

Resume Extraction And Candidate Recruitment System

¹Vipin Kumar, ²Vikram Nath, ³Heena Shaikh, ⁴Payal Piwalatkar, ⁴Sneha T. Shinde

^{1,2,3,4,5}Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Pune, 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.

IndexTerms -Clustering, Classification, Data Mining, Data Preprocessing, Decision Tree.

I.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 1.1, we introduce the existing system. In Section II, the proposed system. In Section III, describes mathematical model. In Section IV, describes results of our system. In section V, we will discuss the conclusions and possible future work.

1.1 EXISTING SYSTEM

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

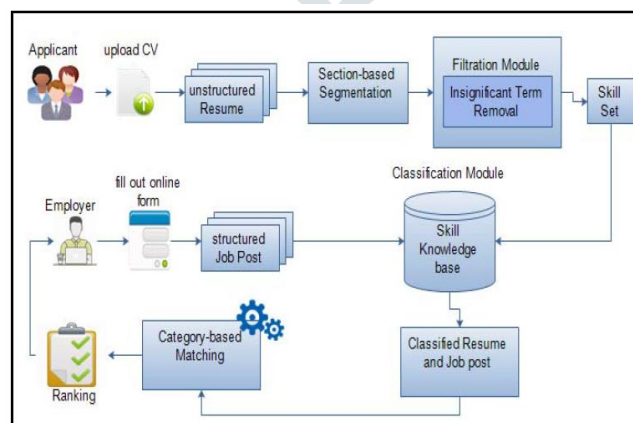


Figure -1: Architecture of the Existing System

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.

II. PROPOSED SYSTEM

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

The proposed system is a web-based application developed from XAMPP and Java.

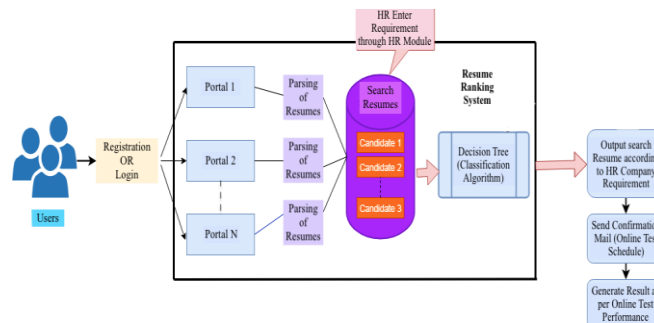


Figure -2: Architecture of the Proposed System

Users can register or login into our application. Through our application users add there different job website accounts, which is integrated by our system into single account. Which can be manage easily.

Decision Tree is used for Resume Extraction. The system can extract resumes based on keywords(Skill Sets provided by Recruiter). The first step is to Register or login by recruiter. Then, the recruiter has to specify the requirements and submits it to the application. The application accepts the requirements and passes it to the Decision Tree which performs the extraction based on given requirements.

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.

III. MATHEMATICAL MODEL



Figure -2: MATHEMATICAL MODEL

- $M = \{Q, \Sigma, \delta, q_0, F\}$
- $Q = \{S1, S2, S3, S4, S5, S6, S7, S8\}$
- $\Sigma = \{Resume\}$
- $q_0 = S1$
- $F = S8$
- $\delta(S1, Upload Resume) = S2$
- $\delta(S2, Submit Resume) = S3$
- $\delta(S3, HR Enter Requirements) = S4$
- $\delta(S4, Submit Requirements) = S5$
- $\delta(S5, Resumes Extracted) = S6$
- $\delta(S6, Send Online Test Schedule to Extracted Candidates) = S7$
- $\delta(S7, Passed Candidates are Selected for Interview) = S8$
- $S8 = Candidates Selected for Interview$

IV. RESULTS

The main page of the web-based application is as follows:

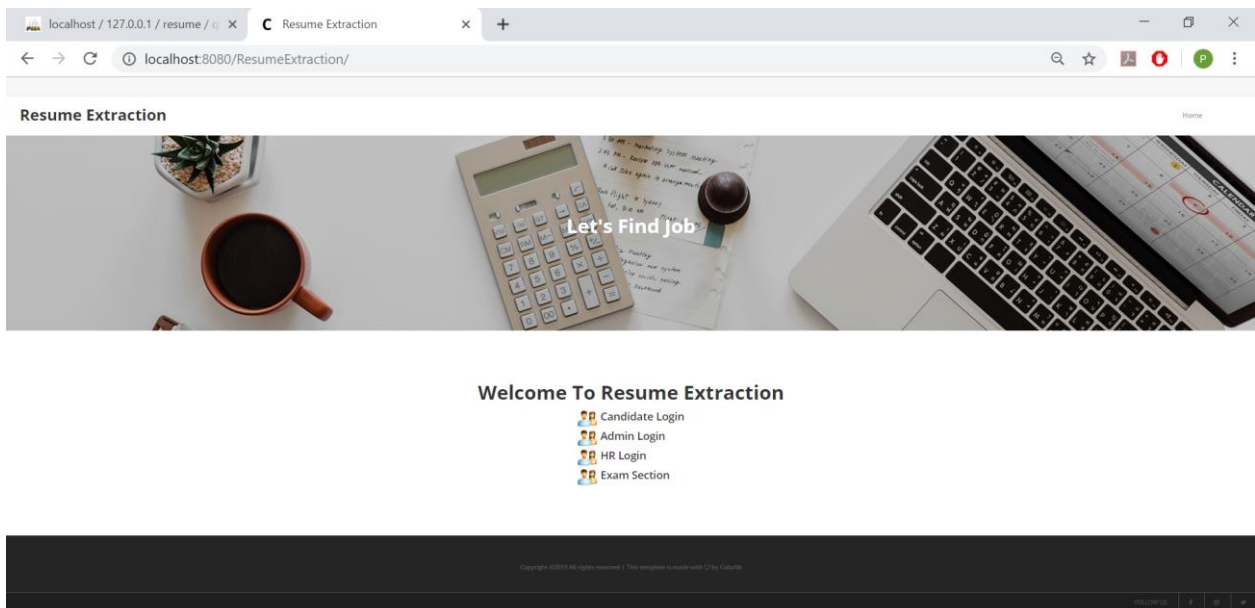


Figure : Home Page

The user selects a Login option from the given options.

