Volume 3, Issue 11, November 2018 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

A Review on Automated CV Classification

Bhamare Priyanka
Dept. of Computer
G.C.E., Nagaon
Dhule , M. H. India

Mali Puja
Dept. of Computer
G.C.E., Nagaon
Dhule , M. H. India

kachave Kalyani
Dept. of Computer
G.C.E., Nagaon
Dhule , M. H. India

Prof. S. U. More
Asst. Prof. Dept. of Computer
G.C.E., Nagaon
Dhule, M. H. India

Abstract-: The task of finding the right candidate for a particular job can be a very tedious task for the HR department of an organization. Going through hundreds of resumes is not an easy task. No one has enough time to go into the details of any resume. Shortlisting resume based on job requirement is also important task from point of view of HR department. As job designation varies the requirement also changes and HR members need to find out all these from hundreds of resumes which include lots of efforts and time. To simplify this process we propose a system which will automatically calculate a score for each resume and will classify them in clusters which will make easy for HR department to select appropriate candidate. The score for resume will be calculated based on the weighted scheme set by HR department at time of publishing the job requirement. Thus the system will prove to be an effective system for all organization and will greatly reduce efforts and time of HR department.

Keywords: Resume parser, resume analyser, text mining, K-Mean

I. INTRODUCTION

Corporate companies and recruitment agencies process numerous resumes daily. This is no task for humans. An automated intelligent system is required which can take out all the vital information from the unstructured resumes and transform all of them to a common structured format which can then be ranked for a specific job position. Parsed information include name, email address, social profiles, personal websites, years of work experience, work experiences, years of education, education experiences, publications, certifications, volunteer experiences, keywords and finally the cluster of the resume (ex: computer science, human resource, etc.). The parsed information is then stored in a database (NoSQL in this case) for later use. Unlike other unstructured data (ex: email body, web page contents, etc.), resumes are a bit structured. Information is stored in discrete sets. Each set contains data about the person's contact, work experience or education details. In spite of this resumes are difficult to parse. This is because they vary in types of information, their order, writing style, etc. Moreover, they can be written in various formats. Some of the common ones include '.txt', '.pdf', '.doc', '.docx', '.odt', '.rtf' etc. To parse the data from different kinds of resumes effectively and efficiently, the model must not rely on the order or type of data.

II. Literature Review

1) Resume Parser with Natural Language Processing paper 2017: To design a model this can parse information from unstructured resumes and transform it to a structured JSON format. Also, present the extracted resumes to the employer based on the job description.
Limitation in system:
1) The system works in natural language processing which requires dataset of keywords. So the system efficiency is directly proportional to the quality of dataset of keywords.
2) The system only transforms the unstructured data into meaningful data it does not provide any classification of resume.
3) The three stages discussed are inputs to each other so if wrong output is generated in any stage the system output will change greatly and system will not work correctly.

2) Online Resume Parsing System Using Text Analytics (2015): To design a system that will pare and shortlist the resume based on text analysis on the data present in the resume.
Limitation in system:
1) The system depends on the knowledge base which is set of keywords which are pre stored in dataset. IF the system knowledge is not properly trained or maintained it will not calculate score for resume or calculate wrong score and thus system will not work as expected.
2) If the keywords are not properly found out by the system the calculation of score become difficult.
3) Managing the dataset of keywords is very difficult task.

III. PROBLEM STATEMENT

To develop a system which will allow HR department to publish job opening with customized weighted scheme according to job requirement and provide a portal for candidate to upload their CV/Resume and finally provide the HR team with a classified view of all the CV available for particular job.

IV. SYSTEM OBJECTIVES

1) To provide a classified view of CV to HR department

2) To eliminate the manual process of CV classification and shortlisting

3) To reduce time and efforts of HR department in CV selection process

4) Provide HR team option to customize each and every job before uploading as per requirement and skill set required.

Hardware Requirements

1. Processor	:	Pentium IV
2. Hard Disk	:	12GB
3. RAM	:	512 MB or more

Software Requirements

1. Operating System	:	Windows XP or More
2. User Interface	:	HTML
3. Client-side Scripting	:	JavaScript
4. Programming Language	:	Java
5. Web Applications	:	JDBC, Servlet& JSP
6. Database	:	MYSQL 5.0

V. PROPOSOSED APPROACH

The proposed system is a client server architecture based system which will use java servlets for client server communication. The front end of system will be developed using the java swing components and backend of system will be developed in java. The system database will consist of relational MYSQL database. As shown in the above figure the system will consist of 3 major modules as follows:

1) HR/Admin Module
2) Candidate Module
3) Web Server

HR/Admin Module:
This module will be run at the HR or admin end. This will be desktop software whose front end will be developed using java swing components. This module will be well protected with username and password and the password will be secured with SHA-1 algorithm. Admin can customize the job as per requirement and assign weight and upload the job. Admin will get a classified view of all the CV submitted by candidates for which k-means clustering algorithm will be used.

