

Resume Analyzer Using Text Processing

B.Kelkar¹, R.Shedbale², D.Khade³, P.Pol⁴, A.Damame⁵

¹ Professor, Dept. of Computer Science and Engineering, Sanjay Ghodawat College, Shivaji University, India

² Student, Dept. of Computer Science and Engineering, Sanjay Ghodawat College, Shivaji University, India

³ Student, Dept. of Computer Science and Engineering, Sanjay Ghodawat College, Shivaji University, India

⁴ Student, Dept. of Computer Science and Engineering, Sanjay Ghodawat College, Shivaji University, India

⁵ Student, Dept. of Computer Science and Engineering, Sanjay Ghodawat college, Shivaji University, India

Abstract: An abstract summarizes recruitment of candidates becomes more challenging when lots of resumes are uploaded for the same employment position. Companies many a times not able to decide on which resume is best-fit for that position. This paper presents an effective Company Recommender System to help recruiters to find the best candidate for given employment title using text mining and machine learning tools. When candidates upload their resumes, then according to the companies requirements for the employment positions the candidates' resumes get ranking. The ranking can be used by the company to get most preferable candidates.

Key Words: Text Mining, Lexical Analysis, Natural Language Processing, etc.

1. INTRODUCTION

Text Mining is the discovery of new and previously unknown information, by automatically extracting evidence from different written resources. A key element is the connecting together of the extracted material together to form novel truths or new hypotheses to be discovered further by more conventional means of experimentation. Text mining is dissimilar from what we're familiar with in web search. In web search, the user is typically watching for something that is previously known and has been written by someone else. The goal is to keep aside all irrelevant to the query needs in order to find the relevant information.

Text mining is a deviation on a field called data mining. Data mining tries to find interesting patterns from large databases, whereas in text mining, the goal is to discover unknown

information, something that no one yet knows and so could not have yet written down. Databases are designed for programs to process mechanically; text is written for individuals to read. A "real" text mining is that discovers new pieces of knowledge, from approaches that find overall trends in textual data. There are programs that can, with sensible accuracy, extract evidence from text with somewhat regularized structure. For example, programs that read in resumes and extract out people's names, addresses, job skills, and so on, can get accuracies in the high 80 percent. This process rather falls into an area called information extraction.

With the rapid growth of Internet-based hiring, there are a abundant number of personal resumes among recruiting systems. To increase more consideration from the recruiters, most resumes are written in diverse formats, including changeable font size, font color, and table cells. However, the diversity of format is harmful to data mining, such as resume information extraction, automatic employment matching, and candidates ranking. Supervised methods and rule-based methods have been expected to extract truth from resumes, but they great depend on categorized structure information and huge amounts of labelled data, which are rigid to gather in reality. The mission of finding the right nominee for a particular job can be a very exhausting 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. This may result in short listing of wrong candidate or rejection of a right candidate which may result in significant loss of money and other resources. Sentiment Analysis

approach can also prove vital to analyze a candidate's resume on the source of the description he or she delivers. To shorten this procedure, we recommend a Text Analytic approach to judge resumes on the basis of their content.

Resumes are an abundant source of unstructured data which can be usefully analyzed by the companies to shortlist the right candidate. Numerous talents of a candidate can be recognized based on the content of his resume. Just like humans, a computer can analyze the resume by finding the accurate keywords which will classify the level of each nominee on a scale of 3, Low, Average and High. The Internet-based enrolling platforms play an important role in the recruitment channel [1] with the rapid growth of the Internet. Nowadays, almost every company or department posts its job requirements on various online hiring platforms. There are more than one thousand employment requirements uploaded per minute in Monster.com (<http://www.monster.com/>). Online hiring is immensely useful for saving time to both employers and employees. It allows the job seekers to submit their resumes to many operatives at the same time without travelling to the office and it also saves operatives' time to consolidate a job fair. Meanwhile, there are countless portals acting as a third-party facility amid job seekers and company human resources, so that tons of resumes are collected by these portals. For instance, LinkedIn.com (<http://www.linkedin.com>) has collected more than 300 million personal resumes uploaded by users. Because of the growing quantity of data, how to effectively analyze all resume is a severe problem, which fascinated the attention of researchers.

