Towards partial fulfilment for Undergraduate Degree Level Programme
Bachelor of Technology in Computer Engineering

A first stage project Evaluation Report on:

Job Recommendation System

Prepared by:

Admission No.	Student Name
U18CO009	Tarsariya Keyur
U18CO010	Jadhav Ganesh
U18CO018	Shekhaliya Shubham
U18CO039	Nainuji Jigar

Class   :   B.TECH. IV (Computer Engineering)   7th Semester

Year   :   2021-2022

Guided by   :   Dr. Sankita J. Patel

DEPARTMENT OF COMPUTER   ENGINEERING
SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT
- 395 007 (GUJARAT, INDIA)

Student Declaration

This is to certify that the work described in this project report has been actually carried out and implemented by our project team consisting of

Sr.	Admission No.	Student Name
1	U18CO009	Tarsariya Keyur
2	U18CO010	Jadhav Ganesh
3	U18CO018	Shekhaliya Shubham
4	U18CO039	Nainuji Jigar

Neither the source code there in, nor the content of the project report have been copied or downloaded from any other source. We understand that our result grades would be revoked if later it is found to be so.

Signature of the Students:

Sr.	Student Name	Signature of the Student
1	Tarsariya Keyur	
2	Jadhav Ganesh	
3	Shekhaliya Shubham	
4	Nainuji Jigar	

Certificate

This is to certify that the project report entitled <u>Job Recommendation System</u> is prepared and presented by

Sr.	Admission No.	Student Name
1.	U18CO009	Tarsariya Keyur
2.	U18CO010	Jadhav Ganesh
3.	U18CO018	Shekhaliya Shubham
4.	U18CO039	Nainuji Jigar

Final Year of Computer Engineering and their work is satisfactory.

SIGNATURE:

GUIDE   JURY   HEAD OF DEPT.

Abstract

Talent acquisition is most import task for the success of the company. In current situation for a given job thousands of job seeker apply which make hard for the hiring team to go through each and every resume manually and check for the credibility of the applicant. Similar for the job seeker in the large market thousands of the jobs are available which makes finding suitable job difficult for the user. This project aims to solve this problem by making automation of the resume matching process by using various technique for the data extraction from the given text or description and finding similarity between the job seeker's profile and job description. Similarity can be found using the cosine similarity or Euclidean distance. Also data can be extracted from the resume to finding keywords in bags of words.

Keywords: Recommendation System - Job Recommendation System - Content based Filtering – Hybrid approach – Cosine similarity

contents

List Of Figures

List Of Acronyms

CF Collaborative filtering
CBF content-based filtering
RS recommendation System
JRS Job Recommender Systems

List Of Symbols

$\sum$ Summation

$\sqrt{}$ Square Root

Chapter 1

Introduction

On the Internet, there is an overwhelming amount of information available for every domain. In the Internet era, we have been overwhelmed by immense amounts of information from various sources, making finding functional or suitable or fruitful items/objects such as newspapers, websites, products, songs, movies, books, or even jobs a big challenge. For this reason, more and more applications have been broadly developed and new techniques have emerged to support human decisions suggesting services, products and various types of information to customers. One field of research in this direction is that of Recommender Systems. Recommender Systems are tools that use various techniques and algorithms to isolate irrelevant information from a huge amount of data and generate personalized suggestions of a small subset of them that a user can examine in a reasonable amount of time. The increasing usage of the Internet has heightened the need for online job hunting. The critical problem is that most job-hunting websites display recruitment information to website viewers. Many websites on the Internet give employment opportunities, but the task is tedious as students need to go through a large amount of information, taking lots of time and energy and suffering from unwanted or less helpful information. Jobseekers have to retrieve all the information to find jobs they want to apply for. The whole procedure is tedious and inefficient. One field of research in this direction is that of Recommender Systems.

1.1   Applications

Dealing with the tremendous amount of recruiting information over the Internet, a job seeker always spends hours finding useful ones. With a huge number of different job roles existing today along with the typically large number of applications received, short-listing poses a challenge for the human resource department. This is only further worsened by the lack of diverse skill and domain knowledge within the HR department, required for effective screening. Being able to weed out non-relevant profiles as early as possible in the pipeline results in cost savings, both in terms of time as well as money.[16]

1.2   Motivation

Talent acquisition is an important, crucial, complex, essential task in industries that requires a significant amount of time. Talent acquisition has the most challenging part. The lack of a standard structure and format for a resume makes a short listing of desired profiles for required roles very tedious and time-consuming. Effective screening of resumes requires domain knowledge to understand the relevance and applicability of a profile for the job role. With a massive number of different job roles existing today and the typically large number of applications received, short-listing challenges the human resource department.

In addition, in most recommendation systems, the most general application of recommendation algorithms uses CF algorithms without considering the user's resume and job description. That means candidates' resumes and details of recruiting information. So we proposed an improved algorithm based on CBF. Our aim is to give an effective method of recommendation for online job hunting and talent hunting. We hope to offer candidates a personalized service that can help them find ideal jobs quickly and conveniently.

1.3 Objectives

The e-recruiting platforms are usually based on Boolean search and filtering techniques that cannot sufficiently capture the complexity of a person-job fit as selection decisions. Much literature has applied the recommender system concept to the job problem. Recommendation between entities of the domain: users and opportunities

The job recommendation problem is a bidirectional recommendation between job-seeker and job.[4] Two viewpoints are distinguished: from recruiters and job seekers. The recruiters generate the job description by determining the set of requirements and constraints on skills, expertise levels, and degrees. The job-seeker, on the other hand, generates the candidate's resume by specifying the academic background, previous work experience and skills[4].