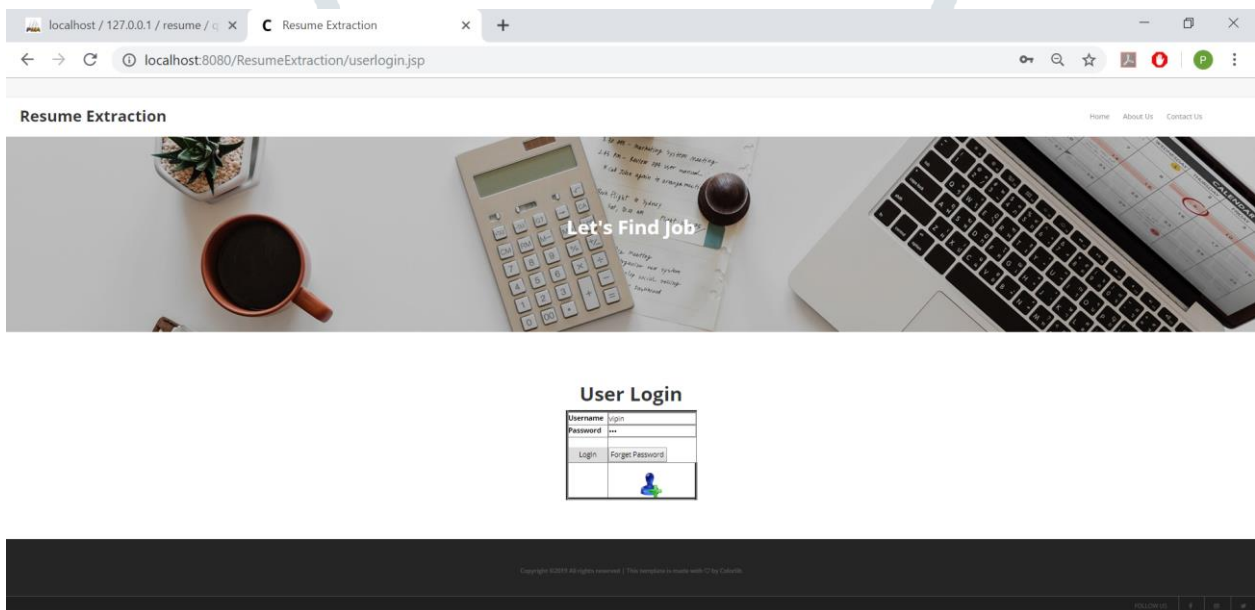


Figure : User(Candidate) Login/Register Page

Here, user login if they had account with us. If not then they just register it with some simple form.

This account is managing different Job accounts(Naukri, Monster, Shine, etc) which user already have. They just have to login with that account.

