Deanship of Graduate Studies

Al-Quds University

Employment Recommendation System Using Content-Based Filtering Approach and Fuzzy Logic

Amani Basem Mohammad Habboub

M.Sc. Thesis

Jerusalem – Palestine

1441 / 2020

Employment Recommendation System Using Content-Based Filtering Approach and Fuzzy Logic

Prepared by:

Amani Basem Mohammad Haboub

B.S. Information and Communication Technology, Al-Quds Open University, Palestine

Advisor: Dr. Badie Sartawi.

“A thesis Submitted to the Faculty of Science and Technology, Al-Quds University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer”

1441 / 2020

Al-Quds University
Deanship of Graduate Studies

Computer Science Department

Thesis Approval

Employment Recommendation System Using Content-Based Filtering Approach and Fuzzy Logic

Prepared by: Amani Basem Mohammad Haboub

Registration No.: s1412354

Advisor: Dr. Badie Sartawi

Master Thesis submitted and accepted, Date: 19 / 8 / 2020

The names and signatures of the examining committee members are as follows:

1- Head of Committee:	Dr. Badie Sartawi	Signature	Badie Sartawi
2- Internal Examiner:	Dr. Yousef Abuzir	Signature	
3- External Examiner:	Dr. Nidal Kafri	Signature	Nkafri

Jerusalem – Palestine

1441 / 2020

Dedication:

This work is dedicated to my father soul, to my mother for her endless Love, Support and Encouragement.

Amani Basem Mohammad Haboub

Declaration:

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed: 

Amani Basem Mohammad Haboub.

Date: 19 / 8 / 2020.

Acknowledgments:

I thank God the Almighty for sustaining me until the completion of this work.

I would also like to express my sincere gratitude to my supervisor Dr. Badie Sartawi who was abundantly helpful and offered invaluable assistance, support and guidance through my Master study and research.

Amani Basem Mohammad Habboub.

Abstract:

Nowadays, there is a gap between the requirements of the labor market and the scientific specializations, which are available in universities. The existence of similar disciplines with different names in universities leads to a large number of specializations, especially for a specific number of jobs. These things we mentioned before make it difficult to find and match the most suitable specialization, which based on the educational outcomes for the available jobs in the labor market.

The importance of the study reveals in matching Automatic job with an available job positions is a big problem for organizations and applicants. There are many techniques and algorithms to help job seekers find the right job. Some are traditional algorithms while others have been found recently, also there are a large number of hybrid algorithms that are a combination of many algorithms. All these algorithms aim to find the best job for the candidate and to make recruitment processes faster, more accurate and transparent.

The literature review focused on how we can help job seekers find the most suitable jobs using techniques such as analysis, extraction, and web crawling. The problem with current method is that it calculates the similarity between job applications (job profile) and CV (user profile, this happens without measuring whether the recommended candidate is satisfying or meets the employer's goals or needs.

The main question of this work (problem statement) is how to achieve the best matching between candidates and offered jobs. So, we suggest a hybrid system based on NLP tasks, BabelNet dictionary, content base filtering approach and finally fuzzy logic, all that allow system users (i.e., applicants and employers) to search for suitable jobs with matching

them with job offers, which leads to best presentation for each candidate. We also suggest building a database contains different specialties and appropriate jobs, according to the opinion of stakeholders, such as academics who teach that specialization and the opinion of some workers in the labor market.

Our main contribution is proposing a model to manage the recruitment process by evaluating the applicants' CVs, which contents weighting match to the offered/posted (open jobs) requirements by recognizing the employers' needs.

In order to make our finding approach closer to reality, transparent and to reduce effort and time applicants and employers spend, we take into account the opinions of academics working in Palestinian universities in different disciplines and their relationship to the labor market.

Also, the study provides an opportunity for applicants, who were not accepted when the education section was evaluated from 100%, or getting a low match ratio to be acceptable, and has a good chance to employ. The expert knowledge base we created effects on results by giving a wider chance to those who had a low employing chance to be accepted, also the study shows that the employee opinion effected in the candidate list.

Finally, as appears from candidate lists, the system is well effective in extracting lists according to the proportion of matching required for each section in the employment announcement, and in terms of accuracy. The results are satisfactory, according to seekers qualifications, there are very close matching rates in some cases and spaced out in other cases. Also, it is easy to use through the simple user interface, and flexible to deal with by consistent with the requirements of the employer expressed, by a certain percentage ratio.

Table of Contents

List of Tables……..……..……..... vii
List of Figures……..……..……..... viii
List of Abbreviates……..……..……..…….. ix
Chapter 1: Introduction..……..……..…… 1
1.1. Introduction……..……..……..…….. 1
1.2. Problem Statement……..... 1
1.3. Hypothesis……..……..... 2
1.4. Our Contributions……..……..…….. 3
1.5. Thesis Structure……..... 5
Chapter 2: Literature Review……..……... 7
2.1. Introduction……. 7
2.2. Automatic Match Job Offers and Candidate Profiles……..... 7
2.3. Effective Job Recruitment System Using Content-based Filtering……. 9
2.4. Job Recommendation System Using Profile Matching and Web-Crawling……..... 10
2.5. Implementation of an Automated Job Recommendation System Based on Candidate Profiles……..…….. 12
2.6. Fuzzy AHP approach for employee recruitment…….. 14
2.7. A Fuzzy Expert System Tool for On-line Personnel Recruitment……..…… 16
2.8. Summary……..… 18
Chapter 3: Background……..……..…… 20
3.1. Introduction……..……..……..……..... 20
3.2. Job Board and Search Engine……..……..……..... 21
3.3. Best Job Sites……..……..... 22
3.4. Advantages and Disadvantages of e-recruitment……..……...… 24
3.5. Recommendation system..……..……..…..…. 25
3.5.1. Collaborative filtering approach……..……..……..…. 27
3.5.2. Content base filtering approach……..……..……..….. 29
3.5.3. Knowledge-based approach……...……..…….....…... 31
3.5.4. Hybrid filtering approach……..……..... 32
3.6. Fuzzy logic……..……..…..... 35
3.6.1. Fuzzy Set……..… 35
3.6.2. Membership function……..……..…….....…. 35
3.6.3. Fuzzy logic operations……..…….. 37
3.6.4. Fuzzy Logic Architecture……..……..…..… 37
3.6.5. Fuzzy interface system ……...…..... 38
3.6.6. Characteristics of fuzzy logic……..…… 40
3.6.7. Advantages and disadvantages of Fuzzy Logic System..... 40
3.7. BabelNet……..……...…... 41
3.7.1. Wikipedia……..……..……..…... 42
3.7.2. WordNet……..…..……..…… 43
3.7.3. BabelNet advantages……..……..……..…… 44