The paper is separated in following sections – Section 2 terms the associated work which has been done. Section 3 will present the full explanation of the methodology we recommend to approve. Section 4 shall give an vision of the prototype we have executed till now. Section 5 shall present the conclusion and the offered upcoming work which can be done.

2. RELATED WORK

There are three main extraction methods to deal with resumes in previous research, including keyword search based method, rule-based method, and semantic-based method. Since the fine points

of resume are tough to extract, it is an different mode to accomplish the goal of job matching with keywords search approach. Enthused by the way of extracting the news web page, numerous rule-based extraction methods give the resume text as a web page and then extract detailed truths based on the DOM tree structure. For the past kind of methods, investigators treat the resume extracting task as a semantic-based object extraction problem. Some investigators use sequence labelling process or text classification methods to predict the tags for segments of each line. Though, most of these methods intensely rely on hierarchical structure information in resume text and large amounts of labelled data. In actuality, learning of text extraction models frequently relies on data that are labelled/ annotated by a human expert. Moreover, the more expertise and time the libeling process requires, the more costly it is to label the data.

Resume RDF ontology has been presented by Uldis Bojars and John G. Breslin [3] which customs an RDF data model to model a resume. Resume RDF describes resume information with its profligate set of classes and properties. Uldis Bojars further prolonged FOAF with resume information [4] for an even more upgraded description of information. In 2002 and 2003, Turney and Littman offered a scheme which would infer the semantic orientation or evaluative character of a word from its huge hundred billion-word corpus corpora taking into thought the semantic associations with the other words, stated as paradigms by him. [5][6].

Ujjal Marjit et al. [7] proposed a dissimilar method which recovered resume information using the idea of Linked Data allowing the web to share data with dissimilar sources allowing it to discover several kinds of information. An ontology based approach was proposed by Maryam Fazel- Zarandi et al. [8] which would match job seekers talents with the help of a deductive model which determined a match between the talents of a job seeker and the skills required by the recruiter. One more system to mechanize resume information extraction was established by Kopparapu of TCS Innovations lab [9] which performed speedy search of resume extracting valuable information from a free format resume with the help of numerous NLP methods.

Online Chine resume parser was offered by Zhi Xiang Jing et al. [10] which used rule based and statistical algorithms to extract data from a resume. Zhang Chuang et al. [11] worked on a resume document block analysis which was based on

pattern matching and multi-level information identification building the largest resume parser system. elik et al. [12] intended a system which transformed a resume into an ontological structural model which simplified the analysis of Turkish and English resumes. Di Wu et al. [13] succeeded to extract information from resumes more excellently by the idea of ontology using WordNet for similarity calculation. Numerous research articles were found on information extraction from resumes Yu et al. offered a resume information extraction methodology with cascaded hybrid model [9]. Chandola et al. established an online resume parsing system using text analytics [10]. Kopparapu et al. presented a method for automatic extraction of valuable information from unstructured resume and claimed great precision and recall [11]. Several other efforts for extracting the information from the resumes have been made by Zhi Xiang et al. [12], Zhang Chuang et al. [13], Celik Duygu et al. [14]. Mayuri Verma et al. proposed a Cluster based Ranking Index (CBR) to rank the resumes to find the appropriate nominee and thus enhance the employment process [15].

Very few efforts have been made in the past to make a Company Recommender System. YixinCai et al. executed a feature-based multinomial sorting system as a Company Recommender [16]. Then again numerous research articles have been found on Recommender Systems in other areas or fields. Douglas Eck, Thierry Bertin-Mahieux and Paul Lamere. [17] developed music auto-tagging system using meta-learning algorithm. Claudio Biancalana[18] planned a method to recommend movie via neural networks. Yukun Cao and YunfengLi[19] planned a fuzzy based system for consumer electronics to regain optimal products. Lihong Li, Wei Chu, John Langford and Robert E.Schapire[20] developed a personalized news article approval system by developing contextual information about the users and articles. Rafael Sotelo, Yolanda Blanco-Fernandez, Martín López-Nores, Alberto Gil-Solla and José J. Pazos-Arias [21] proposed a TV program recommendation methodology, which formed personalized TV schedules for crowds of users based on TV-Anytime descriptions of TV contents and semantic reasoning techniques. Text Analytics also stated to as Text Mining is a quickly rising field which can be defined as a statistical machine learning methodology for translating the unstructured raw information into structured information which can