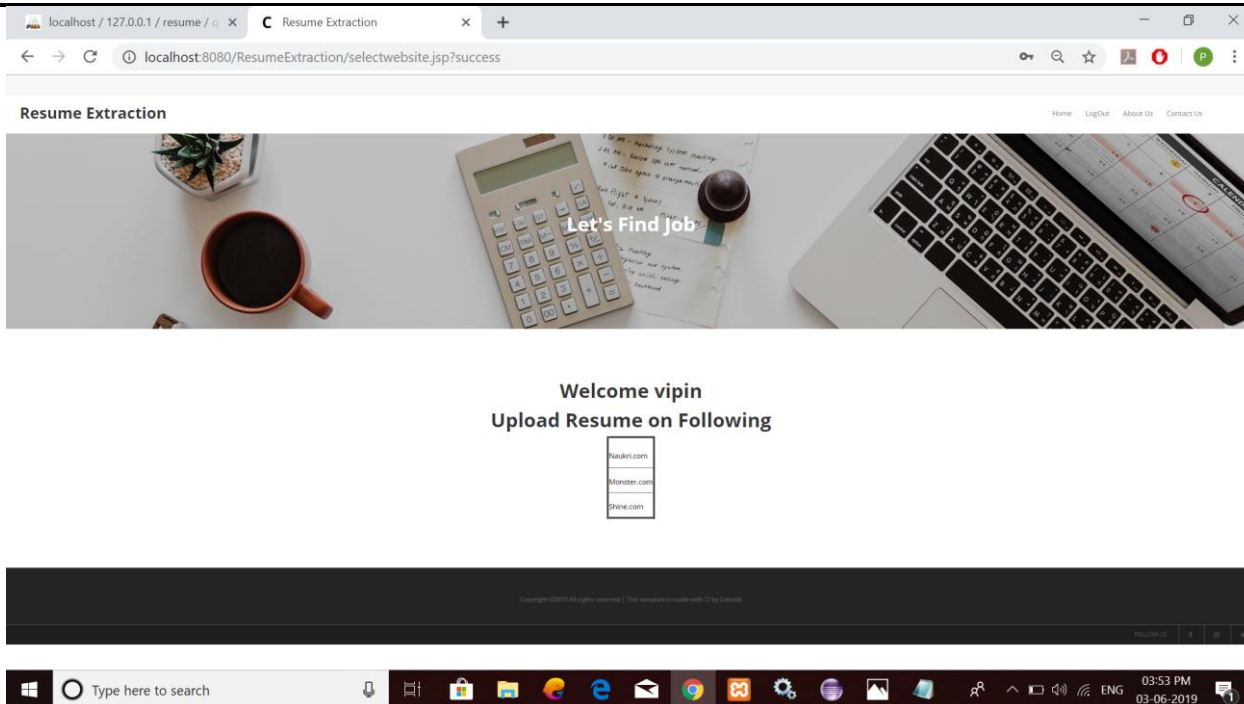


Figure : After Main Login Page Selection of Job Sites to Continue

User have to login with Job accounts they have or just register if they don't.