3.8. Natural language process…… 45
3.8.1. Natural Language process model “bag of words” steps……. 45
3.8.2. Classification of Stemming Algorithms…….. 46
3.8.3. Porters Stemmer…….. 47
Chapter 4: Recommendation Model ……..... 48
4.1. Introduction……...……...……. 48
4.2. The proposed system Architecture..……..……. 50
4.2.1. The natural language processing ……..……..…. 50
4.2.2. Bable dictionary……..……..….... 51
4.2.3. Content base filtering approach……...……...……..…. 52
4.2.4. Mamdani fuzzy Logic…….…….... 60
4.3. Conclusion…….…… 67
Chapter 5: Implementation …….. 68
5.1. Introduction……...…… 68
5.2. Discussion…….…..… 69
5.3. The implementation of the proposed system…….…... 73
5.4. Conclusion…… 80
Chapter 6: Experiment Results and Analysis…….…… 82
6.1. Introduction…….…..... 82
62. Data Collection……. 82
6.3. The proposed system experiment…… 93
6.4. Performance Evaluation…….. 99
6.5. Result discussion……. 103
6.6. Precision, Recall and F measure…….. 108
6.7 Case study……. 109
6.8. Summary…….. 116
Chapter 7: Conclusions and Future Work…….……..… 117
References…….……..……... 121
Appendices……..... 125
Appendix A: The Survey hardcopy……..…….. 125
Appendix B: The survey results …….……..... 129
Appendix C: The survey results example in charts for the computer science specialization……..… 131
Appendix D: Specializations of Computer domain in Palestinian Universities in the West Bank……..... 135
Appendix E: ISO 9001 – The Job Description……. 137
.....الملخص بالعربية 138

List of Table

Table 3.1: The differentness between some e-recruitment sites…..….. 24

Table 3.2: The Definition of recommendation algorithm……..… 33

Table 3.3: The Comparison between recommendation algorithms…... 34

Table 3.4: The differences between Fuzzy sets and Crisp sets ……. 36

Table 4. 1: Bag of words from bable.net for position skills …… 52

Table 4.2: the words appearance and repetition in the two reviews 56

Table 4.3: the calculation of IF for User skills, and the position skills 57

Table 4.4: IDF values for each word. 58

Table 4.5: The TF-IDF values 59

Table 6. 1: The population distributing……. 86

Table 6. 2: The sample size for each university ……. 88

Table 6. 3: The sample size for each group……. 88

Table 6.4: calculate the $\bar{X}$ and S…… 89

Table 6.5: the survey results for a programmer job……... 95

Table 6.6: the survey results for a Computer science specialization…. 95

Table 6.7: The results of matching specializations in job application with candidates CV…… 102

Table 6.8: the of candidate list for programmer job according to the education percentage……. 103

Table 6.9: Another specializations not in advertisement, and have match ratio within 50%..... 105

Table 6.10: The sections percentages……..… 106

Table 6.11: The match percentages results and the final match ratio… 106

Table 6.12: bag of words for User experience……. 110

Table 6.13: bag of words for position experience…… 111

List of Figures

Figure 3.1: Fuzzy logic operations…….. 37
Figure 3.2: Fuzzy logic architecture…….……... 37
Figure 3.3: Fuzzification……...……...……... 38
Figure 3.4: Membership values…….…….……. 39
Figure 4.1: The proposed system architecture……...……. 49
Figure 4.2: The similarity and cosine relationship …...……. 55
Figure 4.3: the candidate CV……... 61
Figure 4.4: the job post……..... 62
Figure 4.5: the values of fuzzy sets……. 64
Figure 5.1: The structured CV template…….. 71
Figure 5.2 Figure 5.2: The structured job template…..…… 72
Figure 5.3. CV (Applicant) flow…….… 74
Figure 5.4. Employer flow (Job flow)……. 77
Figure 6.1: The job application…….……. 83
Figure 6.2: Expert knowledge base generation……...…. 91
Figure 6.3: ER diagram for knowledge database……... 92
Figure 6.4.: The proposed system home page……..… 93
Figure 6.5: CV structure form……..……... 94
Figure 6.6: The list of jobs according to the user CV…….… 96
Figure 6.7: Job application……..... 97
Figure 6.8: The manage jobs operations……..…….....**.** 97
Figure 6.9: the candidate list and the evaluation rate……... 98
Figure 6.10: The job application…...…….. 106
Figure 6.11: Candidate CV……...……..……..….. 109
Figure 6.12: The job application…...…….. 110

List of Abbreviations

CF	Collaborative Filtering
CBF	Content-Based Filtering
HR	Human Recourses
TF	Terms' Frequency
IDF	Inverse Document Frequency
COG	Center Of Gravity
FAHP	Fuzzy Analytic Hierarchy Process
MCDM	Multi-Criteria Decision-Making
FES	Fuzzy Expert System
NLP	Natural Language Processing
BOW	Bag of words

Chapter 1:
Introduction

1.1. Introduction

Previous research has focused on how to help job seekers find their suitable jobs through their CVs (publishing, uploading and updating) using techniques such as analysis, Information Extraction, organization, web crawling, etc. [1] [4] [5], With no consideration if the recommended candidate achieve the employer's goals or needs or even the appropriate of specialization for the job, according to university literatures.

Getting a suitable job on an accurate scientific basis is a challenge to the recruitment systems and sites deployed on the Internet. There are multiple researches deal with the recruitment process in terms of techniques used without looking at it as an integrated and interdependent model go through several stages to extract the desired results.

1.2. Problem Statement

There is a gap between education and the labor market. The job market needs of jobs are not met by the specializations offered in universities, but there are also many similar specializations in terms of title and studying plan in universities without looking at job opportunities, and linking them with the education outcomes of these specializations.

Restriction of the employer role in the employment process is only to receive the results of the employment announcement, without express the real needs in qualifications employers

wants to find in their future employee, to work in the best possible way and achieve the best results.

