

Candidate Ranking and Evaluation System based on Digital Footprints

Vinay Dandwani, Vinita Wadhvani, Raveena Chawla, Nitin Sachdev,
Mrs. Arthi C.I

Department of Computer Science & Engineering Vivekananda Education Society's Institution of Technology
Mumbai, India

Abstract: Digital resume provides insights about a candidate to the organization. This paper proposes a system where digital resumes of candidates are generated by extracting data from social networking sites like Facebook, Twitter and LinkedIn. Data which is relevant to recruitment is obtained from unstructured data using Data Mining algorithms. Candidates are evaluated based on their digital resumes and ranked accordingly. Ranking is done based on the requirements specified by an organization for a key position. The key aspects of this paper are a) Specification and design of system. b) Generation of digital Resume. c) Ranking of candidates. According to the ranking provided by this system, Recruiters can shortlist candidates for interviews. Thus, it revolutionizes the traditional recruitment process.

Keywords: Recruitment, Digital Resume, Digital Profile, Digital Footprint, Candidate Ranking.

I. Introduction

A digital footprint is created by an entity's interaction in a digital environment. These digital footprints are valuable in targeted marketing, sentiment analysis and other social graphing services. Google Ads uses our digital footprint to show relevant advertisements. Such targeted advertisements are beneficial for both, the users and the companies. Facebook and YouTube also uses our digital footprints to give us better experience by showing relevant information [1]. This paper states that these digital footprints can also be used to revolutionise the traditional recruitment approach by designing a Candidate Ranking And Evaluation System. This system will rank each candidate on the basis of their digital resume which is created from their digital footprints. Social footprints like endorsements, achievements, projects, skill sets of a candidate will be extracted from Social websites like LinkedIn, Facebook and Twitter. The recruiters can shortlist the applicants for interview based on the ranking provided by this system.

Moreover, using the Twitter streaming API (Application Programmable Interfaces) a system can be built to recruit candidates. The authors of paper [2] created a system that recruits candidates for smoking cessation nicotine patch study using various twitter's features like tweets, retweets, hashtag etc. The authors designed a data stream processing software system that intercepts an incoming tweet in real-time to recruit users.

II. Proposed System

This section describes the detailed architecture of the proposed system. System Architecture is divided into three modules. 2.1) Extracting data from LinkedIn, facebook and twitter. 2.2) Data Mining and storing refined data in database. 2.3) Generating the digital resume based on the refined data and ranking the candidates. The first module involves deep digging into these social networking sites to find the relevant attributes for recruitment. Priorities are assigned to these attributes based on their importance in evaluating candidates. For example, To recruit a Java Developer, projects and achievements for java are more important than comments related to java. Using social media mining, data is extracted from candidate's LinkedIn profile, twitter account and facebook account. Social Media Mining is the process of representing, analyzing and extracting actionable patterns from social media [3]. This raw data is stored in the data warehouse for further processing. The second module is designed to refine the data. It gives a structure to the data so that digital resume can be generated. This digital resume consists of candidate's technical skill set, Projects, Achievements etc. The most important step of this architecture is ranking the candidates based on company's requirements so that best fit can be determined. For Ranking, digital resume would be rated. Rating would be done by mapping the digital resume and the company's requirements. While rating the profile, Priorities of the attributes are considered. Thus, Recruiters get direct access to the digital profiles of the applicants along with their ranks. This will help them to shortlist candidates for interviews. And it also serves as an aid for pre-interview verification as the recruiter gets a fair idea about the candidate prior to the interview.

