

Applicant Tracking and Scoring System

Ankit Tiwari¹, Sagar Vaghela², Rahil Nagar³, Mrunali Desai⁴

¹Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T., Sion, Mumbai, Maharashtra, India

²Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T., Sion, Mumbai, Maharashtra, India

³Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T., Sion, Mumbai, Maharashtra, India

⁴Assistant Professor, Department of Computer Engineering, K.J. Somaiya Institute of Engineering & I.T., Sion, Mumbai, Maharashtra, India

Abstract - Text in a resume file stored in any format is cleaned, extracted and fed to a Natural Language Processing (NLP) module. The NLP module is fine-tuned with the help of Named Entity Recognition (NER) to accommodate the varying needs of different companies and their technical jargons. The extracted, intelligent data is stored in a ".csv" format. The Human Resource Officer (HR) can define constraints on which the HR wants the applicants to be judged. These constraints are used to score each traits from the resume. At last the cumulative score is used to showcase the eligibility of the candidate to the HR.

Key Words: NLP, NER, Resume Parser, text mining, JSON Resume.

1. PROBLEM STATEMENT

To design a system to extract the information, parse the extracted information from unstructured to structured JSON format, and rank those resumes according to the skill sets of the candidate and on the job description.

2. INTRODUCTION

The biggest challenge of screening resumes by far is its volume. The number of resumes received is one of the biggest time-consuming factors during the recruiting process. An average job opening receives 250 resumes and up to 88 per cent of them are considered unqualified. This means a recruiter can spend up to 23 hours screening resumes just for a single hire. An ATSS (Applicant Tracking and Scoring System) is a must-have software for recruitment and talent asset departments because it organizes all the resumes received for each role. With the growing competition on a daily basis and the growing skills of the applicants, it gets difficult for the companies to sort the resumes according to their needs and this activity is very time-consuming. Recruitment is an important activity in any organization or a company. Therefore, instead of going through each of the applicants, the company can get a sorted list of applicants that fulfil the company requisites. ATSS is an application that caters the company for sorting the resumes and helping the HR team to shortlist the applicants based on various criteria.

Applicant Tracking and Scoring System tracks job applicants as they go through each stage of the hiring process. Many ATSS systems offer additional features, such as note taking, bulk emailing and job posting.

3. LITERATURE SURVEY

3.1 Shakya, Sujan, Web-based Employment Application Processing Support System, Master of Software Engineering, May 2008, (Dr Thomas Gendreau, Dr Kasi Periyasamy)

It is a powerful online recruitment and application processing support system which is capable of storing and maintaining different types of user accounts, resumes, applications, jobs, and keeping track of the steps in the hiring process. It allows applicants to search for jobs based on different criteria and to post application. In addition, WEAPSS allows the applicant to create online resumes which can be posted for multiple jobs. In the two years, following the first approval for recruitment for a position, WEAPSS allows HR staff to re-initiate the pool search and re-advertise the position for a number of times in case of a vacancy. The focus of WEAPSS is to streamline advertisement, hiring processes, save administrative time, eliminate redundant processes, and accelerate communication with candidates. Throughout each phase of the recruitment process, WEAPSS facilitates a much more streamlined, standardized approach than the existing, paper-based recruitment process. Tasks such as sorting, coding, filing, and routing application materials which were previously performed manually, can now be performed automatically.

3.2 Bazzi, Issam, Richard Schwartz and John Makhoul (1999) An omnifont open-vocabulary OCR system for English and Arabic. Pattern Analysis and Machine Intelligence 21 495-504

The early experimental OCR systems were often rule-based by the '80s these have been completely replaced by systems which are based on statistical pattern recognition. For clearly segmented printed materials such techniques offer us

virtually error-free OCR for the most important alphabetic systems including variants of the Latin, Greek, Cyrillic, Hebrew alphabets and many more. However, when the number of symbols is large, as in the Chinese or say Korean writing systems or the symbols are just not separated from one another, as in Devanagari print, OCR systems are still far from the error rates of human readers, and the gap between the two is also very much evident when the quality of the image is compromised for example by fax transmission. Until these problems are resolved, OCR cannot play a pivotal role in the transmission of cultural heritage to the digital age that it is often assumed to have. In the recognition of handprints, algorithms with successive segmentation, classification, and identification stages are still said to be in the lead. For cursive handwriting, Hidden Markov Models that make the segmentation, classification, and identification decisions in parallel have proven superior and topmost, but performance still leaves much to be desired, both because the spatial and temporal aspects of the written signal are not necessarily in lockstep (discontinuous constituents arising, for example, at the crossing of t-s and dotting of i-s) and because the inherent variability of handwriting is far greater than that of the speech, to the extent that we often see illegible handwriting but we barely hear unintelligible speech.