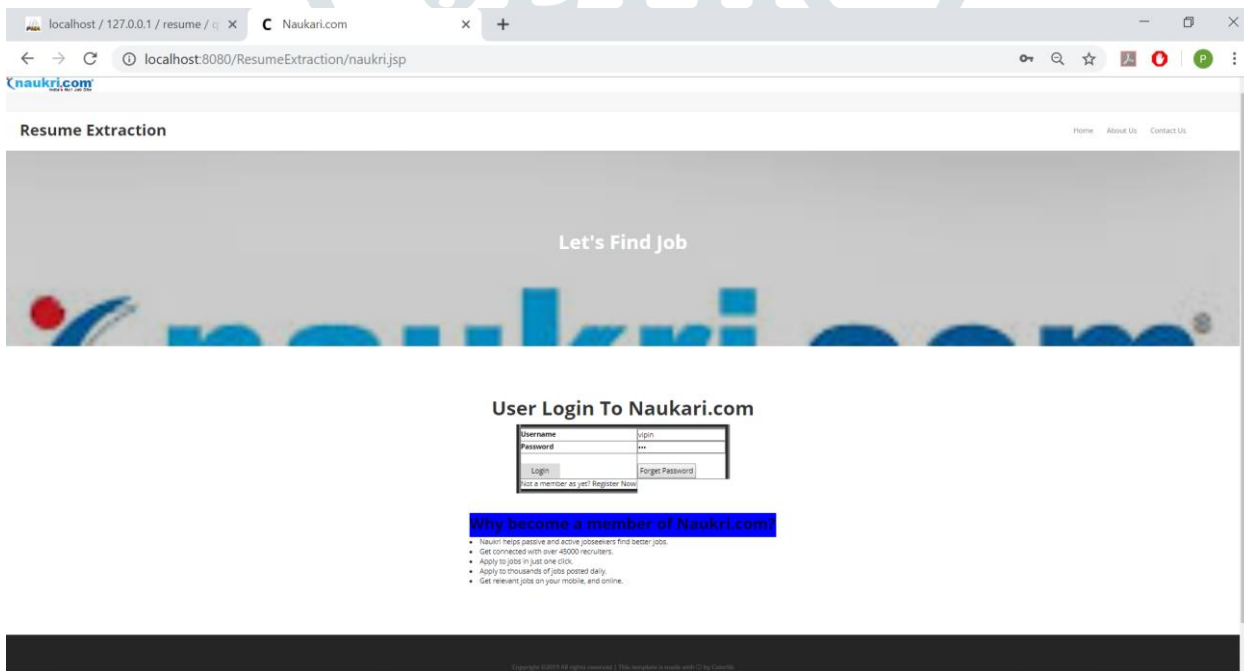


Figure : Naukari.com Login/Register Page

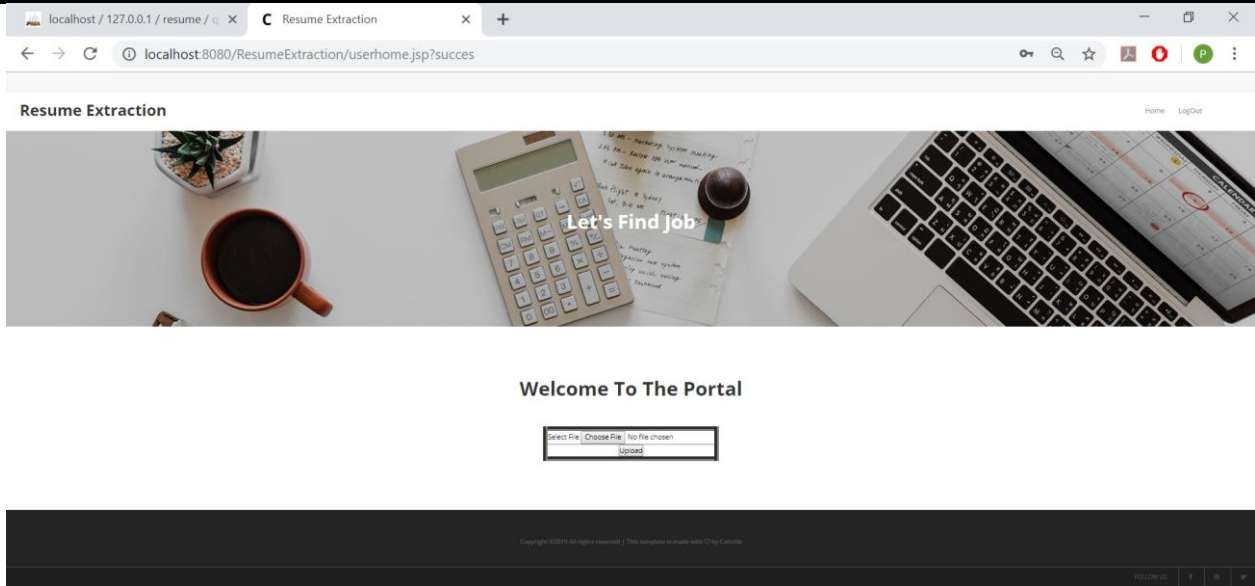


Figure : After Login to any Job Site Resume Upload Page

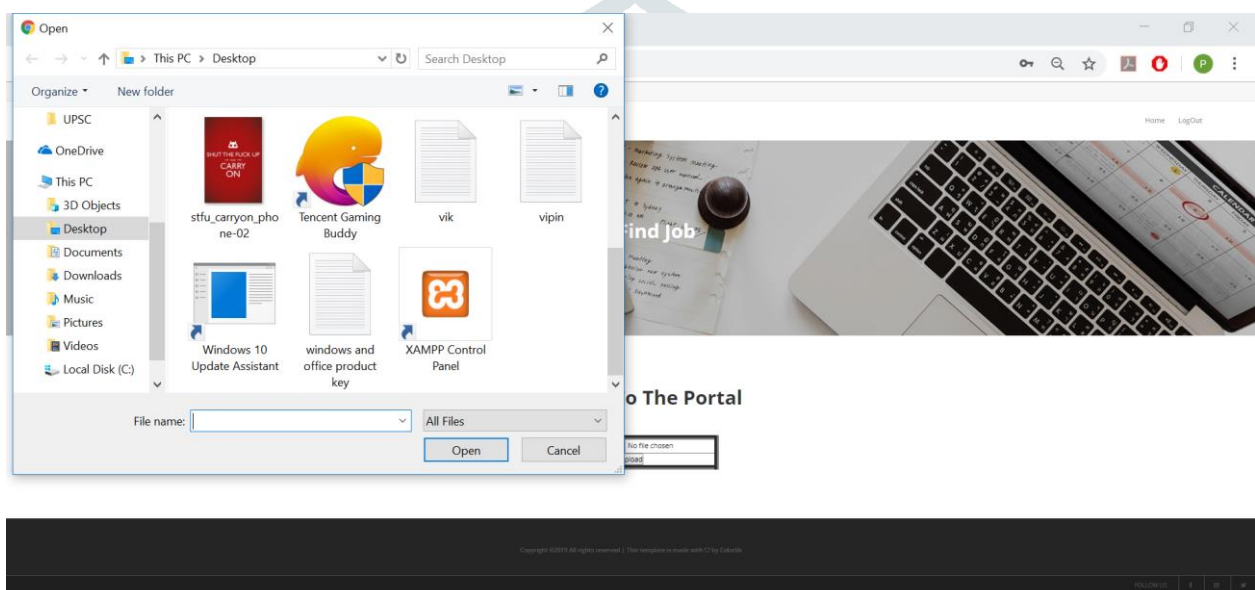


Figure : Resume Upload

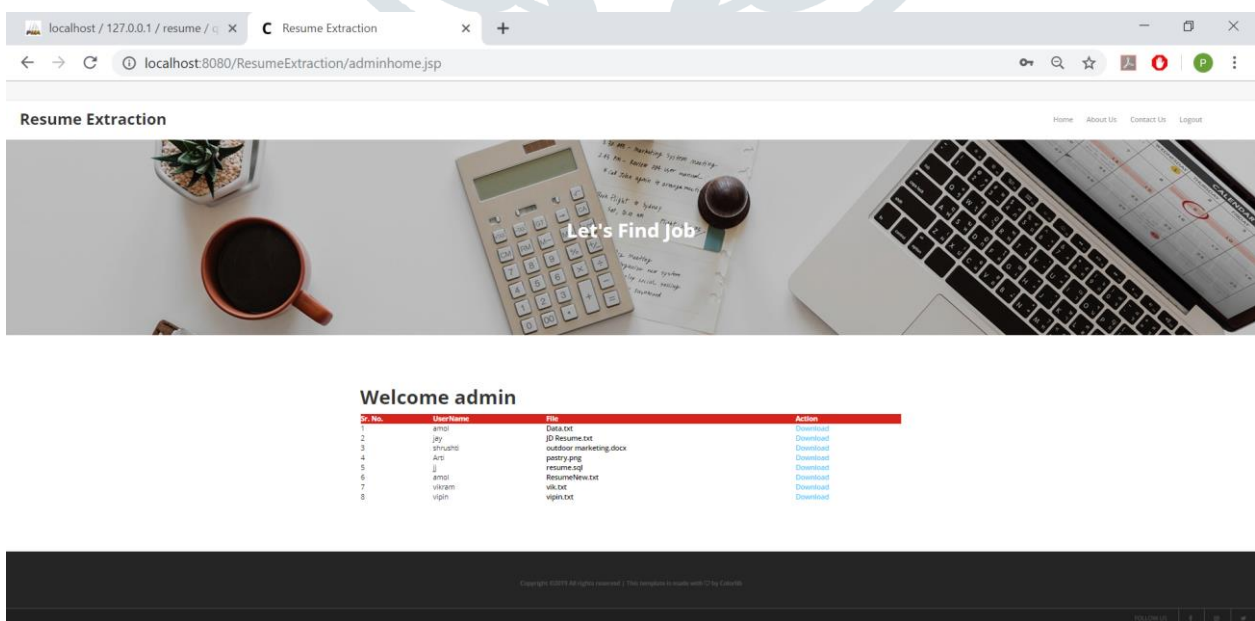


Figure : Admin Page

This login is for admin of this site.
 From here admin can see all users profile, resume, etc.
 Download the resume of particular user.