Also, there is a limited interaction between the companies or the employers and the applicants. Most of the current recruitment systems display recruitment information to website viewers, without any matching or ranking for job suitability for the applicants' CVs. Thus, the main question of this work is how to achieve a best matching between candidates and offered jobs. So, find the best candidate to the best job.

In addition, in our country, the political and economic situation directly affects the recruitment process, and it is difficult for the employer to find the ideal employee to work a lot. Therefore, employers waive some sections in the job application or prefer one section to another for a good employee. In addition, there are two types of education: vocational technical education and traditional education.

All these factors increase the needs of knowing the opinion of the employer about the sections in the job application, and the impact of his opinion on the results of the recruitment process in the list of candidates most appropriate, if we take the opinion of the employer in the same job, do we get different results?

1.3. Hypothesis

The main hypothesis that will be assumed to achieve the main object of this research study, is that there is a relationship between experts (Lecturers) view about specializations in universities and the success of the employment process.

Usually, a job advertise contains many specializations which apply for the same job, with no consideration for the differences between these specializations in terms of study plan, courses and Credit Hours, therefore there are differences between applicants and disparity

in the proportion of their suitability for the job on the basis of the difference between specializations in the same domain.

Therefore, our system classifies and arranges applicants for a job according to employer needs and according to the learning outcomes of different specializations. Measure the impact of the employer's opinion on the terms of the job application according to the following hypothesis:

- H1: The employer's view for the job application should be reflected in the recruitment process.
- H2: The employer's view for the job application should be reflected in the applicants' recommendation list.
- H3: The employer's view for the job application should be reflected on the employment application and employee qualifications structure.
- H4: Weighting the applicants' CVs based on the job needs will enhance the employment process results.
- H5: The adjacency of the structured CV with the structured job should facilitate the ranking and matching processes.

1.4. Our Contributions

We suggest a new technology represented in a hybrid system based on BabelNet dictionary, content base algorithm and fuzzy logic in order to obtain the best possible accuracy in analyzing, and evaluating the recruitment process (CVs and job applications). Using these algorithms will increase the efficiency and the accuracy of our proposed system, according to their efficiency dealing with documents, terms and text analyzing. Also calculating similarities will be according to the CV and application contents not by user behavior. Also,

the system could make decision in matching and ranking the results by using mamdani fuzzy logic, which uses synthesizing a set of linguistic control rules obtained from experienced human operators. Also, we generate an expert knowledge database, which contains the opinion of lecturers in universities in different specializations in the same domain, and its relation with the job position in the labor market, and the evaluation of the employers for the employee specialization after hiring. This process is the core part of our system; it evaluates the education section in candidates' CV template then categorizes them under certain conditions determined by the employer in order to accurately identify their needs in job application.

The study aims to measure the suitability of different specializations in the same domain to a specific job by referring to the expert knowledge database, which based on the opinion of lecturers in Palestine universities, by focusing on the relationship between the labor market and specialization. Also, study the impact of the employer's opinion on the recommended candidate lists for the job by measuring the important of each section or part (education, skills, and experiences) in the job application, according to the employer's offer and then calculate the similarity between these parts and the same parts in the curriculum vitae.

In this research, we propose a solution to the recruitment process by involving universities in the recruitment process by identifying suitable jobs for different disciplines, and linking education outputs to job opportunities in the market.

Also, we propose an interactive system, between both universities and the labor market in order to participate in determining the most appropriate job for the various disciplines, so that employers after hiring can provide feedback about the employee and the suitability of the job according to his specialty.

In short, the contributions of this research are:

- Reflecting the domain experts and employers' feedback on the hiring process.
- Build a system (web application) that applies Content-based filtering approach CBF, which will allow the employers to determine their needs accurately, while posting a job by specifying the job need with a weight for each need, each category will have a percentage according to its importance.
- Provide applicants' recommendations list for the employers with the percentage that expresses the suitability for each applicant and the job.
- Provide jobs recommendation list for the candidates that shows the jobs with the percentage that expresses the suitability for each job according to the applicant's CV.
- Candidates can check the categories in the job app separately, it helps them to find

1.5. Thesis Structure

In addition to the introduction chapter, this thesis includes six other chapters, Chapter 2: (Literature Review) It discusses and analyzes the previous research on the job recommendation systems and the Matching techniques of Job Offers and Candidate Profiles. This chapter mainly reviews job recommendation systems and its variants and explains the main differences among them and their relevancy to work presented in this thesis document.

Chapter 3 (Background) contains an introduction in e-recruitment, recommendation systems and fuzzy logic and lists their main characteristics and challenges. It concludes examples of some best Job Sites with their classification. Chapter 4: (recommendation model), it is describing the theoretical model; Architecture, terms, algorithms, justification of algorithms use, explanation of efficiency.

Chapter 5: (implementation), it describes the system implementation such as components, technology, input, output, etc. Chapter 6: (Experiment and Analysis), in this chapter we

experiment the system and Analysis the result, we introduce a case study, and it include the data collection process.

Chapter 7 (Conclusions and Future Work) Based on the results of the comparative study, this chapter concludes the main findings derived from this research. It concludes with a list of some future work research lines.

Chapter 2:

Literature review

2.1. Introduction

Over the past years, many research papers have been presented new ways to improve the performance of recruitment systems. In this chapter, we will review some recruitment systems and their methodologies discuss some of their performance improvements and weaknesses points.

2.2. Automatic Match Job Offers and Candidate Profiles:

The main research question in [10]is on what basis could more realistic technology, be built into employment matching systems. The study seeks to create a system that simulates human resource systems in selecting candidates for jobs, by considering that the human resources staffs are experts in the field of ranking and selecting the most suitable candidates for the available jobs. The idea of the study is to adopt a model that is capable of repeating the behavior of human experts when classifying candidates for jobs trying to predict new job ratings. So, these automatically generated ratings can completely repeat the rating made by human experts, the idea appears to solve several problems that human resources (HR) suffer, such as dealing with large number of job applications, failure to deal fairly with all job candidates. The expert knowledge is not always accurate in determining the top positions, and the current methods of matching profiles and job offers, depending on some kinds of grammatical matching.