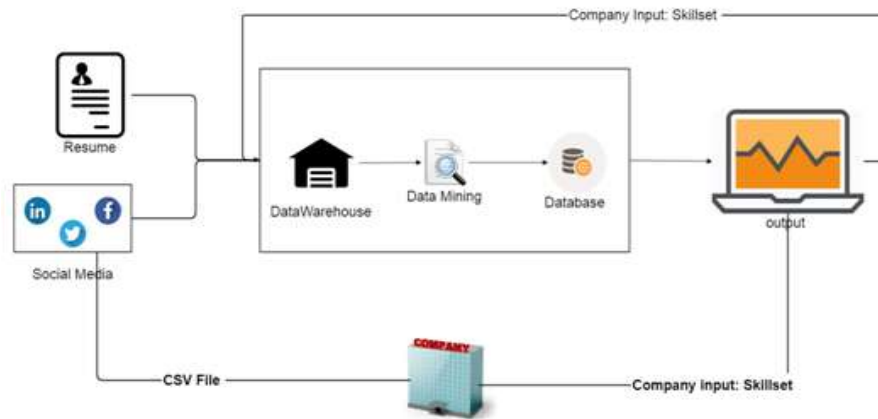


Figure1: Proposed System

III. Implementation Details

There are 2 parts of this system.

3.1) When Recruiter search for candidate in our system.

3.2) When Company requires candidate ranking (List of candidates given by the company). Input:

1. Excel file which contains the list of all candidate's details along with their account links of Gmail ID, LinkedIn, Twitter and Facebook .
2. The requirements in terms of skill sets / key position from the organization itself.

Output: The dashboard, where it presents candidates digital resume with respect to footprints collected.
Processing: Processing is comprised of 3 sub-parts: 3.2.1) Data extraction. 3.2.2) Data Mining 3.2.3) Candidate Ranking.

3.2.1) Data Extraction: Data Extraction involves Creation of application on their developer website which grants us the permission to access their data

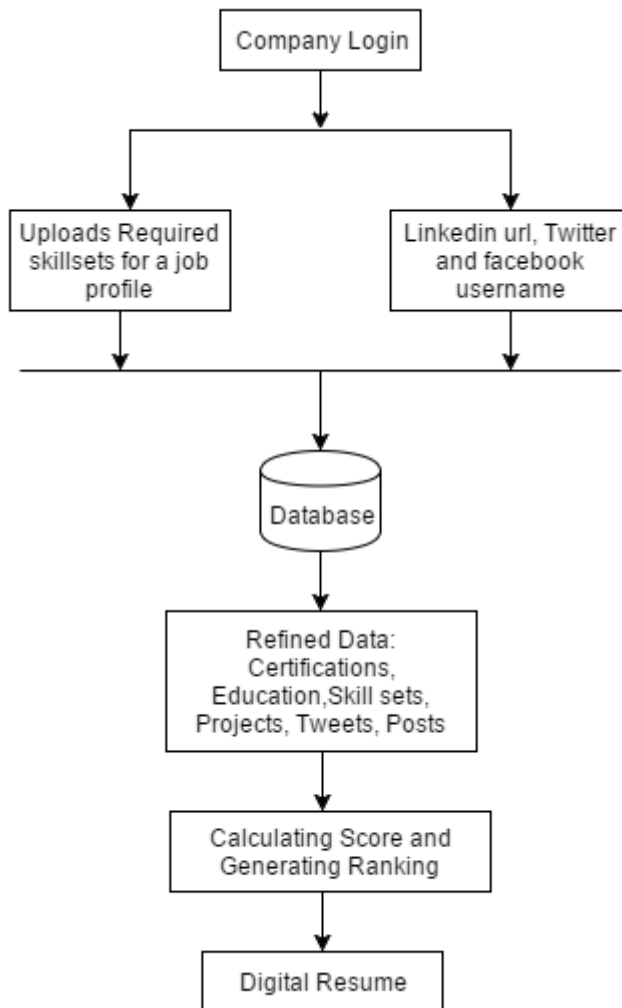
3.2.1.1) LinkedIn: LinkedIn is a professional networking site. Recruiters often check LinkedIn profiles of applicants for quick background verification. Due to this, the credibility of data shared by users on LinkedIn is higher than other networking sites [4]. Users share their academic information with minute details. LinkedIn data also includes Projects, Achievements and Certifications Python has many packages available to access LinkedIn like PyLinkedInAPI 0.1.6, linkedin 0.1.5, linkedin-scraping 0.1.2, parseli 0.0.5. These API's provide functionalities like getProfile, postComment, postMessage etc

3.2.1.2) Twitter: Tweet is a short way to convey the message to the world. So people tweet their achievements, which programming languages they know and what projects they have done. Packages in python which let us extract tweets are python-twitter 3.2 and tweepy.