Based on the requirement that a good match between jobs and persons needs to take into account both the preferences of the candidate and the preferences of the recruiter to recommend the job.

1.4 Contribution

We are going to job seeker's skill and preferences for job and the job description in to the consider to recommend the job for that we need to find the similarity between the job seeker's profile and job's profile. So, we can create a vector of words containing the keywords for the job and later we can use any similarity algorithm like cosine similarity to recommend the job to the job seeker.

1.5 Organization of project report

This chapter covers the introduction to the project along with its application, motivation, objective and overview. Chapter 2 presents a theoretical background of the terminologies in the job recommendation as well as other important concepts needed to understand the project better. Chapter 3 is the Literature Survey which summarises the work done in the job recommendation to recommend the job. Later in Chapter 3, a review of various job

recommendation algorithms is discussed. Our proposed methodology and logic development of the same is covered in Chapter 4. Chapter 5 discusses the brief overview about our data. Chapter 7 ends the report with conclusion and future work proposed.

Chapter 2

Theoretical Background

This chapter discuss various recommendation technique and how this technique works and the advantages and limitation of the various recommendation technique. Finally, it contains overview of different RS.

2.1   Overview of Recommendation System

We often seek suggestions from friends, colleagues or known ones whenever we want to buy something like a refrigerator, TV, mobile phone or washing machine or even when planning for the Trip or which book to refer to or which movie or song for entertainment. Even with their best intentions, these friendly suggestions sometimes do not fit us or are effective in our case. Not just in decision making plays an imperative part to settle on choices which help to pick up benefits by connecting the best alternative as a suggestion. The point is that it is very arduous to highlight a precise suggestion on the items on which we might be interested.

One field of research in this direction is that of Recommender Systems. RSs are tools that use various techniques and algorithms to isolate irrelevant information from a huge amount of data and generate personalized suggestions of a small subset of them that a user can examine in a reasonable amount of time. An RS is an intelligent computer-based technique that predicts on the basis of users' adoption and usage and helps them to pick items from a vast pool of online stuff[2], Or it identifies the users' needs automatically by inferring the needs from the user's item interactions. Alternatively, the recommender system asks users to specify their needs by providing a list of keywords or through some other method[3]. RSs are a useful alternative to search algorithms since they help users to.

These are the systems that help us to select similar things whenever we select something online. The concept of understanding a user's preference by their online behaviour, previous purchases, or history in the system is called a recommender system [12]

The recommender systems techniques can be used to address the problem of information overload by prioritizing the delivery of information for individual users based on user preferences. Recommender Systems are tools that use various techniques and algorithms to isolate irrelevant information from a huge amount of data and generate personalized suggestions of a small subset of them that a user can examine in a reasonable amount of time. So The task of the recommendation system is to help the user to concentrate on the area of interest.

Following are the approaches of the job recommendation System.
I. Collaborative Filtering recommenders
II. Content Based Filtering recommenders
III. Knowledge-based recommenders
IV. Hybrid Recommenders

2.2   Collaborative Filtering Recommenders

Collaborative filtering uses similarity between users and items simultaneously to provide recommendations. CFR finds users with similar interests as the target user and suggests recommendations to him/her based on their liked items. The key function in CFRs is the computation of similarities among users.[13]
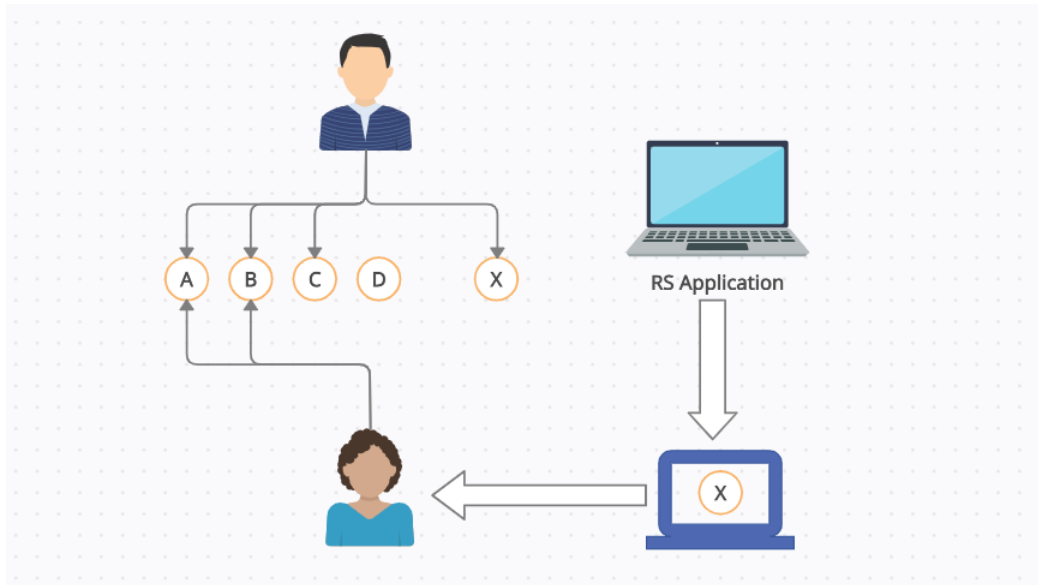
Figure 2.1: Collaborative Filtering System

As show in figure 2.1 as User P likes A, B and X the recommender system will try to recommend item X to the user Q

Advantages

- CFR does not require processing of the items so it is content independent.
- CFR takes feedback from the users and the history of the users to recommend the items to the user.
- CFR recommendations are based on user similarity.

