computer Applications (0975 – 8887) Volume 28– No.3, August 2011.
- [4]. Mauricio A. Valle , Samuel Varas , Gonzalo A. Ruz “Job performance prediction in a call center using a naive Bayes classifier”, Facultad de Ciencias Económicas y Administrativas, Universidad de Valparaíso, Santiago, Chile, 2011.
- [5]. Gla’ucia M. Bressan “Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop Universidade de São Paulo, Departamento de Engenharia Elétrica, 13566-590 São Carlos, SP, Brazil 2009.
- [6]. S.L. Ting, W.H. Ip, Albert H.C. Tsang. “ Is Naïve Bayes a Good Classifier for Document Classification?”, International Journal of Software Engineering and Its Applications Vol. 5, No. 3, July, 2011.
- [7]. Yasin Uzun “ Keyword Extraction Using Naive Bayes”, Bilkent University, Department of Computer Science, Turkey, yasinu@cs.bilkent.edu.tr
- [8]. Binal A. Thakkar, Mosin I. Hasan, Mansi A. Desai ”Health Care Decision Support System For Swine Flu Prediction Using Naïve Bayes Classifier”, International Conference on Advances in Recent Technologies in Communication and Computing, India, 2010
- [9]. http://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [10]. https://en.wikipedia.org/wiki/Decision_tree