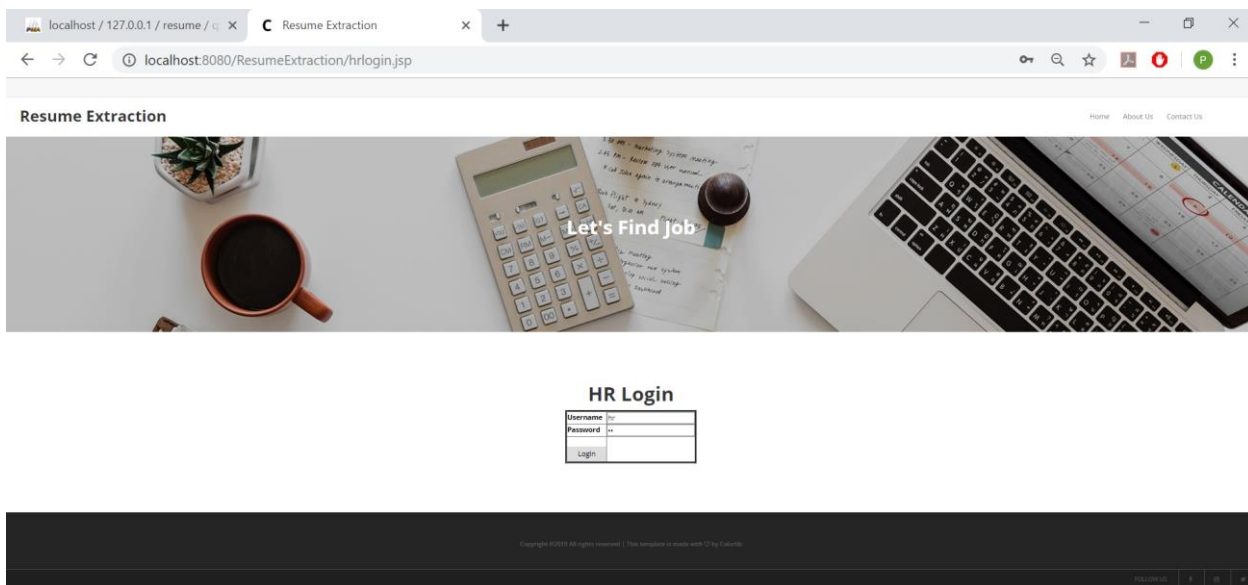


Figure : HR(Recruiter) Login/Register Page

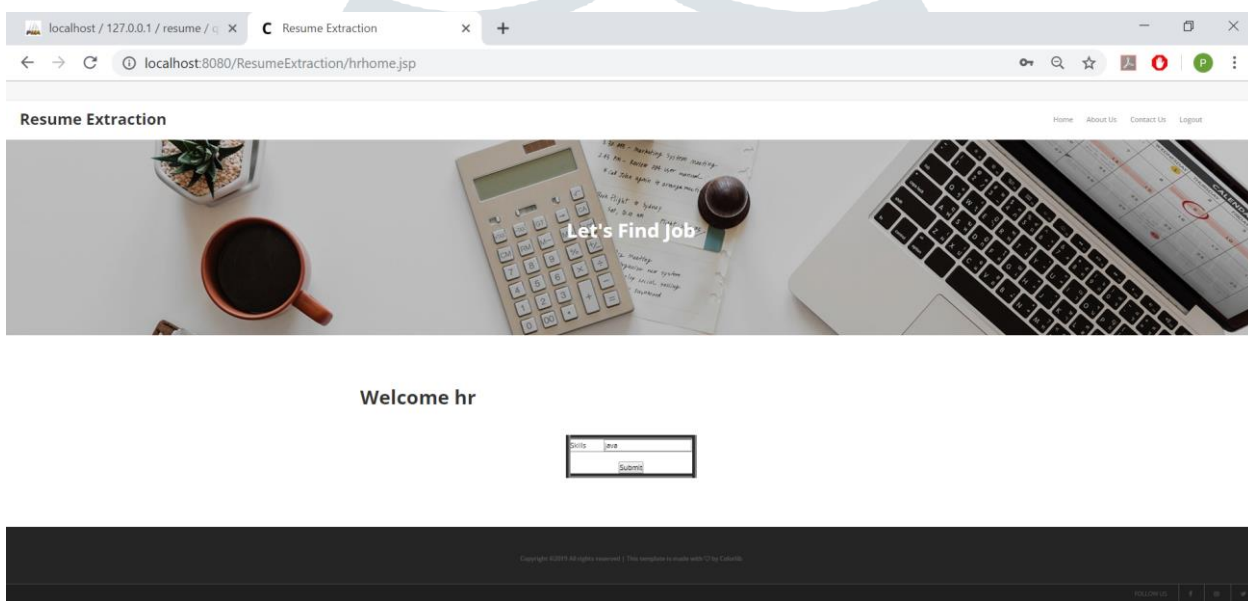


Figure : HR(Recruiter) Page

Here, Recruiter sets the skill sets as per there company requirements.