Limitation

- A general problem of CFR is the cold start problem, which may occur in three situations: new users, new items, and new communities or disciplines. If a new user rates few or no items, the system cannot find like-minded users and therefore cannot provide recommendations. If an item is new in the system and has not been rated yet by at least one user, it cannot be recommended. In a new community, no users have rated items, so no recommendations can be made and as a result,

- As it runs on the user's feedback, false feedback from the user can also cause the wrong recommendation.
- Makes the criticism that collaborative filtering systems are black boxes that cannot explain why an item is recommended except that other users liked it.

2.3   Content Based Filtering

Content-Based filtering is based on a description of the item and a profile of the user's preferences. Items are recommended having similar content information to those a user has. CBF analyses the similar characteristics of the item and target users based on that build the profile for the user. In this system keywords are extracted from the item and user's description to find similarity between them. only the most descriptive features are used to model an item and users and these features are typically weighted. Once the most discriminative features are identified, they are stored, typically as a vector that contains the features and their weights. The user model typically consists of the features of a user's items. To find recommendations, the user model and recommendation candidates are compared in e.g. the vector space model To abstract the features of the items and user TF-IDF algorithm can be used to show similarity.[14]

TF-IDF
- Stands for term frequency and inverse document frequency.
- TF-IDF assumes a document is just a "bag of words".
- TF just measures how often a word occurs in a document.
- DF is how often a word occurs in an entire set of documents.
- Words with high TF and DF both might not be an important measure relevant to the document.
- So, a measure of relevance of a word to a document might be TF/DF or Term frequency * Inverse document frequency.[15]
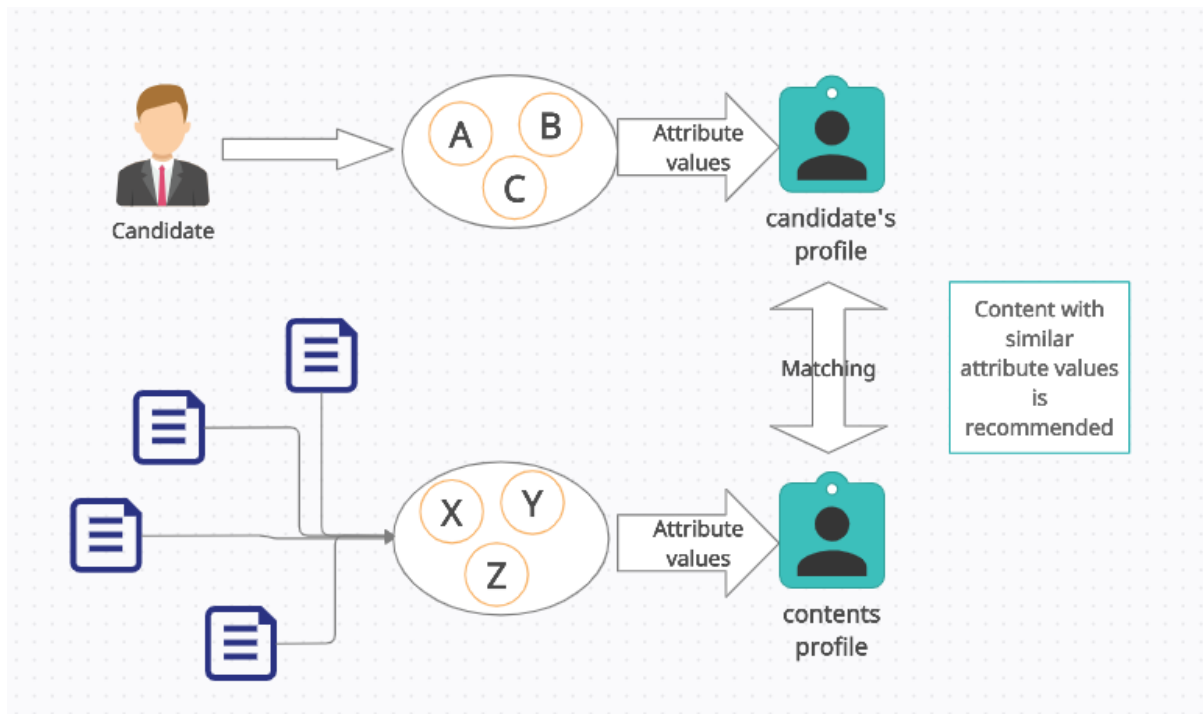
Figure 2.2: Content Based filtering System

As show in figure 2.2 it extracts attributes from the user and also from the job description later find similarity between them using various known technique.

Advantage

- The model doesn't need any data about other users, since the recommendations are specific to this user. This makes it easier to scale to a large number of users.
- The model can capture the specific interests of a user, and can recommend niche items that very few other users are interested in.
- The user gets recommended the types of item they love.
- The user is satisfied by the type of recommendation
- New items can be recommended; just data for that item is required

Limitation

- The model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.
- The user will never be recommended for different items.

- Business cannot be expanded as the user does not try a different type of product.
- If the user matrix or item matrix is changed, the cosine similarity matrix needs to be calculated again.