The study suggests approach based on machine learning and some classifications that

represent the primary source of knowledge. These classifications are DISCO1, ISCO2 and ISCED3. The approach tested by using data from real employment scenarios; they work with samples that obtained from four main areas: information technology, law, logistics, and marketing, and focus on three types of elements: skills, education and languages. From the user file to compare with job offers, then compare the results of the proposed system with OKAPI BM25 algorithm, which they consider it as the primary way to compare a new approach to this problem because it represents as the latest software solutions based on technologies retrieve traditional information and the most accurate method of calculating using a bag-of-words model.

This approach should be able to calculate the cost of converting a profile into a job offer by:

1. Determine the basic distance between the two groups, such as the skills in job offer and applicant profile: they define the distance between them as the minimum number of basic operations required to convert the applicant's profile into a job offer. These basic operations are three: insertion, deletion or alternatives.
2. Replacement costs: by creating appropriate knowledge bases that cover relevant terms from various fields in employment scenarios.
3. Insertion and deletion costs: by considering the minimal cost of insertions and deletions that could transform the profile into the given job offer.

After all the partial costs (replacement costs, insertion and deletion costs) for each of the different categories have been collected, the study combines all of them following an aggregation strategy consists of calculating the transformation cost.

For ranking the approach follow is: the higher the overall transformation cost the worse the position of the candidate in the ranking and vice versa, the lower the transformation cost

the better the position of the candidate in the final ranking.

Finally, the results that achieved can be summarized as follow the following:

1. Improving the traditional matching, and Reduction of the efforts of cost and time to find appropriate links between job offers and applicant profiles.
2. Reduce the need of professionals in specific knowledge concerning a professional field or skill from the HR.

The weaknesses in the study are:

1. Knowledge from experts is not always consistent.
2. Higher number of candidates to rank cause to be more difficult to be completely accurate.
3. Knowledge bases covering the relevant terminology from the different thematic areas in recruitment scenarios do not exist.

2.3. An Effective Job Recruitment System Using Content-based Filtering

In [21]; the main challenge is “to improve the quality of job recommendation. Authors suggest a user model (Content-Based Filtering) and social interaction (Collaborative Filtering) improve the goal of this project, which attempts to select the right graduates based on industry needs, and will focus on reducing limitations of existing job portals, and provides better and efficient job recommendation. The main aim of this portal is to connect to the industries and acts as online recruitment, to support the students to find the right job after graduation.

The study explains the idea of using social interaction (Collaborative Filtering), one of drawbacks that traditional collaborative filtering algorithm faces, which is high sparse data.

This makes the precision and quality of the recommendation algorithm unsatisfied. And because of the fast development of social networks. It became possible to selectively fill the missing value in the (user-item matrix) by using the friendship or trust relationship information of social networks.

The key problem is that most of the job-hunting websites just display recruiting information to job applicants. Job Seeker has to retrieve among all the information to find jobs they want to apply. The whole procedure is tedious and inefficient, and there are many drawbacks of existing job systems as dealing with graphics environment, content insufficient, time response and reliability.

The proposed system is Web-Based Applications, which allows applicants and employers to register their details. Applicants can browse through the vacancy details that are posted and can apply for the job online. Employers can browse through the posted resume and select suitable candidates; also, job seekers can update their CVs daily, sending resume and save a job according to their needs.

The Limitation of the proposed system is scalability and sparsely.

The Advantages of the new job portal are as follows:

1. Achieve the main targets of the project
2. Standard content services and display
3. High-level management and flexibility.

2.4. Job Recommendation System Using Profile Matching and Web-Crawling

The proposed system in [4] is a job recommendation system for students and campus in universities. The developed system is solution for problems the current campus recruiting systems faced, like the weak result, as lower matching degree, recruiting process take a

long time, and higher recruiting cost.

The study briefly telling that the problem is caused because of three points, they are:

1. Problem in the requirement descriptions, which are not clear and definite.
2. Students may confuse and misunderstanding for the requirement description and apply for inappropriate positions.
3. The students under greater employment pressure always looking for a new job, this would increase the cost of candidate selection.

The developed system aims to help the placement office in universities, to match between company's profiles and student's profiles with higher precision and lower cost. The system used different matching methods for profile matching which are: semantic matching, tree-based knowledge matching, and query matching.

These methods integrated according to representations of attributes of students, and companies to find the profile similarity degree, each student in the university has different education levels and skills based on student's cv and background details, each student applies for a job he expect that he/she get only these job recommendations or he/she will get highly relevant recommendation.

The system has two phases, first phase in this system based on the profile similarity degree, preference lists of companies and students are generated. Second phase: if two students have similarity in profiles. Job recommendation system defined a solution called job taste, which is the preference criterion considered before applying for a particular job. The job taste (the job that candidate prefer) can be getting a job in a highly ranked company, or offers higher payment. So, the recommendations are provided to the respective customer according to his job taste or preferences.

Also, students can perform a keyword-based search for job profiles from various job recruitment sites. For obtaining data from online recruitment sites system uses web crawling. With loop matching, matching results would be further optimized and provide more effective guidance for a recommendation.

Through the proposed system the students can search for a job profile job recruitment sites by using a keyword-based search, and received data from online job sites by using web crawling. Students get all the data in one central place. They can access this information from anywhere with proper logging.

As a result, the system improved the existing campus recruitment recommendation system. This system has improved matching results further optimized, and provided more effective and relevant recommendation for the students, where they search for a job using keyword-based search or web crawling method. Also, the system lesser chances of data loss, and the required time for maintaining all the data will be reduced with this system because of the maintained will be automatically, so they save time.

2.5. Implementation of an Automated Job Recommendation System Based on Candidate Profiles

In [27], a recommendation system for online job-hunting is designed based on a collaborative filtering approach. The system used CF with its two approaches, user-based and item-based to choose a the better performed one. In addition, the study explained the reason for conducting the new system that most jobseekers are new graduates from the undergraduate and graduate levels, and most recruitment sites on the Internet offer jobs as information only. Site users view them and they must retrieve the necessary information

about jobs that suit them.

The study tested the user-based and item-based CF algorithms on the data set respectively using three methods of similarity calculation, which are Log-likelihood, city Block and Tanimoto are. It compared the similarities result from the three methods of similarity calculation and find that using Log-likelihood with the item-based algorithm is the more accurate and the best performance. The result shows that the item-based algorithm is more the better performed than user-based, because finding similarity between users is more difficult, and the attributes of items do not change easily it is a stable preference. Also, item-based CF Deals with Boolean Data (1,0).