3.3 Santosh Kumar Nanda, Department of Computer Science and Engineering, Eastern Academy of Science and Technology, Development of Intelligence Process Tracking System for Job Seekers

At the present time to getting a good job is a very intricate task for any job seekers. The same problem also a company can face acquiring intelligent and qualified employees. Therefore, to reduce the problem of manually searching for the right candidate, there are many management systems were applied and out of them, the computer-based management system is one of an appropriate solution for this problem. In the computer management system, software is made for job-seekers to find their suitable companies and as well as made for companies finding their suitable employees. However, the current solutions in the market are not Artificial Intelligence (AI) based, and to make privacy, security and robustness, the solution should be made with the application of an AI system. In this proposed study, an attempt has been made for finding the solution for job seekers and companies with the application of expert systems.

4. PROPOSED SYSTEM

4.1 Text Extraction

Many of the resumes that are stacked inside a company is already in a soft copy format, on which a software extraction technique can directly be applied. But to make this system

more holistic, an Optical Character Recognition (also optical character reader, OCR) module is added to include the applicants who have sent a hard copy of their resume. OCR is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image[6]. For our project, we will be OCR to convert the digital formats like PDF, JPG etc. to a text format. Fig - 1 shows the process of converting a hard copy document into a text blob[1] that can be fed into a Natural Language Processing (NLP) module.

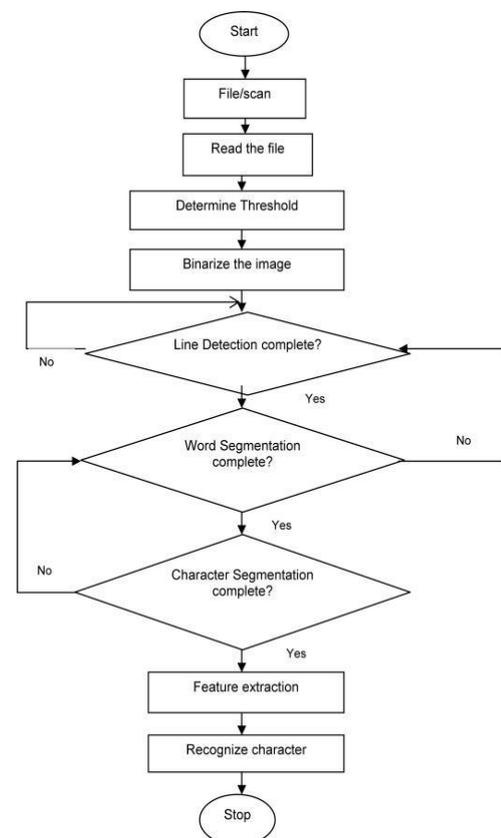


Fig -1: OCR Flow Chart

4.2 Natural Language Processing (NLP)

Natural Language Processing is an algorithm that takes the plain text as input and can convert into meaningful data. Using NLP, we are going to parse the resume, NLP requires the following for parsing:

- 1) Lexical Analysis
- 2) Syntactic Analysis
- 3) Semantic Analysis
- 4) Named Entity Recognition (NER)

Lexical Analysis:

Lexical analysis is the first phase of NLP parsing, as shown in Fig -2, the plain text input is segmented into words and paragraphs and then the tokens are created.

Syntactic Analysis:

In Syntactic analysis the analysis of the grammar and the arrangement of words in a meaningful manner is checked, sentences like "College goes to girl" is rejected.

Semantic Analysis:

Semantic analysis checks the exact meaning of the text, sentences like "Sunny is raining" will be rejected by the English semantic analysis.