further be categorized, mined, trained for well and great feature information [22]. Techniques of Pattern Recognition, Information Extraction, Data Mining, Parsing are involved in Text Mining and denoted to as Natural Language Processing (NLP) [23]. Rest of the paper has been organized in the order of the steps used in this work. The figure below shows the model proposed in this paper.

Even if there are numerous other existing websites which offer advanced services like searching on the basis of keywords, domain, location etc., their search does not take into consideration, the skill level of a specific candidate. For example, if a company searches for a applicant who can work in C language, they can easily search for applicants who have C language mentioned in their resumes. But how will they know the expertise of that specific applicant in C language.

In our pattern, we use additional information like the projects in which the candidate was involved as well as the project description. This information will be taken as input from the candidate and by analyzing the text used by him or her; we will categorize the candidate into various expertise levels. So if a company wishes an employee who necessarily has a great proficiency level in C language, simply then his resume will be eligible. A knowledge base of numerous keywords will be planned which will form the basis of categorization. This will also help in ranking of numerous resumes to convey which one is better or worse than the other giving the applying nominees a chance to present themselves in the finest probable way.

3. METHODOLOGY

To discover appropriate literature on e-recruiting and data mining from resumes, we shortened the techniques of prior research and wisely selected the articles that are most relevant to our research. According to the approved structures, there are three kinds of famous approaches about resume information extraction in earlier research, which can be termed as follows. The first group of methods takes keywords recovery in consideration. In [3, 5], only the particular data are designated to filter resume streams. Both of them ambition to accelerate the efficiency of search applicants for the job. Some of the vital queries were generated to

filter the resume set so that they can help to improve the work efficiency of the team. Even if these varieties of methods are easy to implement, the raw text content gets too many noises into the index, leading to low accuracy and inadequate ranking results.

The second set of techniques based on the DOM (Document Object Model) tree structure, in which tags are internal nodes and the full text, hyperlink, or images, are leaf nodes. Ji et al. [19] offered a tag tree algorithm, in which they identified and detached the shared part among web pages with the similar template, and then the main text is retained. Also certain other methods extract the information with Regex rules from the HTML pages. Jsoup (<http://jsoup.org>) and Apache POI (<http://poi.apache.org>) can be used to parse resumes that follow some specific template file. Jsoup is a Java library for in work with real-world HTML. It delivers a actual suitable application interface for extracting and manipulating data based on the DOM structure. Furthermore, POI is a valuable Java library for working with Office file, attentive on extracting the file content. It is easy to build a particular program to extract the information from those resumes which follow the particular template file. In [20], the system executed the extraction by annotating texts using XML tags to identify elements such as name, street, city, region, and email. These techniques based on template file with DOM tree are limited by human hard work. Meanwhile it is impossible to know how many groups of resumes follow the same template; these techniques are tough to scale out in big data.

The third group of methods gives extracting information as a semantic-based object extraction job. In [17], a cascaded information extraction framework was planned to support automatic resume organization and routing. The first pass is used to fragment the resume into successive blocks with labels showing the information varieties. Then full information, such as Name and Address, are recognized in certain blocks without searching globally in the entire resume. In [16], a system that aids in the shortlisting of candidates for jobs was designed. Their scheme integrates table analyzer, CRF predictor, and content recognizer into the entire part of parsing resumes. The arrangement of table cells in the file was considered by the table analyzer, and the CRF predictor was used to predict

the label of the text order; then the content recognizer was used to mine named objects in the nominee resume text. In [14], they recommended an ontology driven data parsing system that was intended to operate on lots of resumes to change their structured format for the purpose of professional finding through the semantic web method.