A recommendation system for online job-hunting applied using item-based with Log-likelihood method to compute candidate items' similarities in the Student Job Hunting recommendation system. The system used data of students' resumes and details of recruiting information. Also, the weights of the users who had applied for the jobs and called them co-apply users, and weights of student used-liked jobs to recommend jobs for users. To experiments and implement the proposed system the study used Apache Mahout. For each student the system recommended just 2 to 3 jobs, because the data they used is sparsely. Mahout just focuses on users who have recommended items, while automatically ignores users that cannot be recommended. So, the result of evaluation seemed very good. The proposed system limitation was cold start, scalability and sparsely, which is the CF disadvantages.

Finally, to improve the result, the proposed system used real data, and they are more relevant to users' preferences.

Both next studies [12] [28] share the motives for proposing an e-recruitment system, which is matching user's file with job profile to choose the most suitable and appropriate employee for the job within certain criteria. The two studies suggest the fuzzy logic as solution.

2.6. Fuzzy AHP approach for employee recruitment

The study [12] focuses on the importance of choosing the appropriate employee for the vacant position, because of its impact on the employee's relationship with the institution who belong to, such as dedication to work and commitment to the laws in force and developing the institution. If the employee was chosen with integrity and transparency, and this selection process is considered as one of the most challenges facing human resource management.

The study considers choosing the ideal employee for the suitable job not easy, but a complicated process. This is due to the great diversity in their skills, and this reflects the difficulty in choosing the employee and moving as far as possible from personal judgments in the recruitment process.

The study proposes developing a computer application and using the Fuzzy Analytic Hierarchy Process (FAHP) techniques as one of the most popular multi-criteria decision-making techniques (MCDM). This solution is presented to the Human Resource Management to assist them in making employment decisions and evaluating and arranging job applicants, by transferring Language and verbal expressions of numeric values.

The AHP is a powerful and flexible decision-making process to assist people set priorities, and make the best decision when both qualitative and quantitative aspects of a decision need to be considered. FAHP method is a mathematical approach for alternative selection and justification problem by using the concepts of fuzzy set theory and hierarchical structure analysis. The decision maker can specify preferences in the form of natural language, or numerical value about the importance of each performance attribute. The system combines these preferences using FAHP with existing data. In the FAHP method, the pair-wise comparisons in the judgment matrix are fuzzy numbers and use fuzzy arithmetic and fuzzy aggregation operators; the procedure calculates a sequence of weight vectors that will be used to choose main attribute.

The algorithm steps for evaluating personnel selection problem by FAHP are:

1. Step 1: The first step of FAHP consists of developing a hierarchical structure of the assessment problem. After developing the performance hierarchy, decision makers have to determine the relative weights of each criterion. In the AHP, weights are determined using pair-wise comparison between each pair of criteria. To determine relative weights, decision makers are asked to make pair-wise comparison using a 1-9 preference scale. The pair wise comparison data is organized in the form of fuzzy triangle numbers.
2. Step 2: If decision makers cannot utilize the preferences by the form of fuzzy triangle numbers, they can give preferences by linguistic terms, and use look-up tables for values, they can easily derive corresponding value of fuzzy numbers
3. Step 3: After setting up the hierarchy and pair-wise comparisons of criteria of alternatives, it is necessary to calculate global value of priority of alternatives.

To calculate the steps of FAHP, first the outlines of the extent analysis method on fuzzy AHP are given and then the method is applied for a catering firm selection problem. Let X = $\{x_1, x_2, \ldots, x_n\}$ be an object set, and U = $\{u_1, u_2, \ldots, u_n\}$ be a goal set. According to Chang (1996) in extent analysis, each object is taken and extent analysis for each goal g_i, is performed, respectively. Therefore, m extent analysis values for each object can be obtained, with the following signs:

- Step 1: The value of fuzzy synthetic extent with respect to the ith object.
- Step 2: The degree of possibility of $M_2 = (l2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ is defined as V ($M_2 \geq M_1$).
- Step 3: The degree possibility for a convex fuzzy number to be greater than k convex fuzzy numbers M_i (i=1, 2, …, k).
- Step 4: The normalized weight vectors are W' = $(d(A_1), d(A_2), \ldots, d(A_n))^T$ where W is a non-fuzzy number.

2.7. A Fuzzy Expert System Tool for On-line Personnel Recruitment

The key problem in [13] about the tasks of many job online recruitment portals, are just matching the qualifications of applicants with the requirements of employers without ranking them according to their relative suitability for the jobs applied to. The study proposed a Fuzzy Expert System (FES).The system works as the human intelligence to handle instances of approximate reasoning, and fuzziness that can be introduced into computer systems using the mathematical concept of fuzzy logic. The proposed system aims to minimize the strictness and subjectivity while choosing the right employee for the job that suits him/her, and taking initiating intelligent decision making in the ranking and evaluation of the credentials of job candidates.

The architecture of the FES tool consists of following:

1. System input: this consists of the candidates' profile database, and the job vacancy database. The candidates' profile database contains information that pertains to individual candidates as obtained from the candidate profile form filled by each candidate online. The job vacancy database contains the specific employment requirements which a prospective job seeker must fulfill in order to qualify for consideration for the available jobs.
2. The fuzzy inference system: This is the fuzzy logic engine of the FES which consists of:
    - fuzzifier component: This handles the conversion of input values obtained from candidates' profile into fuzzy values within requirements parameter fuzzy set.
    - fuzzy rule base: This contains the set of fuzzy IF-THEN rules that embodies the knowledge used by the fuzzy inference system. Each of the rules has an antecedent (IF) and a conclusion (THEN) part that prescribes what should be done when certain conditions are true. The specific criteria for selection for each specific job position are represented as IF-THEN rules which determine the eligibility of candidates.
    - Fuzzy inference engine: This implements the fuzzy reasoning of the FES by combining the fuzzified inputs with the fuzzy rules using the mamdani-style fuzzy inference process.
    - Candidate ranking components: This computationally evaluates the closeness of each candidate's fuzzified credentials to the ideal requirements for a specific job using the fuzzy hamming distance function. This fuzzy distance metric is used to rank candidates' profile in the order of their eligibility for the job.