2.4   Knowledge Based Recommender

To recommend the items which are less frequently used. In this technique, the relationship between user and item can be explicitly modelled. By using the knowledge of an item based on rules and patterns, we can recommend how a particular item is suitable for the user.[4]
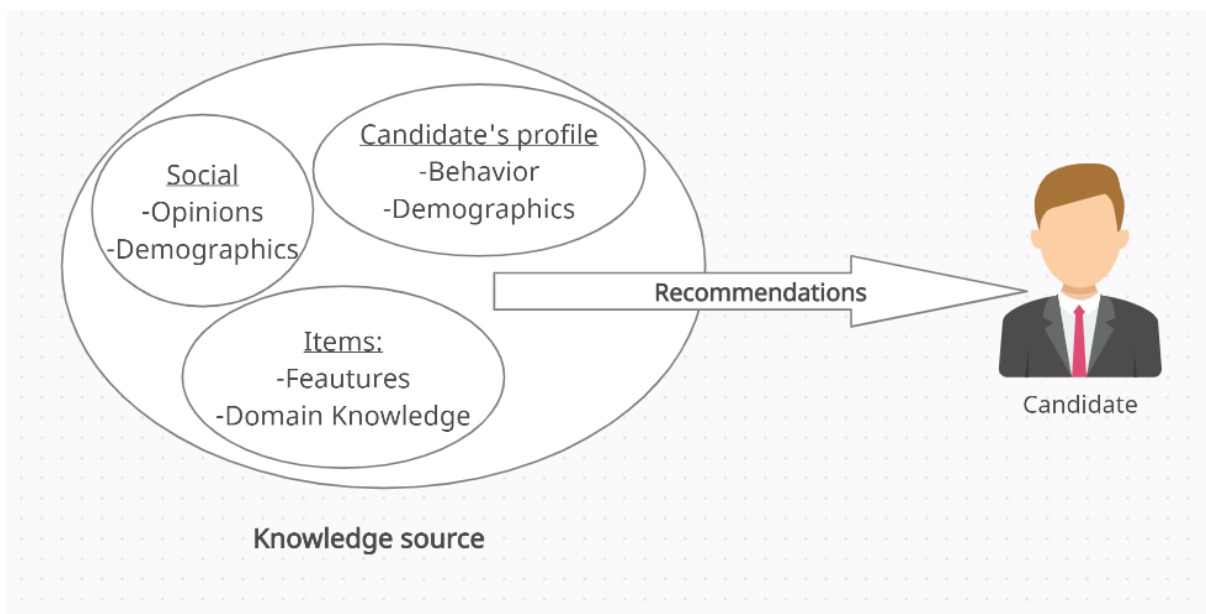

Figure 2.3: Knowledge Based Recommender System

Advantage

- It can recommend the new item to the user even when item is new in the system as it solves the problem of cold start

2.5   Hybrid Recommender System

Hybrid recommender technique is a mix of other techniques to override the drawback of the existing techniques.
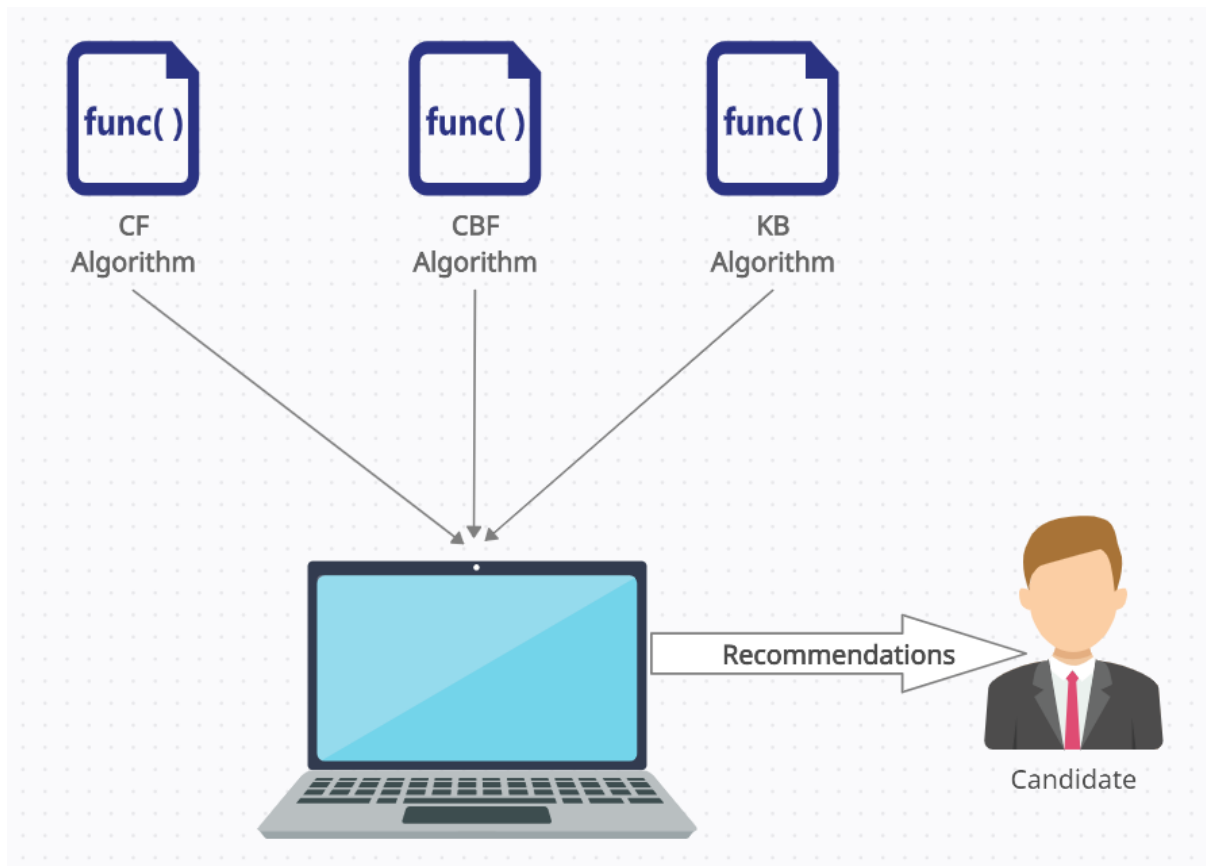