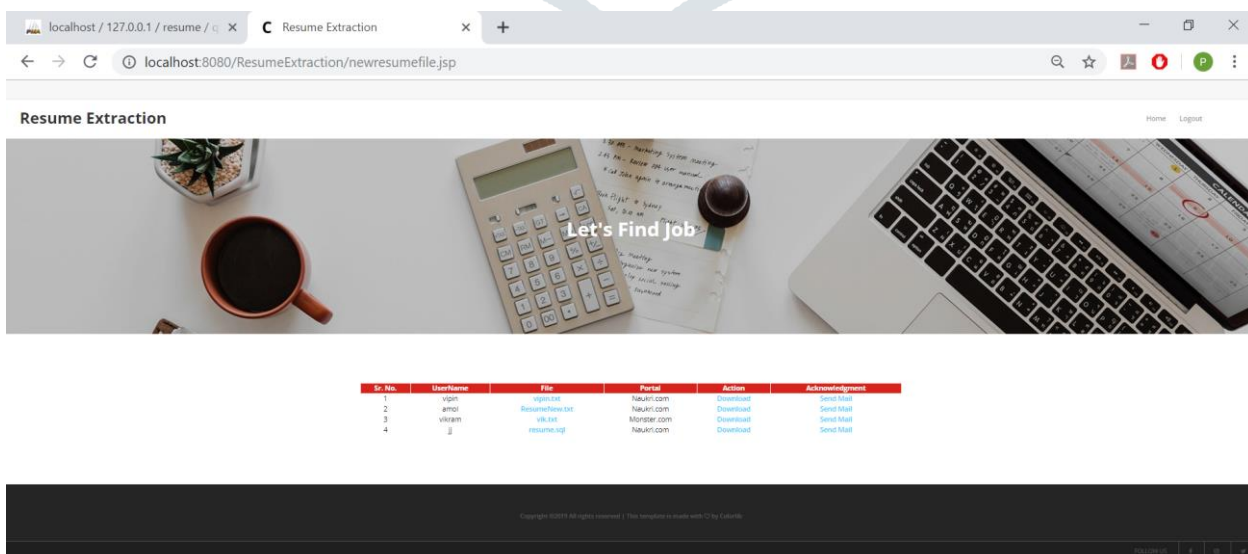


Figure : Extracted Resumes

These are the list of resume which are matched the requirements of recruiter. These can be extracted with the help of Decision Tree Algorithm from all of the Job sites, means recruiter would not go through same work for all the Job Sites to post Job details. In this way, the system provides accurate result for all types of requirements provided by Recruiter.

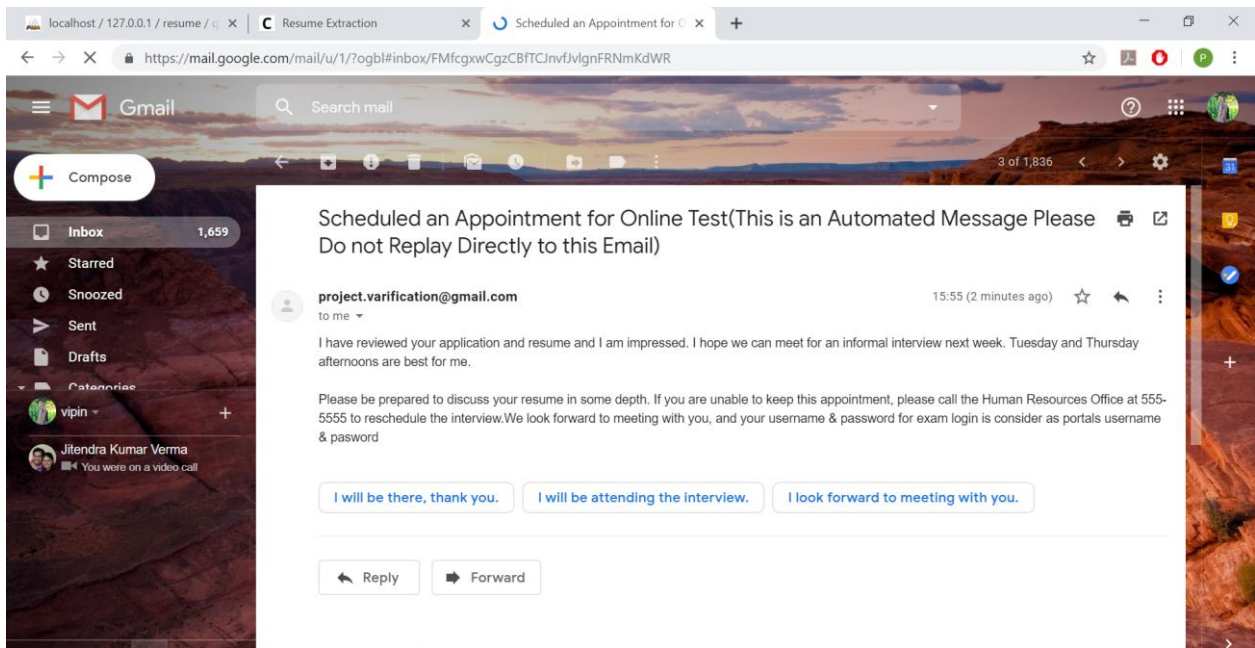


Figure : Confirmation Mail for Online Test

This is the confirmation mail for online test to user. Therefore, user can appeared for this online test & good performance in the online test leads to the Interview.