Finally, the validation of the proposed system by determining whether the system's outcome is consistent with the conclusions of the human experts, validation focuses on evaluating the outcomes rather than the process, by which the outcomes are determined.

The study let ten human experts from Akintola Williams's group to obtain a copy of the software to be evaluated, and to perform a simple benchmark problem on the software. Based on past experience, the human expert answers a set of 14 questions about the software. The questions are quantitative and are based on a 0 (very false) to 5 (very true) numerical scales, the rating (0 to 5) of their assessment of the performance of the FES tool covering seven aspects. The evaluation undertaken yielded satisfactory result to confirm the viability of a FES approach to enhancing the existing capabilities of many online recruitment systems.

2.8. Summary

In this chapter, we discussed the Related Work to discover the gaps in e-recruitments systems and to work on them in our research.

Most studies in e-recruitments systems seek to repeat the behavior of HR employees in employment, which is not always consistent, also it seems difficult to rank a high number of candidates to be completely accurate. Thus, the confused and misunderstanding of requirement descriptions and extracted needed data from user profile and apply for inappropriate positions are also problems in e-recruitments.

According to our review, we chose to work on a hybrid system based on the babel dictionary, content base algorithm, and the fuzzy logic to obtain the best possible accuracy in CVs, and job applications analyses with limited employment for expert knowledge base for evaluating education sections in CV template, and efficient final result as a ranked recommendation lists.

The proposed system helps in understanding employer needs, and determining the most appropriate specialization to occupy jobs, especially among those similar disciplines degree by creating an expert knowledge database, also we use BabelNet to find synonym words for user profile and job offers. We used content base filtering to calculate the similarity between user profile and jobs offers, and to calculate an efficient result we used fuzzy logic.

Chapter 3

Background

3.1. Introduction

Over the past few decades, many new terminologies have appeared in the internet world, especially after the popularity that some websites like YouTube, Amazon and Netflix have enjoyed. One of these terms is recommendation systems and e-employment or e-recruitment.

There are a various understandings of online recruitment system, there are five conceptualization of online recruitment online recruitment as a technology tool, online recruitment as a process, online recruitment as a system, online recruitment as a service and online recruitment as a proxy all that make it hard to find an exact definition of e-recruitment, because each of these conceptualization has its own definition, it's good to mention that most studies relay on mixed definition between two or more conceptualization. In my study I used some of this conceptualization fist e-recruitment as a technology because our study presented tools to automate the ranking of applicants in recruitment. Second-recruitment as system, because it is containing independent elements which are related to each other. Third e-recruitment as a service because we consider it is offering a service for both job seekers and job owners.

e-recruitment system has many advantages that help it to be easy to used, we can consider it as a good tool to track applicants and hiring process, make it easy to examine database records, match appropriate candidates with available jobs, prepare database records for all

applicants and keep a historical records for them and be in touch with each applicants for new jobs opining.

The online job sites have witnessed a revolution in employing business owners and job seekers alike, and they have been able to increase their efficiency to the point that they have contributed effectively to making the appointment decision.

Despite some disadvantages of electronic recruitment, it is considered a revolution in the world of employment, as it saves a lot of effort, time and money as well and it is considered effective and serves both job seekers and companies' owners as well.

3.2. Job Board and Search Engine

There are two ways used in job websites to search for a job: a Job Board and Search Engine. A job board is a website that posts jobs supplied directly by employers, from different companies, whereas job search engines scour the web and aggregate job listings from job boards and employer websites from all over the internet [2].

The job search engines typically list more jobs than job boards. So, using a job search engine can save you lots of time, since you do not have to search for jobs on multiple websites.

With job boards, companies have specifically listed their open positions and often accept job applications directly through the job board. Employers typically pay a fee to the job board to list their jobs on the site.

The largest and best-known job board is Monster [17], which is a general board with postings across a broad range of industries. Also Simply Hired is an example of Search Engines and it is the most popular job search engines, and they collect millions of listings on their platforms, but indeed is both a job search engine and a job board.

3.3. Best Job Sites

There are a large number of recruitment sites in the world, but there are some sites that have won user confidence for several advantages, including the easy user interface, accurate results, and lists of available recent jobs.

We review the most important employment sites, and the most famous of them, depending on whether they use the search engine or job board; they are in order from the best as follows:

1. CareerBuilder: CareerBuilder is one of the largest job boards, providing job local listings directly from employers and from newspapers, resume posting, and career advice and resources to job seekers, also career builder providing useful information, and advice about careers. The site uses the latest AI technology. Your Resume can also upload, and the site will show you job listings that match your experience. You can upload your resume from a file, Drop box, Google Drive, and any .DOC; dot.DOCX; PDF; RDF; TXT; ODT, or WPS up to 1000kb. You can search CareerBuilder by city, state, ZIP code, job title, skills, or military occupational code, as well as browse by job title, job category, or state [17].

2. Dice: is one of the larger job boards and lists mostly technical positions for tech job seekers. Dice focus on all types of technical openings. You can search by company, job title, keyword, employment type, and location. Registered users can upload a resume; get salary information, store resumes and cover letters. Career advice and news relevant to tech job seekers is also provided on the site.

3. Glass door: is a career community that helps people find jobs and companies that recruit top talent. It is an industry leader on salary information by company. Glass

door members can see the latest job listings and reviews, interview questions, and company details as contains company reviews by former and current employees, ratings, company information, salaries, CEO approval ratings, competitors, content providers, and other company details. In glass door you can search by job type, title, companies, salaries, interviews, keywords, experience, and location, and then upload a resume to apply. Also, it possible to post jobs that you know about.

4. Google for Jobs: it is a product from Google, it is a job search engine that compiles listings from many different sources, including other job search engines, and you can search a job by typing into Google search bar. Also, you can search by type of job, location, company type, date posted, and more. Also, check Salary Information is available with Google for jobs.

5. Idealist: specialized in the full-time, internship, and volunteer positions within the non-profit sector for example organizations homelessness or the environment. Idealist allows users to identify groups of interest and communicate with other members for networking purposes. The site consists of over 120,000 nonprofit and community organizations in over 180 countries. Table 2.1 summarizes the difference between some e-recruitment sites [17].