In [15], investigators presented EXPERT, an intelligent tool for screening nominees for employment using ontology mapping. EXPERT has three stages in screening collects applicants' resumes and constructs ontology document for the features of the applicants. Job requirements are denoted as ontology in the second stage. And in the third stage, EXPERT maps the job requirement ontology onto the applicant ontology document and retrieves the qualified nominees. Tang et al. [21] also hire CRF as the tagging model. DOM tree structure is used to infer the characterised structure; then content features, features, and term features are joined to train the model. Uldis Bojars introduced ResumeRDF (<http://rdfs.org/resume-rdf>) ontology to model resume. Further, he prolonged FOAF (<http://xmlns.com/foaf/spec>) to support more description of resume. Chen et al. [18] offered a framework based on text classifiers, which are trained with data corpus from the Internet as an different of manual annotation. Though, these works are limited by file formats and the huge human efforts, which cost in labeling the arrangement data for ontology, CRF, or semantic web model.

The model we propose has four steps:

- Collection of resumes.
- Searching for keywords stored in information base in the resume text.
- Fetching new keywords from the resumes to construct the knowledge base further.
- Ranking and Categorization of candidate based on a rating score.

3.1 RESUME COLLECTION

The purpose of this research is to compare the performance of the system using machine learning algorithms that will provide with a short list of recommended resumes from the large number of resumes that will be given as input. However, any of the machine learning systems make their decisions based on some string assumption, something they discern from the well-

structured and standardized previous data of the similar field that they have been provided with. Our prototype deals with applicants for IT companies but this can be generalized for numerous other regions by using an even more wide knowledge base.

The nominees will identify the languages they know along with the projects on which they have worked. This will help the employing company as they can easily filter out the nominees who do not have the knowledge of the language which is required by the company. Most websites use this as their filter technique by examining with a keyword. For example, if they want a nominee who knows Java, they can simply search for 'Java' in the resume to filter out nominees who do not know Java. But this technique does not tell the company anything about the proficiency level of the particular nominees in the language he or she knows. There is no way to tell how good the nominees is in Java.

3.2 KEYWORD SEARCHING

This is one of the most vital steps of our model. A knowledge base containing of numerous keywords is made from the initial training data. The input text which is received wants some pre-processing before it can be used. For this persistence, we use a POS tagger and a chunker which are used to divided the text into sentences, which are then analyzed by a syntactic parser which labels all the words with their part of speech information. Using a chunker supports in providing a flat structure of extracted data [14]. Lexical Analysis [15] can also be done to tokenize the words which can be then characterized for the purpose of parsing.

The keywords are pull out from the analyzed set of words. The nouns, verbs and adverbs are the portion of speech tags which are targeted for extraction while others can be dropped.

Pull out words are then compared with the keywords stored in the base. Every word stored in the knowledge base has a value linked with it. These values are well-defined based on the importance of the word. Meanwhile our prototype agreements only with resumes for jobs in IT companies, we have used numerous keywords which are extracted from the description of projects in which the nominees was involved. A huge set of valued keywords can be prepared and used to rate the nominees on the basis of words extracted from their resumes. The sum of all the keyword values is calculated to get a rating score which will be used

further to rank the resume and sort the nominee on the basis of rating.

3.3 ADDITION IN KNOWLEDGE BASE

Though the keywords found in resume text will be matched, the words which are not found in knowledge base are further analyzed and if found related, is added to the knowledge base. Since the data from which knowledge extraction has to be done is unstructured, we follow traditional techniques of information extraction. Apart from that, Ontology based Information extraction can also be completed by Semantic Annotation [16] in which we expand the natural language text into metadata which can be denoted in form of RDFa (Resource Description Framework in attributes) [17]. The procedure is distributed into two subtasks – Terminology Extraction and Entity Linking.

After getting the rating score of the resume, a nominee can be ranked on the basis of his resume's score. This will be useful in comparing two nominees while short listing them. Whenever the company searches for a nominee keeping in mind certain requirements, the nominee who is ranked above will be presented to the company first which would be adding to his advantage in cases where the vacancies available may not be high. More important process which has to be followed is of categorization. The sentiment analysis categorizes the people's thoughts as Positive, Negative or Neutral to derive results. Similar to that, our model would categorize nominees as Low, Average or High on the basis of their resume.