Candidate Module:
This module will be run at the candidate end. This will be desktop software whose front end will be developed using java swing components. The candidate will first need to register into the system and then will be presented with an interface where he will answer and select few options which will be part of the candidate CV. The candidate on submission of resume will get to know the score of CV.

Web Server:
The webserver will be developed using java servlets and will be run by using Apache Tomcat application server. Web server will be responsible for all the request and response handling mechanism of system. It will manage the database and provide appropriate data to admin of the system.

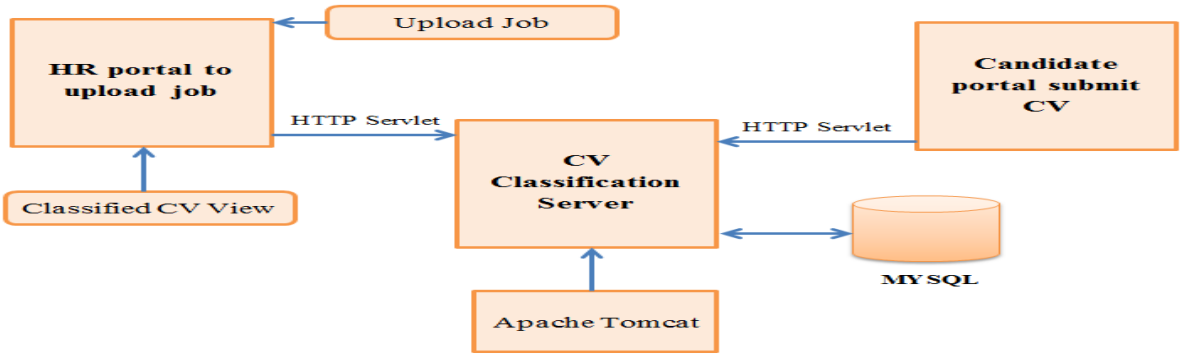

Fig: 1-System Architecture

Algorithms:

K-Means Clustering:

1) K means is unsupervised machine learning algorithm.

2) K means divides the given dataset into k clusters. Where k means number of clusters.

3) This algorithm will cluster candidate resume score in k clusters.

Steps included in clustering:

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4) Recalculate the new cluster center by taking means of data points in that cluster

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3)

CONCUSION

The problems that the HR team face while shortlisting candidates CV are very time consuming and involves lots of human efforts. This manual approach sometimes even give rise to human errors. So using our proposed system both efforts and time can be reduce in CV shortlisting process along with increase in efficiency and accuracy. Considering all the above features our system will definitely prove beneficial to organization and save some precious time and efforts of their HR department.

REFERENCE

1. Swapnil Sonar, Resume Parsing with Named Entity Clustering Algorithm, IEEE Research, may 2012, http://www.slideshare.net/swapnilmsonar/resume-parsing-with-named-entity-clustering-algorithm
2. Sovren Resume/CV Parser, http://www.sovren.com
3. Connectifier, http://www.connectifier.com
4. Rchillies, http://www.rchillies.com
5. Belong.co, http://www.belong.co
6. ALEX System , http://www.hireability.com/alex/
7. http://www.turborecruit.com.au/intelligent-searching/
8. http://www.revolvy.com/main/index.php?s=Parse%20tree
9. Student Thesis "Information Quality Management in Information Extraction:A Survey"
10. http://www.rn.inf.tu-dresden.de/uploads/Studentische_Arbeiten/Belegarbeit_Jansen_Nicolas.pdf
11. F. Ciravegna, "Adaptive information extraction from text by rule induction and generalisation," in Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI2001), 2001.
12. A. Chandel, P. Nagesh, and S. Sarawagi, "Efficient batch top-k search for dictionary-based entity recognition," in Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE), 2006.
13. S. Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data. Morgan-Kauffman, 2002
14. M. J. Cafarella, D. Downey, S. Soderland, and O. Etzioni, "KnowItNow: Fast, scalable information extraction from the web," in Conference on Human Language Technologies (HLT/EMNLP), 2005.
15. M. J. Cafarella and O. Etzioni, "A search engine for natural language applications," in WWW, pp. 442–452, 2005.
16. https://www.ijircce.com/upload/2016 /april/218_ Intellig ent.pdf
17. https://www.tutorialspoint.com/compiler_design/images /token_passing.jpg
18. http://www.nltk.org/book/tree_images/ch08-tree-6.png