Figure 2.4: Hybrid Recommender System

As show in figure 2.4 Hybrid Recommender System takes input from the all-suitable recommendation technique for the job recommendation problem.

2.6 finding similarity

Let say There is two Vector A and B then we can find similarity between them using cosine similarity formula.

$$similirity = cos(\theta) = \frac{A \cdot B}{||A|| \, ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

(2.1)

Similarity can also be find using Euclidian distance using below formula

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

(2.2)

Chapter 3

Literature Survey

This chapter briefly discusses the existing literature in the field of Job recommendation (JRD) system. Extracting the data from the resume of the user and job profile and relate them to show recommendations.

3.1 Overview of the recommendation system

Because of the Internet, companies have changed their hiring process by using the online platform. Companies choose to use online platforms because recruiting the appropriate person is a challenge faced by most companies. The unavailability of specific candidates in some skill areas has long been identified as a significant obstacle to the company's success.

3.2 Boolean matching technique

Online channels like Internet job portals, social media applications, or a firm's career website have driven this development. While the companies established job positions on these portals, job-seekers use them to publish their profiles. For each posted job, thousands of resumes are received by companies. Consequently, a huge volume of job descriptions and candidate resumes are becoming available online. This vast volume of information gives a great opportunity for enhancing the matching quality; this potential is unused since search functionality in recruiting applications is mainly restricted to Boolean search methods. The need increases for applying the recommender system technologies that can help recruiters to handle this information efficiently.[6]

3.3 Context to recommend

We must consider unary attributes such as individual skills, mental abilities, and personality that control the fit between the individual and the tasks to be accomplished [7], as well as the relational attributes that determine the fit between the individual and the upcoming team members.

In this context, literature usually distinguishes between
(1) person-job
(2) person-team
(3) person-organization fits

Many types of research have been conducted to discuss different issues related to the recruiting problem as well as the application of recommender system technologies. However, job recommendation is still a challenging domain and a growing area of research.[4]

Some of the followings are existing systems for a job recommendation.

- Hybrid job recommender System
  - A probabilistic hybrid approach
  - A proactive job recommender system
- Content-based job recommender systems
  - Machine learned recommender system

3.4 Hybrid job recommender systems

3.4.1 A probabilistic hybrid approach

The recommen-dation approach used both concepts: content-based filtering and collaborative filtering simultaneously. Its understands the individual preferences as a combination of preference factors. In a basic approach for collaborative filtering, we look at each value of user/

object pairs (x, y), where x is a set of users and y is a set of objects. The model can then be represented as a variable z which is associated with each value of (x, y), assuming that x and y are independent conditioned on z. The model parameters are then estimated using the Expectation Maximization (EM) algorithm.[8]

This model produced a rating matrix that assigns assessed values to candidate``s profile containing the probability that recruiter x rates candidate y with value v. Later, they defined v = {"qualified", "not qualified"}. Then, they transformed the rating matrix by replacing variable y with a variable a to represent the attributes that were extracted from the candidate resumes. As many attributes are assigned to several profiles, we will see the attribute several times with different values v.

To improve the match between people and jobs: a CV-recommender and a job recommender, separately. In the first step, they built a system recommending CVs that are similar to resumes previously selected by the recruiter for a specific job profile. In the second step, they developed a second recommendation system that recommends jobs to candidates based on their preference profiles which are in turn based on previous preference ratings.[8]

Limitations
- It answers in binary only either 0 or 1 cannot answer in rank wise to give recommendations.

3.4.2   A proactive job recommender system

The proactive recommender system is an adaptive system that attempts to integrate the idea of recommender systems.[5] This system contains five components: web spider, ontology checker, profile analyzer, preference analyzer, and user interface generator. Web spider is a parser that periodically acquires job information from an exterior source. The ontology checker matches information with ontologies and performs the classification. Then, the job data is stored in a pre-designated form. The profile analyzer makes the recommendations whenever the users modify the group of favorites by comparing the weight differences with current open jobs. Then, a list of recommended jobs is generated. Finally, the preference analyzer deduces

the explicitly defined user's preferences and gives a recommendation for preferred jobs after calculating the similarity of jobs to the user's preference.[10]

Limitations

- One way recommendation only recommend to the job seeker
- Cold start problem as user change profile

3.5 Content-based job recommender systems

3.5.1 Machine learned recommender system

The recommendation problem is treated as a supervised machine learning problem. They build an automated system that can recommend jobs to applicants based on their past job histories, in order to facilitate the process of choosing a new job. An item in this learning model represents a person who is hired in an organization. Each item is characterized by a set of features extracted from the candidate's resumes. Given a person who is currently working in an organization, they want to predict the next organization. If the accuracy of such predictions is sufficiently high, the model can be used to recommend organizations to employees who are seeking jobs. This approach uses all past job transitions as well as the data of both employees and organizations to predict an employee's next job transition. They train a machine learning model using a large number of job transitions extracted from person profiles available on the web.[11]

Limitations

- As it takes previous or historic data into the consideration, the problem of sparsity and cold start could occur.