- E.g.: tweet: "java developer since 2 years." This is considered as designation & cross verified with LinkedIn positions specified. If verified then, candidates current key-position/experience is stored as Java Developer.
- E.g. tweet: "java sucks" This is also considered (score is deducted from java field).

3.2.1.3) Facebook: Facebook is a social networking site to connect with people in an informal way. It is not as professional as LinkedIn but we can obtain some relevant information from a facebook profile. People share their education details on facebook. Users express their views through comments. Users also post about the events they have attended recently. All this information about a candidate reveals candidate's skills, interests etc. This data can be extracted using facebook's Graph API. Facebook Graph API gives a JSON file which can be queried to get relevant data. There are various packages in python which are used to get data from facebook's

Graph API. They are Json, urllib etc. For database connectivity, pymysql can be used.



3.2.2) Text Refinement: The gathered raw data is initially stored in data warehouse, henceforth refinement of data take place w.r.t keyword search based techniques.

3.2.3) Ranking System: To distinguish candidate's talent from one another, it focuses on various factors like tweets, projects, certifications, posts, comments, retweet etc. To make the system fair and unbiased certain points can be allocated to those factors. Points may differ based on requirement of an application. With reference to these points and the data available from social media, Algorithm will generate scorecard of candidates which will help in ranking of candidate. Based on scorecard, system will rate candidate i.e Number of stars will define the efficiency of a candidate.

To make the system more efficient, ranking will not only be judged on the basis of digital footprints and projects but also on the basis of the review provided by the interviewer. Interviewer will be asked to rate the candidate after conducting the interview. Interviewer can give feedback in the form of comments or he can also endorse candidate for a particular skill after the interview. From the interviewer's review, system would change the ratings given earlier. This makes the system more dynamic.

Thus, Interviewer's review regarding candidate's performance gives the system a more accurate judgement regarding candidate's skills and knowledge. It helps the system for future analysis of candidate. Considering all the factors candidate will be ranked and stars will be rewarded. Interviewer's review can be used as means to determine the authentication of candidate's data which he/she has submitted i.e from the interviewer's review it can be determined whether the data submitted by the candidate describing the candidate's skills, knowledge, projects, etc are genuine or not. This will help the system in future analysis of candidate. If the data provided by candidate is not genuine then it will affect the rank of candidate. This solution may solve one of the main issue of this system which is data authentication. The above solution eliminates the overhead of authenticating the data submitted by candidate describing the candidate's skills and knowledge.

IV. Conclusion

This paper presents a modern approach of recruitment. This approach results in providing ease to the recruiters. This system proves to be beneficial for large scale recruitments as screening of large number of resumes is difficult and time consuming. This system enables the recruiters to shortlist candidates for interviews. Moreover, proposed system also serves as an aid for pre-interview verification. This approach can be further extended to discover the passive applicants of a job. passive applicants are the applicants who have not applied for a job directly but are unhappy with their current job. Students who have the potential to do internships also come in this category. Thus, by extracting the data of such social networking sites, passive applicants can be identified. The efficiency of the algorithm can be increased by extracting information of candidates from more promising websites like Stack overflow, Hackerrank etc. such websites can provide more insights about the technical skills of a candidate.

References

- [1]. Digital Footprints: Pros, Cons and Future by Prof. Anjana R. Arakerimath, Pramod Kumar Gupta MCA Department, Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University Pune, Maharashtra 411044, India.
- [2]. A Twitter-based smoking cessation recruitment system by Ahmed Abdeen Hamed, Xindong Wui and James R Fingar.
- [3]. Social Media Mining, an ebook by Reza Zafarani, Mohammad Ali Abbasi, Huan Liu.
- [4]. A framework to identify educational relevance in social networking posts by Muhammad Fahim Uddin, University of Bridgeport.
- [5]. Recruiting with Social Media: Social Media's Impact on Recruitment and HR, an ebook by Raj Anand.