Named Entity Recognition (NER):

One of the problems with using the same NLP module for all the companies is the jargons and words that mean something for that company's domain and may mean something else in general. This hindrance is overcome in our system with the help of "Named Entity Recognition" or NER. A named entity is an object that exists in the real world. With NER, we can fine tune our NLP module to understand the real world objects from a domain[2][4][5]. For example, if a company wants to hire developers, they can use our system to differentiate between people who love "Python"- the programming language and the people who love "python"- the snake, based on the context in which the word is used.

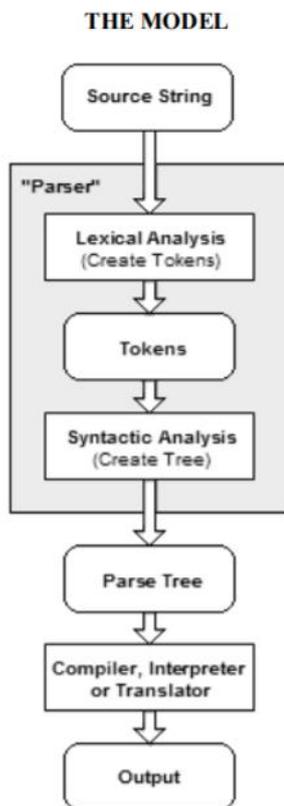


Fig -2: NLP Model

4.3 Applicant sorting and Dashboard interface

A web portal is provided to the HR, to define the constraints and the required skill sets of the company, on which the applicants are to be judged. After getting the output from the NLP, the data obtained will be used for a dashboard which will contain graphs and pie charts based on the data in the resumes. An HR can use this dashboard to prepare his/her query based on the requirement that the respective company has. In order to make this dashboard, all the data will be fed into ElasticSearch. ElasticSearch provides powerful tools to prepare such dashboards. ElasticSearch is included in order to make a more holistic system, which will be ready for basic level industry use. To make the system indifferent to many formats, the data to be fed will be fed by using LogStash.

Using queries inbuilt in ElasticSearch, the resumes will be scored and then they will be sorted according to the constraints that were provided by the HR. The individual traits of an applicant will be provided with a proportionate boost based on the priority of the trait. At the end, the cumulative score will be used to sort the applicants. A final sorted list of applicants will be displayed to the HR (Fig -3).

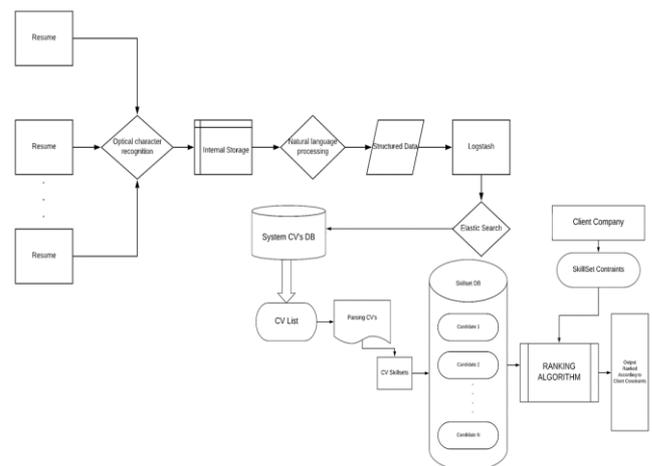


Fig -3: Proposed System

5. EXPECTED OUTCOME

Both employers and candidates will be benefited by our system, our system will parse all the applicants resume and store the summary by extracting key field in the database (Fig -4). The system will use this database to display information about the applicants using pie charts and bar graph (Fig -5) Then it will rank them (Fig -7) according to the constraints defined by the HR (Fig -6), thus reducing the unfair hiring practices and making the hiring system authentic.

[3] M. J. Cafarella and O. Etzioni, "A search engine for natural language applications," in WWW, pp. 442–452, 2005.

[4] 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>.

[5] Sovren Resume/CV Parser, <http://www.sovren.com>

[6] http://vlabs.iitb.ac.in/vlabs-dev/labs_local/machine_learning/labs/exp11/theory.php