Chapter 4

Proposed Work

Project aims to extract useful information from the large number of resume for the particular job and finding the similarity between them to recommend the job to the job seeker or give result to the employer which resume are best fit for the particular job or role.

4.1   Logical Development

We will present one system for screening candidates, which is a decision support tool assisting recruiters to shortlist candidate resumes. It will mine Resume extracting features of candidate profiles such as skills, education, location, and experience. It will use information retrieval techniques to rank candidates for a given job position.

For each job, the system ranks candidates based on the similarity between job profiles and candidate's resumes. The ranking can be processed by adding filtering criteria. These filtering criteria will be defined based on the candidate meta-data and the information automatically extracted from the candidate's resumes.

This system consists of three main components:
- Batch processor
- Query processor
- Resume matcher.

The batch processor initially processes new applications. It First stores the candidate meta-data in the main database and extracts data from the candidate resume, which will be saved in the extracted database. This information is helpful for the query processor and the resume matcher to provide the ranking candidate list for a given user query.
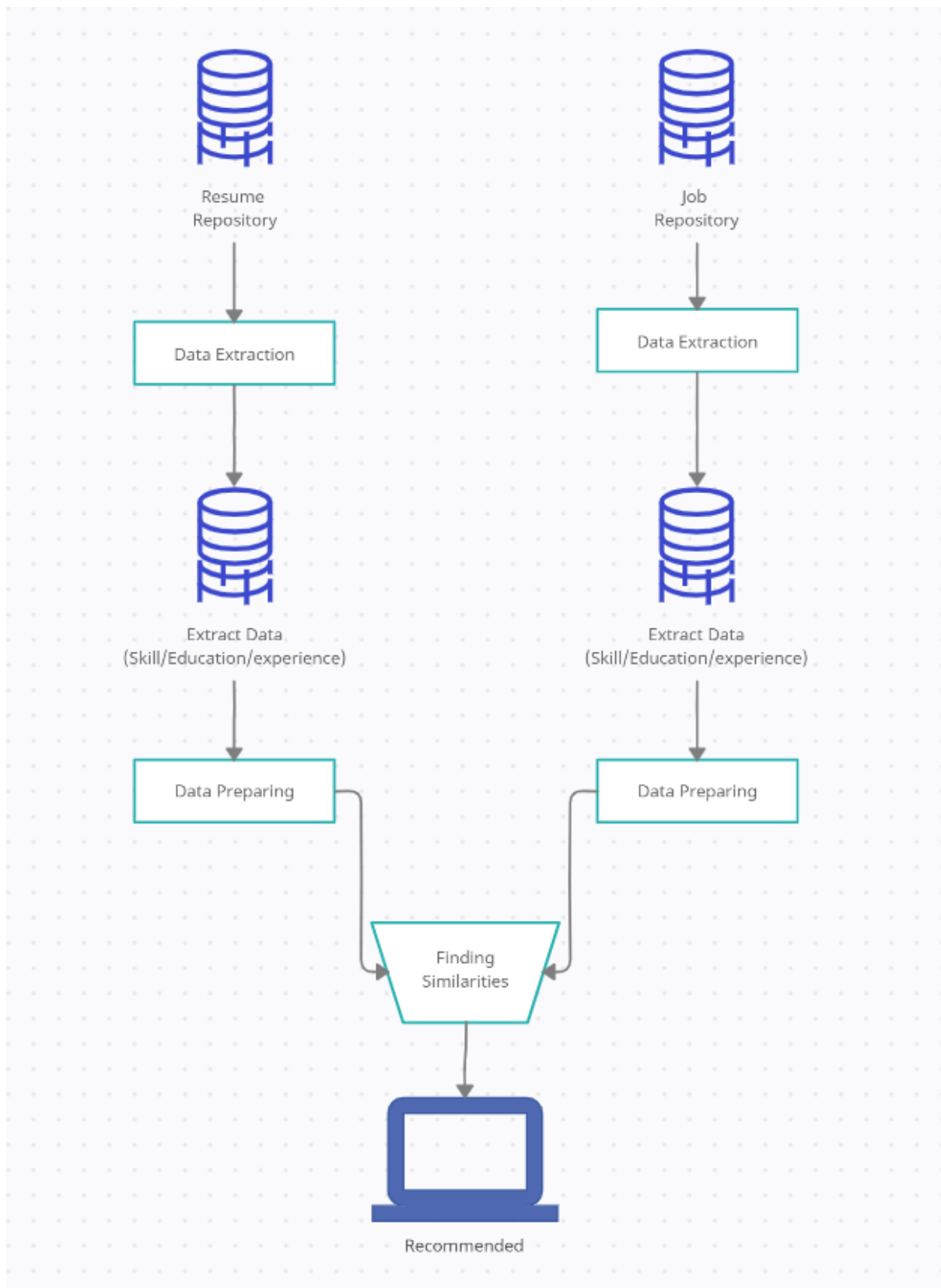
Figure 4.1 High Level approach of the problem

The recommendation process is divided into two parts: job recommendation and job-seeker recommendation. For both parts, the recommendations should be the objects which are the most consistent with their preferences. The useful keywords are extracted from the user's resumes.