Table 3.1: The differentness between some e-recruitment sites

Job website	Search engine /job board	Used for	Best for
CareerBuilder	job boards	Focused more on people with a bachelor's degree.	Company Profiles
Dice	job boards	technical positions	Tech and Engineering Jobs
Glassdoor	job boards	primarily known for providing anonymous company reviews before apply job	company reviews
Google for Jobs	job search engine	pulls jobs from many different sources	Search for jobs
Idealist	job boards	full-time, internship, and volunteer positions	Volunteers job
Indeed	job search engine and job boards	find almost every type and level of job	Number of Listings
LinkedIn	job search engine	Move job search from online to the real world	Networking Tools

3.4. Advantages and Disadvantages of e-recruitment

3.4.1. Advantage of e-recruitment:

1. E-employment is less expensive than traditional recruitment, as placing your ad on social media or on a recruitment site will not cost you anything compared to placing a paid ad.
2. It is characterized by a fast response, as the response to a job posting may come soon after it is placed.
3. It saves a lot of time and effort, as the job seeker can send his CV via e-mail and if it is suitable for the job, he is contacted by the company to determine the date of the personal interview without the need to go to the company's headquarters.
4. Provides you and your company an outreach, which is considered a free advertisement about the nature of your company's work and services.
5. It provides you with many applicants for your company, which means various

experiences, qualifications and skills as you can choose the best for your company [8] [24].

3.4.2. Disadvantages of online recruitment:

1. There is no interaction between the candidate for the job and the company, and it is often difficult for the applicant to contact one of the company representatives to inquire about something.
2. Often, the CVs contain incorrect information.
3. Someone may use your personal information on your CV. Bad use if confidentiality and security are not available [8] [24].

3.5. Recommendation system

There is a wide category of web applications that include predicting user responses to different options, such as products and goods, services, movies, and series.

Recommendation systems are simple algorithms that aim to provide the most relevant and accurate elements to users, so that it filters the information that is entrusted to the customer from everywhere by filtering it dynamically according to the preferences, interests and behavior of the customer, the recommendation system is betting on the ability to predict whether a specific user prefers a product or service or not, depending on the user's profile and behavior in the past [16].

Recommendation systems consist of a set of algorithms that are commonly used to recommend products or services to users, based on previous user information. Recommendation systems have become an essential and important component of many online stores, movie and series bases and job search sites.

The recommender system approaches are classified into four categories which are: Collaborative filtering CF, Content-based filtering, Knowledge-based and Hybrid approaches [20].

The first approach is the collaborative filtering, which predicts the interest of the user to an item by using a database of a group of other user's preferences, This approach is itself divided into two subcategories: The Memory-based collaborative filtering which predicts the interests of the user by assigning him first to a group of similar users, through similarity or correlation measures, then it uses the weighted-notations of the same-group users regarding the items; Model-based collaborative filtering on its side uses the predicted values of a user's notation regarding an item, based on the knowledge that the system has about the user [20].

The second approach is content-based filtering, which focuses on the content similarity between an item and the other items that the user has previously liked. Systems that are based on this approach have a two-steps process: the user's profiling and the items' representation [15].

The third approach is the knowledge-based approach which suggests to the user, items based on the inference of his needs and preferences through the construction of a strong knowledge of the field [21].

Finally, the Hybrid approach is the combination of the three previous approaches by using different technologies [16] [20].

3.5.1. Collaborative filtering approach:

Collaborative filtering CF is based on collecting and analyzing a large amount of information in users' behavior, such as expressing their opinions about the products or activities they are doing and their preferences, then anticipating what the user will like based on the similarity of his activity and preferences with other users. It is based on the assumption that users who previously agreed to their tastes will have similar tastes in the future, and will like the same things that others liked before.

One of the advantages of the cooperative filtering method is that it does not depend on an analysis of the contents of the activity or preference, and in this way it can provide complex recommendations for the activities and preferences of goods and products without understanding the content of these activities or goods such as the recommendations of films without the need to understand the content of this film [14].

A- Memory-Based Collaborative filtering CF:

This method relies on building groups of users who have the same opinions, and choices in a particular topic, assuming that they will tend to get the same opinion in other areas.
Groups are created for users who share the same interests as each user is considered part of the group. Thus, user predictions about new elements can be produced according to the preferences and interests of users in the same group.

Memory-based CF also divided into:

- User - User: The user is the center of this method, where the profiles of the most similar users are determined based on their interaction with the elements. In this method the extent of similarity between the users and their position as neighbors are

evaluated, in order to suggest the most common elements among these neighbors (the most similar are the nearest neighbors), these recommendations are new elements for the user.

- Item – Item: This method is to find elements that are similar to the ones the user has already "positively" interacted with. In this method, the similarity between the elements that users interact with is evaluated in the same manner, and a new recommendation is presented to a user based on the users ’interactions with the elements.

B- Model Base Collaborative filtering CF:

Model-based CF methods, is a method in which a model is produced from the historical rating, and used to deduce the predictions Model based. it only rely on user-item interactions information and assume a latent model supposed to explain these interactions.

The main problems a Collaborative filtering system faces [20]:

1. Cold start: It is for the new user who has not yet classified or selected any item and his profile is almost empty, so it is difficult to make any recommendation to the user because his preferences and taste are unknown to the system.
2. Trust: When recommending it is preferred to use personal files for users who have rich data and many classifications compared to people who have personal files with a short experience. The question of faith arises towards the evaluations of a particular customer with a great and rich experience.
3. Scalability: The system should be able to accommodate the significant increase in the number of users and the various elements that are interacting with it, to form

appropriate recommendations for users and therefore the system needs more resources to process information and use it to identify the similar tastes of users, and goods with similar descriptions and preferences

4. Sparse: It is difficult to determine the interests and tastes of the users who evaluate a specific number of elements and their interaction is limited, so there is little information available about them and the recommendation process.
5. Synonym: synonymy the difficulty of finding differences between closely related elements, and therefore the inability of the recommendation system to calculate similarities between elements using different similarity measures.
6. Privacy: To obtain an accurate and highly effective recommendation, the recommendation systems collect information from user files and study their tastes and interests, and this requires protection of user data and its privacy through specialized programs in information security and reliability.