In our prototype, we have characterized the resumes of nominees applying for IT companies in the same 3 level scale and considered it as their proficiency level in the programming language stated in their project description which would help the company shortlist only those nominees whose proficiency level in a specific language is as required by the company.

In this way, the efficiency of recruitment process of a company could be significantly improved as better nominees would be picked up without needing to give a lot of time in going through the resumes manually.

4. IMPLEMENTATION

The algorithms for matching of keywords have been applied on Python facilitated by MySQL connector which fetches data necessary for matching from the table of Keywords and their associated values which form the Knowledge Base. The algorithm matches the extracted keywords with the keywords existing in the knowledge base and stores them in a different list along with their rating values. The rating scores of individual keywords after being added are returned to the nominees table for the purpose of their ranking on the basis of score. Categorization is performed on the basis of rating score of each nominee.

The instructions for rating and categorization followed in the prototype are as below -

Rating scale for singular keywords –

- 1: Low
- 2: Average
- 3: High

Rating Score = Sum of ratings of all keywords matched

Categorization on the basis of Rating Score –
Below 10: Low 10 to 20: Average Above 20: High

The nominees and the company will use a website based interface to interact. Both of them after getting registered as users shall be added in the database, separate for nominees and companies. The nominee database consists of various fields the nominee would have to fill in while registering which includes the programming languages known and projects in which the nominee has been involved along with its description. These are the crucial fields which will be used to determine the expertise level of the nominee. The fields for expertise level and rating score shall be automatically filled for every nominee once the resume is analyzed.

For terminology extraction, field specific lexicon can be used after tokenizing the text. After that, a link is produced between extracted lexical terms and the ideas from either ontology or the previously present knowledge base. Lastly, the context of the various terms is analyzed so that they can be correctly assigned to the level in which they should belong. In this manner, knowledge base can be repeatedly updated and also, it will be manually examined frequently to eliminate keywords which may no longer be beneficial.

The problem is that the present are not much flexible and efficient and time saving. It requires nominee, to fill the forms online than also you might not get the genuine information of the nominee. Beside where our system which saves the time of the nominee by providing to upload there resume in any format preferable to the nominee beside all the information in the resume our system will detect all its activity from the nominee social profile which will give the best nominee for that particular job and nominee will also be satisfied because he will get job in that company which really appreciates nominees skill and ability. On the other hand we are given that similar kind of flexibility to the client company.

5. Implementation Overview

The work related all the modules as per explained above has been implemented using web pages so below are some of those modules:

In figure 1, the user side page. Here user can upload their resume which is in text format, PDF format or word format. This resume can store in particular directory.

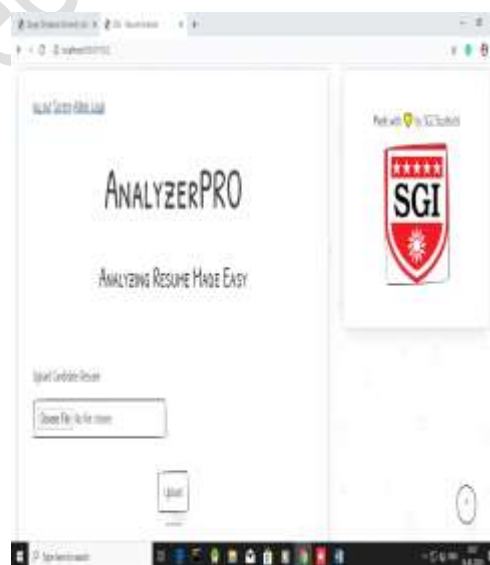


Fig1. Collection of resumes.

In figure 2, this is user side page. Here user can select job description as per their choice and knows which job is best according to their resumes.