The similarity of different preferences is calculated using different methods. Finally, the complete similarity is calculated, and the recommendation is generated. The steps of recommendation are as follows:

> (1) user's preferences are extracted from the content of the user's resumes and then, the attribute is determined and converted to vector.
> (2) The similarity calculated between users in turn then calculates the reciprocal score.
> (3) Finally, the recommendation is generated by ranking the reciprocal scores to present the top-n recommendations

Chapter 5

Simulation

This chapter discusses the visualisation of the data set for. It consists of bar graph for given skill in the market for a job-seeker.
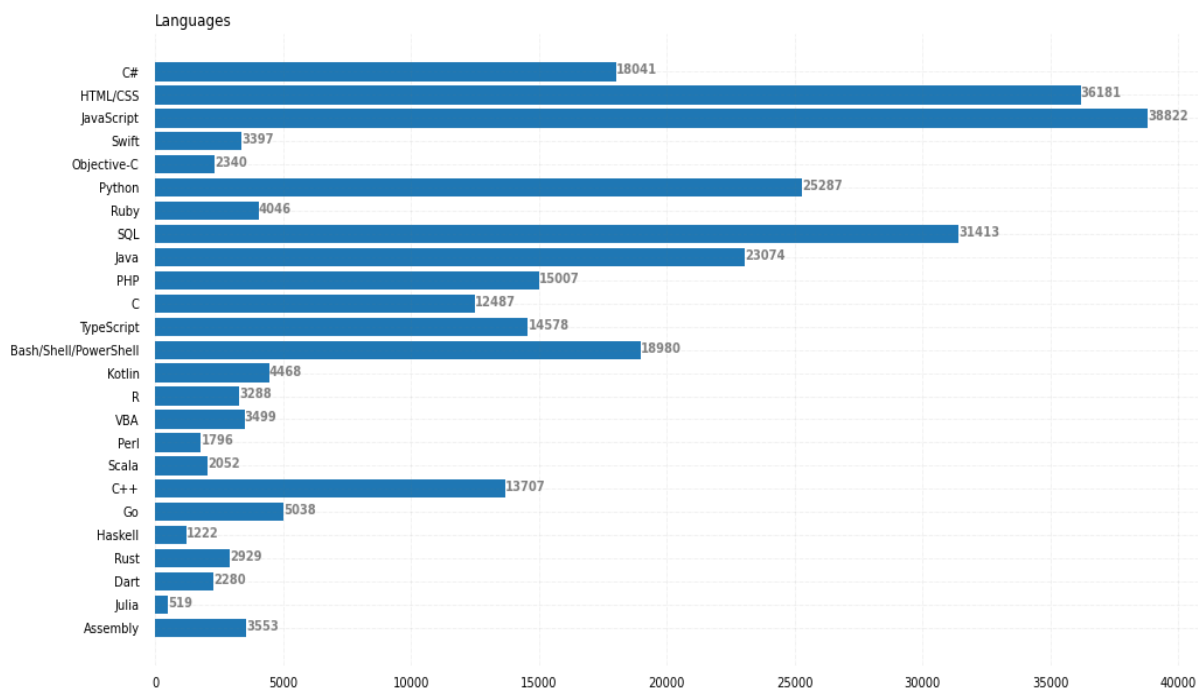
5.1   visualisation

Figure 5.1 Show number of users know particular Language

No of Candidate who knows the particular Framework

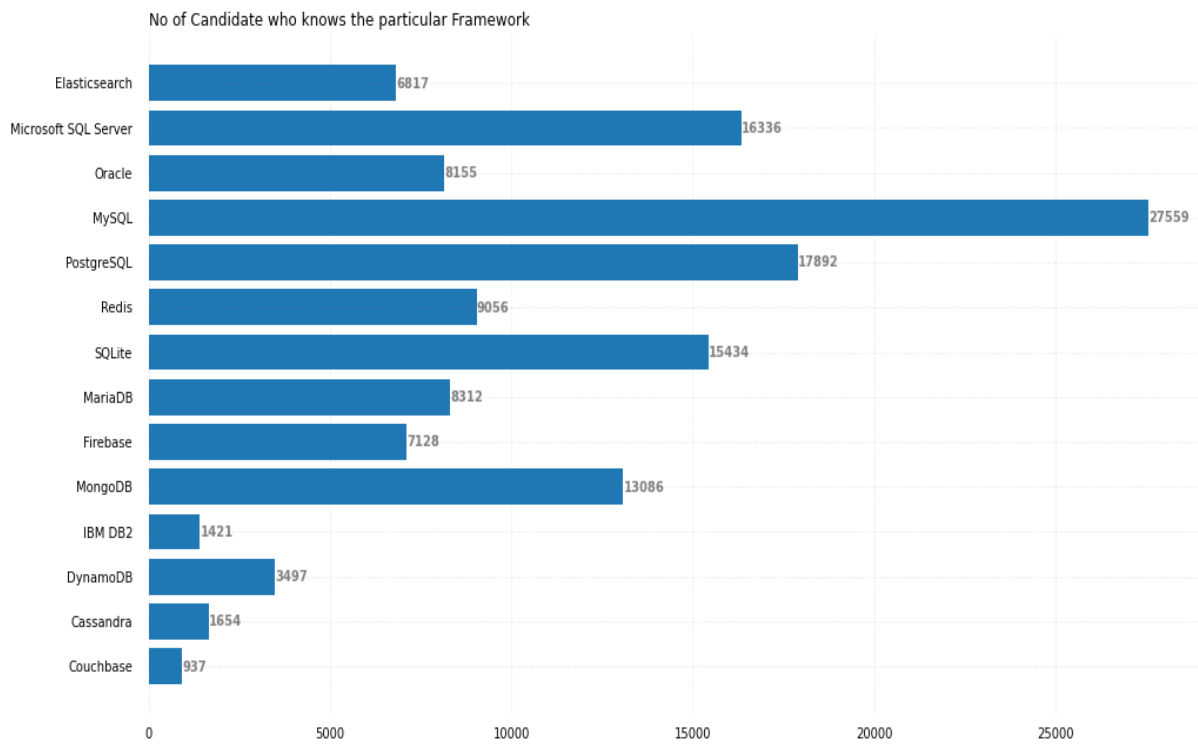| Database	Value
Elasticsearch	6817
Microsoft SQL Server	16336
Oracle	8155
MySQL	27559
PostgreSQL	17892
Redis	9056
SQLite	15434
MariaDB	8312
Firebase	7128
MongoDB	13086
IBM DB2	1421
DynamoDB	3497
Cassandra	1654
Couchbase	937