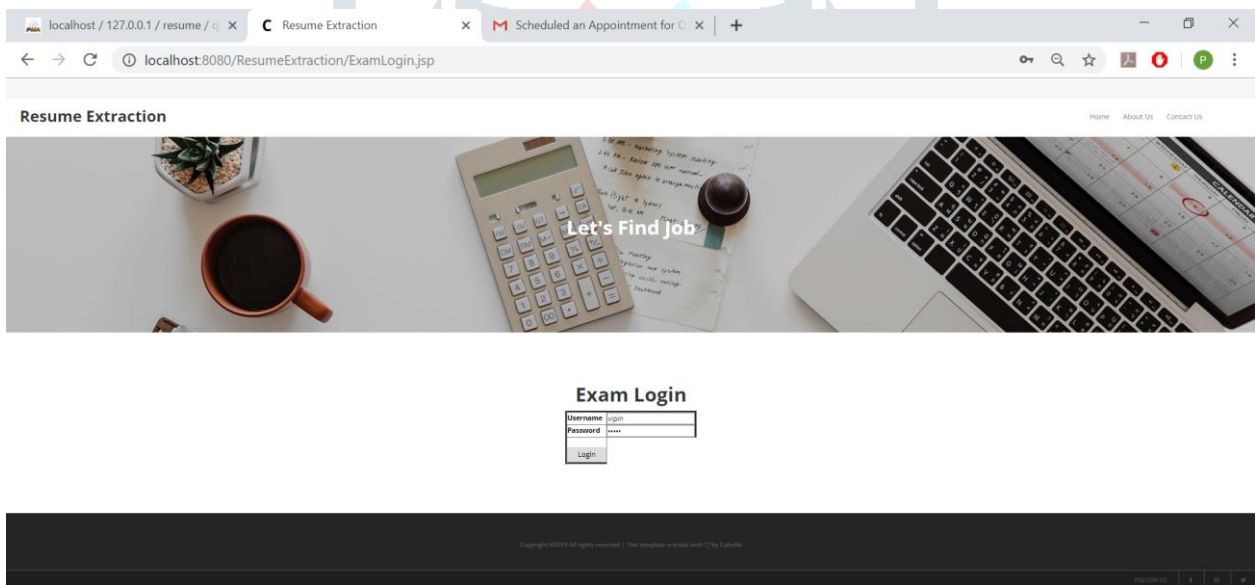


Figure : Examination Section Login

Only candidates who are selected(Recruiter send confirmation mail to everyone whose resume are extracted) by recruiter can eligible for this test otherwise user can't login.

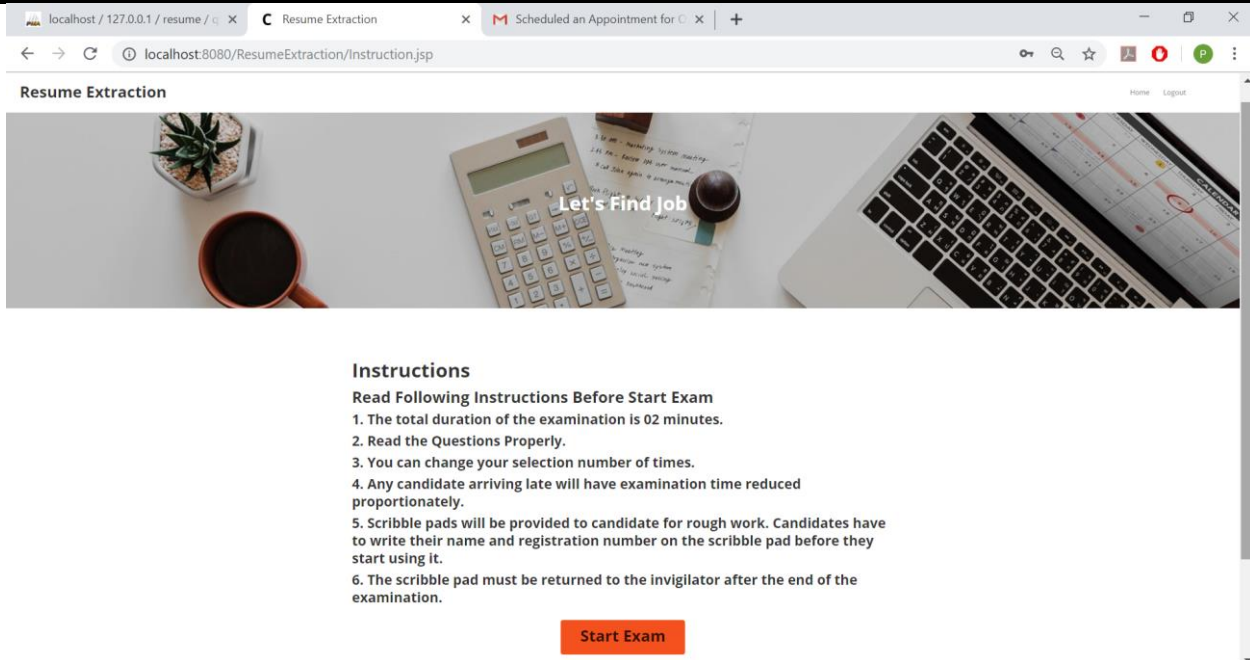


Figure : Test Instructions

These are some basic instructions which should followed by every users.