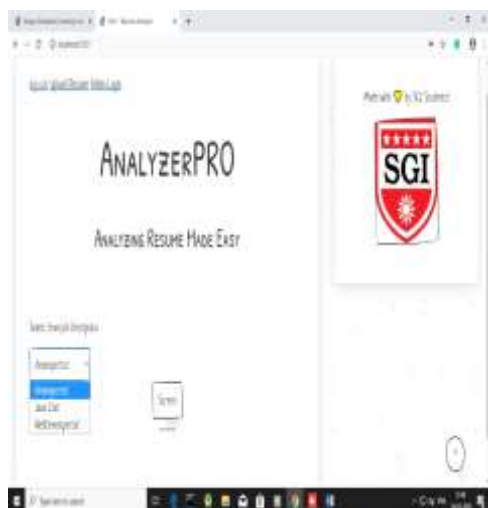


Fig2. Selecting feature based job description.

In figure 3, this is the final output of the project. After clicking on screen button this page will redirect this shows rank of the candidates according to their resumes for particular job.

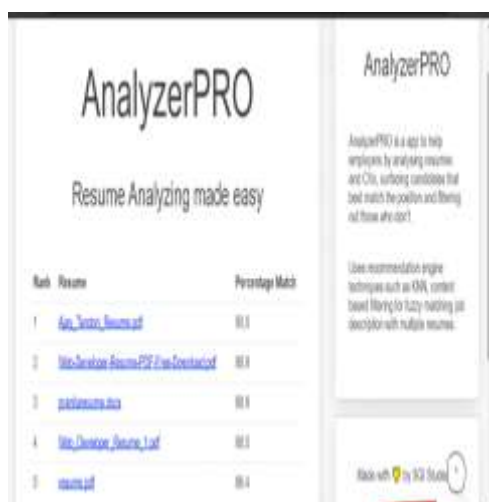


Fig3. Final output with percentage wise ranking.

6. Challenging Issues

Text mining is applied on text document and most text documents are in natural language which is ambiguous for extraction. Natural language complexity is main issue in text mining. One word may have multiple meaning and multiple words can have same meaning. One word or phrase may have several meanings those can lead to indistinctness problem. In arenas like Bioinformatics there are many names for a solo gene or protein that may also lead to indistinctness problem. One more problem with text mining is when we use the social media records i.e. status updates, tweets, comments, reviews etc. maximum persons use words like- "btw" for by the way, "ppl" for

people etc. these words do not exist in dictionary therefore they affect the mining.

REFERENCES

- [1] http://en.wikipedia.org/wiki/Text_Mining
- [2] http://en.wikipedia.org/wiki/Sentiment_analysis
- [3] Uldis Bojars, John G. Breslin, "ResumeRDF: Expressing Skill Information on the Semantic Web".
- [4] Uldis Bojars, "Extending FOAF with Resume Information".
- [5] Turney, P.D., Littman, M.: Unsupervised learning of semantic orientation from a hundred billion-word corpus. Technical Report ERC-1094 (NRC 44929), National Research Council of Canada (2002)
- [6] Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association, ACM Transactions on Information Systems (TOIS), vol. 21, no. 4, pp. 315--346 (2003)
- [7] Ujjal Marjit, Kumar Sharma and Utpal Biswas, "Discovering Resume Information Using Linked Data", in International Journal of Web & Semantic Technology, Vol.3, No.2, April 2012.
- [8] Maryam Fazel-Zarandi, Mark S. Fox, "Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach", International Journal of Computer Applications (IJCA), 2013
- [9] Koppurapu S.K, "Automatic Extraction of Usable Information from Unstructured Resumes to aid search", IEEE International Conference on Progress in Informatics and Computing (PIC), Dec 2010.
- [10] Zhi Xiang Jiang, Chuang Zhang, Bo Xiao, Zhiqing Lin, "Research and Implementation of Intelligent Chinese Resume Parsing", WRI International Conference on Communications and Mobile Computing, Jan 2009.
- [11] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao Bo, "Resume Parser: Semi-structured Chinese Document Analysis", WRI World Congress on Computer Science and Information Engineering, April 2009.
- [12] Celik Duygu, Karakas Askyn, Bal Gulsen, Gultunca Cem, "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs", IEEE 37th Annual Workshops on Computer Software and Applications Conference Workshops, July 2013.