Figure 5.2 show number of users familiar with particular Database

Chapter 6

Conclusion and Future Work

6.1   Conclusion

Talent acquisition is most import task for the success of the company. In current situation for a given job thousands of job seeker apply which make hard for the hiring team to go through each and every resume manually and check for the credibility of the applicant. Similar for the job seeker in the large market thousands of the jobs are available which makes finding suitable job difficult for the user. This project aims to solve this problem by making automation of the resume matching process by using various technique for the data extraction from the given text or description and finding similarity between the job seeker's profile and job description.

6.2   Future works

In this report we have currently described the present literature survey and the technique which can be used to recommend the job. After deciding which technique is better, we can implement the suitable best technique or the mix of techniques. Also keywords and data extraction can also be performed in future to gather data and build profile for user or job.

References

[1] Tanya V Yadalam, Vaishnavi M Gowda, Vanditha Shiva Kumar, Disha Girish And Namratha M "Career Recommendation Systems using Content based Filtering" In Proceedings of the Fifth International Conference on Communication and Electronics Systems

[2] Pradeep Singh, Dutta Pramanik Pijush, Avick dey and Choudhury, Prasenjit. (2021). "Recommender Systems: An Overview, Research Trends, and Future Directions". In International Journal of Business and Systems Research. 15. 14–52.

[3] Beel, J., Gipp, B., Langer, S. *et al.* "Research-paper recommender systems: a literature survey". *Int J Digit Libr* 17, 305–338 (2016)

[4] Shaha Alotaibi, "A survey of job recommender systems" Article in International Journal of Physical Sciences · July 2012.

[5] Schafer J, Konstan J And Riedl J "Recommender Systems in E-Commerce" In Proceedings of the First ACM Conference on Electronic Commerce (EC 1999), Denver, Colorado, USA, ACM pp. 158-166

[6] Frank Faerber, Tim Weitzel And Tobias Keim, "An Automated Recommendation Approach to Selection in Personnel Recruitment" (2003). AMCIS 2003 pp 302.

[7] Tomoki sekiguchi, "Person-Organization Fit and Person-Job Fit in Employee Selection: A Review of The Literature" vol. 54, no. 6 March 2004

[8] Thomas Hofmann And Jan Puzieha "Latent Class Models for Collaborative Filtering" In Proceedings of the 16th International Joint Conference on Artificial Intelligence, Stockholm, Sweden pp. 688-693.

[9] Keim T." Extending the Applicability of Recommender Systems:A Multilayer Framework for Matching Human Resources." In Proceedings of the 40th Annual Hawaii International Conference on System Sci. (HICSS'07). Hawaii, USA, IEEE

[10] Lee DH And Brusilovsky P, "Fighting Information Overflow with Personalized Comprehensive Information Access: A Proactive Job Recommender" In Proceedings of the 3rd Conference on Autonomic and Autonomous System, Athens, Greece, IEEE pp. 21-26

[11] Paparrizos I, Cambazoglu BB, Gionis A (2011). Machine Learned Job Recommendation. In proceedings of the fifth ACM conference on Recommender systems, RecSys"11, Chicago, Illinois, USA, ACM pp. 325-328.

[12] Baptisite Rocca, "Introduction to recommender System" at (https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada) (accessed on 23-09-2021)

[13] "Collaborative Filtering and Matrix Factorization" at (https://developers.google.com/machine-learning/recommendation) (accessed on 23-09-2021)

[14] "Content based filtering" (https://developers.google.com/machine-learning/recommendation) (accessed on 23-09-2021)

[15] Ankur Dhuria, "How to do content based filtering using TF-IDF?" at (https://medium.com/analytics-vidhya/how-to-do-a-content-based-filtering-using-tf-idf-f623487ed0fd) (accessed on 23-09-2021)

[16] Yi, Xing & Allan, James & Croft, W. "Matching resumes and jobs based on relevance models". In Jan 2007.

Acknowledgement

We would like to express our deep gratitude to our project guide, Dr. Sankita J Patel, Associate Professor, Computer Engineering Department, SVNIT Surat, for their valuable guidance, helpful feedback, and co-operation with a kind and encouraging at the initial stage. We would also like to thank Dr. Mukesh A. Zaveri, Head of Department, Computer Engineering Department. We are also thankful to SVNIT Surat and its staff for providing this opportunity which helping us to gain sufficient knowledge to make our work successful. Special thanks and appreciation to our colleagues in developing and people who have willingly helped us out with their abilities.