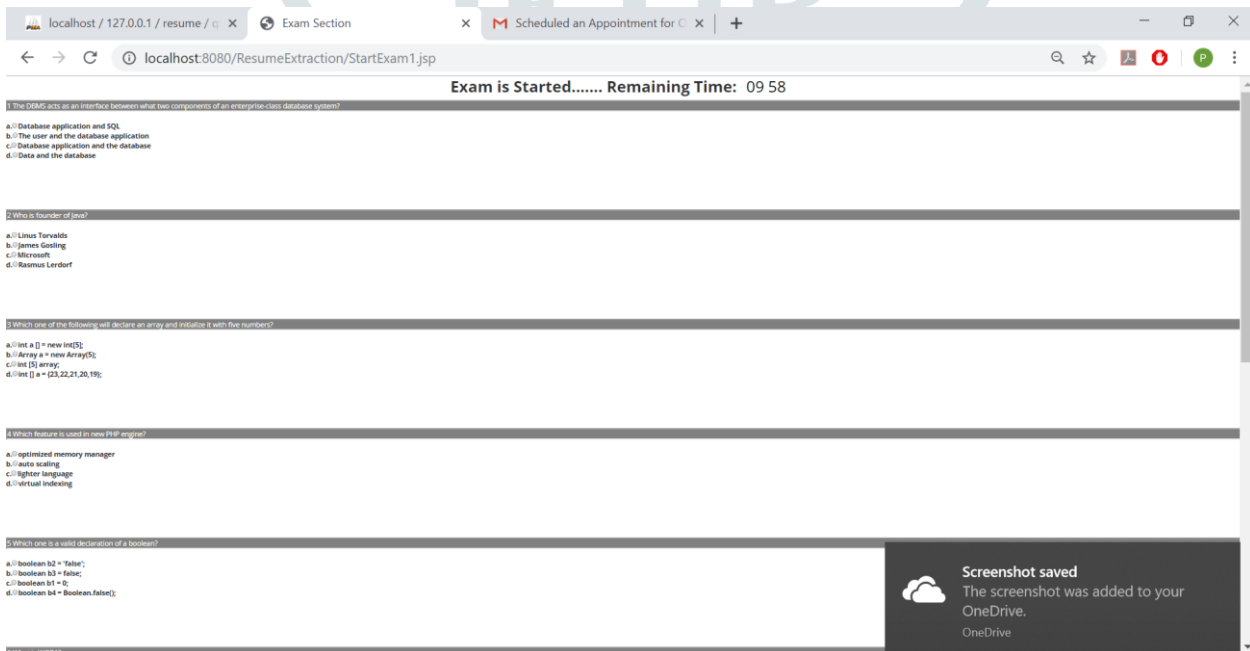


Figure : Test Started

This is how the page look like when candidate is going onto online test.

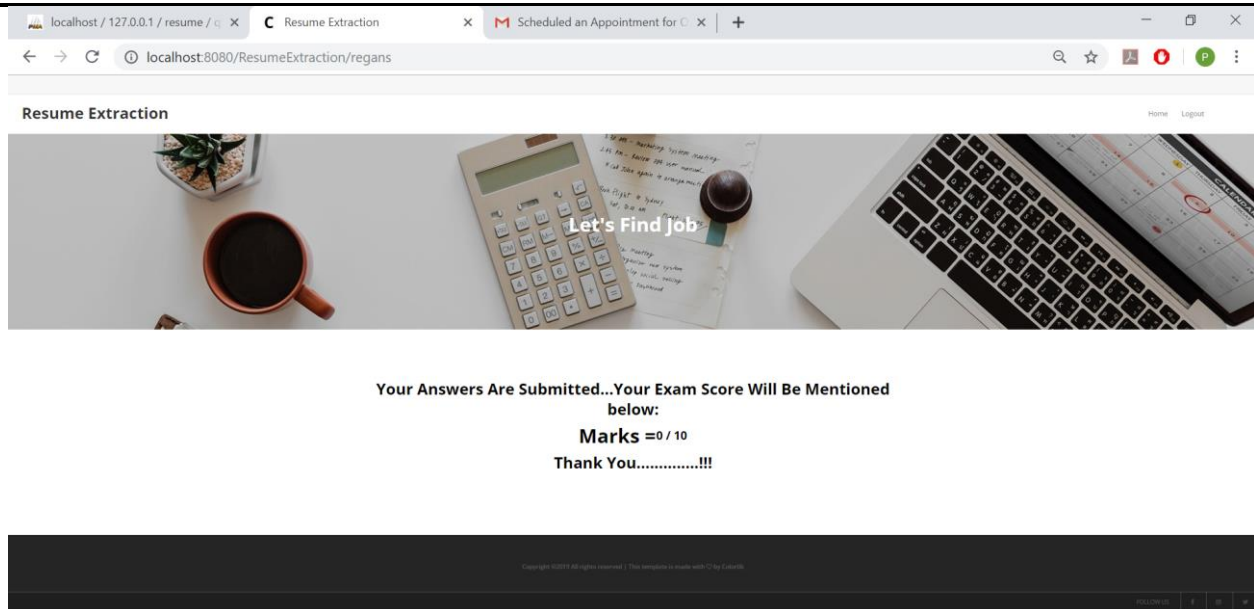


Figure : Test Result Page

When time is over or candidate submit the test, result is generated as per candidate performance in the test.

If he/she crosses the threshold marks then a message says “Congratulation you are successfully passed this online test. You are selected for Interview test. Please come at this place at this timing.”

If he/she not successfully passed the test then a message says “Better Luck Next Time”.

In this way we build this system to reduced the time of manual checking of resumes, short list the number of candidates with the help of online test.

So, large amount of time is save with the help of this system.

V. CONCLUSION

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.

ACKNOWLEDGEMENT

This research was supported by Marathwada Mitra Mandal’s Institute of Technology. We express our gratitude to our colleagues who provided insight that greatly assisted the research and made our work easier, although they may not agree with all of the interpretations provided in this paper.

We are also grateful to Ms. Jagruti Bhagwat for assistance with Resume Extraction and Candidate Recruitment System, and Prof. Sneha T. Shinde, who moderated this paper and in that line improved the manuscript significantly. We have to express our appreciation to everyone who helped us, for sharing their pearls of wisdom with us during the course of this research. We are immensely grateful for their comments on earlier versions of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed professionals.

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] Gláucia 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] https://en.wikipedia.org/wiki/Decision_tree
- [10] http://en.wikipedia.org/wiki/Naive_Bayes_classifier