- [13] Di Wu, Lanlan Wu, Tieli Sun, Yingjie Jiang, "Ontology based information extraction technology", International Conference on Internet Technology and Applications (iTAP), Aug 2011.
- [14] "Learning and Knowledge-Based Sentiment Analysis in Movie Review Key Excerpts" Björn Schuller and Tobias Knaup.
- [15] http://en.wikipedia.org/wiki/Lexical_analysis
- [16] Jalaj S. Modha Prof & Head Gayatri S. Pandi Sandip J. Modha, "Automatic Sentiment Analysis for Unstructured Data"
- [17] Ben Adida, Mark Birbeck "RDF in attributes"
- [1] <https://www.linkedin.com>
- [2] <http://www.monsterindia.com/>
- [3] <https://www.naukri.com/>
- [4] <http://www.indeed.com/resumes>
- [5] Ivens Portugal, Paulo Alencar and Donald Cowan, "The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review" .
- [6] <https://www.google.com>
- [7] <https://www.twitter.com>
- [8] <https://www.netflix.com>
- [9] Yu, Kun, Gang Guan, and Ming Zhou. "Resume information extraction with cascaded hybrid model." Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005.
- [10] Chandola, Divyanshu, et al. "ONLINE RESUME PARSING SYSTEM USING TEXT ANALYTICS."
- [11] Kopparapu, Sunil Kumar. "Automatic extraction of usable information from unstructured resumes to aid search." Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on. Vol. 1. IEEE, 2010.
- [12] Zhi Xiang Jiang, Chuang Zhang, Bo Xiao, Zhiqing Lin, "Research and Implementation of Intelligent Chinese Resume Parsing", WRI International Conference on Communications and Mobile Computing, Jan 2009.
- [13] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao Bo, "Resume Parser: Semi-structured Chinese Document Analysis", WRI World Congress on Computer Science and Information Engineering, April 2009.
- [14] CelikDuygu, Karakas Askyn, Bal Gulsen, GultunCem, "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs", IEEE 37th Annual Workshops on Computer Software and Applications Conference Workshops, July 2013.
- [15] Mayuri Verma, "Cluster based Ranking Index for Enhancing Recruitment Process using Text Mining and Machine Learning", International Journal of Computer Applications (0975 – 8887) Volume 157 – No 9, January 2017.
- [16] YixinCai, Ruixi Lin, Yue Kang, "A Personalized Company Recommender System for Job Seekers" .
- [17] Douglas Eck, Thierry Bertin-Mahieux, Paul Lamere, and Stephen Green, Automatic generation of social tags for music recommendation. Neural Information Processing Systems Conference (NIPS) 20. Vancouver, British.
- [18] Claudio Biancalana, Fabio Gasparetti, Alessandro Micarelli, Alfonso Miola, and Giuseppe Sansonetti, Contextaware movie recommendation based on signal processing and machine learning, Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation, p.5-10, October 23-27, 2011, Chicago, Illinois.
- [19] Yukun Cao, and Yunfeng Li, An intelligent fuzzy-based recommendation system for consumer electronic products, Expert Systems with Applications: An International Journal, v.33 n.1, p.230-240, July, 2007.
- [20] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In Proceedings of the Nineteenth International Conference on World Wide Web, 2010.
- [21] Rafael Sotelo, Yolanda Blanco-Fernández, Martín López-Nores, Alberto Gil-Solla, and José J. PazosArias, "TV program recommendation for groups based on multidimensional TV-anytime classifications," IEEE Trans. Consum. Electron., vol. 55, no. 1, pp. 248-256, Feb. 2009.
- [22] Feldman, Ronen, and James Sanger. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press, 2007.
- [23] J]Manning, Christopher D., and

HinrichSchütze. Foundations of statistical natural language processing. Vol. 999.

Cambridge: MIT press, 1999.

[24] <http://www.indeed.com/resumes>

[25] <https://cran.r-project.org/web/packages/tm/tm.pdf>

[26]

https://en.wikipedia.org/wiki/Uniform_Resource_Identifier

[27]

<https://cran.r-project.org/web/packages/RTextTools/RTextTools.pdf>

[28] <https://journal.r-project.org/archive/2013/RJ-2013-001/RJ-2013-001.pdf>