3.5.2. Content base filtering approach:

In content-based filtering recommendations, the system makes a recommendation to the user of a specific item (product or content) based on two principles: the item description (the properties of the item, such as: a product brand, product color, product size, etc.) and the user profile that contains the user's previous options. This recommendation is made with specific techniques such as TF-IDF (Terms' Frequency-Inverse Document Frequency) [25]. This technique is used in big data and machine learning based on frequency of terms in the documents and measures how important a term is. These techniques used to compare the profile with the elements, determine the similarity between them to make the recommendation and determine the most appropriate element for a user. These systems are

used to recommend various web pages and TV programs such as movies, series and more.

Content-based recommendation systems create user profiles through user-provided notes and data, and the elements that they prefer (item description) by clicking, or by rating, or by searching for it which means that they have an interest in the item and a great ability to purchase the product or content is attested. To determine the new elements that the system will recommend, and here the content-based recommendation systems have solved the cold start problem experienced by the collaborative filtering approach systems, and the user later provides more information or inputs such as accepting or rejecting the recommendations, because the user provides more inputs or takes actions regarding the new recommendations presented to him, therefore, the accuracy of the information in the user file increases, which also means accuracy in the results of the recommendations system [15].

Briefly content-based filtering CBF mechanism including the following steps [14]:

1. Examine the properties and features of the items, whether product or content, in preparation for a recommendation.
2. Compare feature properties with user interests and preferences.
3. Recommend the elements whose characteristics are consistent with the user's interests and needs.

A- Advantages of content-based filtering [6]:

1. Do not depend on other users, any need for other user data.
2. May recommend new and unpopular items with a unique taste.
3. Transparency: the result of the recommendation can be easily explained.
4. It provides user especial needs that other users not care about, and achieve in dependence from other users through the exclusive classifications that an active user uses to build his or her profile.

B- Disadvantages of content-based filtering [6]:

1. Hard to extract from sound, movies or photos.
2. CBF defends the same types of elements because it suffers from a problem of over specialization.
3. Limited content analysis: If the content does not contain enough information to accurately distinguish elements, the recommendation will not be ultimately accurate.
4. The user will never be recommended for different items.
5. Business cannot be expanded as the user does not try a different type of product.
6. If the user matrix or item matrix is changed the cosine similarity matrix needs to be calculated again.

3.5.3. Knowledge-based approach:

A Knowledge-Based Recommendation System focused on item features to meet the user needs, and on specific queries made by the user without looking for other users have the

same behaviors, and it does not need a large amount of data or depends on a user's rating history in his profile to generate recommendations.

A Knowledge-Based Recommendation System depends on the domain knowledge; it can generate recommendation for complex product such as expensive items like diamonds, and the ones which have few ratings.

Knowledge-based approach has many advantages such as solving cold start problem [21].

3.5.4. Hybrid filtering approach:

It is the combination of two recommendation systems, to obtain better results in generation recommends, and reduces the weaknesses that found in each recommendation system separately. For example, combine content-based filtering with collaborative recommendation [20].

Table 3.2 and Table 3.3 summarize details and descriptions of different techniques of recommendation algorithms.

Table 3.2: The Definition of recommendation algorithms

Recommendation techniques	Short description	Pros	Cons
	Collaborative filtering approach		
Memory based techniques • User-based technique	find other users whose past rating behavior is similar to that of the current user and use their ratings on other items to predict what the current user will like	• Work with complex objects • Domain knowledge not needed. • Quality improves over time	• Cold-start • Performance • Limited scalability • Data sparsely • Synonymy
• Item-based technique	predict preferences, uses similarities between the rating patterns of items		
Model-based techniques	Employs the previous ratings to learn a model in order to improve the performance of Collaborative filtering Technique.		
	Content-based filtering approach		
Content-based filtering	based methods, features of items are abstract and compared with a profile of the user's preference	• Domain knowledge not needed. • Quality improves over time. • recommend to user with unique tastes • Users get recommendations without sharing their profile, and his privacy. • Able to provide explanations.	• Ramp-up problem for new user. • Performance limited by the features that associated with recommended objects. • limited content analysis • Content overspecialization
	Knowledge-based approach		
Knowledge-based approach	Suggest objects based on inferences about user's needs and preferences. it assists users in the determination of suitable solutions from complex product and service assortments.	• No need to gather information about a particular product. • No ramp-up problem	• Need knowledge acquisition. • Knowledge engineering difficulties.
	Hybrid filtering approach		
Hybrid filtering	Combines different recommendation techniques in order to gain better system optimization to avoid some limitations and problems of pure recommendation systems.	• Increase the benefits of both techniques that are combined • Can implement serendipity, diversity • Often outperforms CF and CB alone	• Decrease the weaknesses of both techniques that are combined. • Can be a lot of work to get the right balance

Table 3.3: The Comparison between recommendation algorithms

Recommendation techniques	Input	Process	Similarity calculation	output
Collaborative filtering approach				
Collaborative	Rating from user u of item i.Depends only on usage data (e.g. rating, downloads, user preferences).	Try to Predict the user opinion (rating score) on different items to be able to recommend the best item to other users in neighborhood	Jaccard similarity Cos similarity. Centered Cos similarity. Mahout similarity algorithms: 1.Euclidean Distance Similarity 2.Log Likelihood Ratio Similarity 3.Pearson Correlation 4.Tanimoto Coefficient Similarity 6.Spearman Correlation	List of items user like it most(Top N recommendation)
Content-based filtering approach				
Content based	Feature of items. Rating of items from users. Depends only on the content/ description of the item and the user.	Generate a classifier based on user's ratings, and use it to classify new items.	Simple match Dice's coefficient Jaccard's coefficient Cosine coefficient Overlap coefficient	The most similar item as a list :Item id, similar item id ,similarity score The most similar item has the highest score.
Knowledge-based approach				
Knowledge based	User needs A description of user's needs or interests.	Match between items and user's needs. Users specify the requirements and system try to identify solutions.	Set base: Jaccard's Similarity Vector base: Cosine Similarity Asymmetric: sub Sumption Similarity Euclidean Nearest neighbor	a subset of items that satisfy the maximum set of weighted constraints
Hybrid filtering approach				
Hybrid	Uses and item content features well as usage data to benefit both types of data.	Separate implementation and joining the results. Utilize some rules of content-based filtering in collaborative approach. Utilize some rules of collaborative filtering in content-based approach. Create a unified recommender system that brings together both approaches.		

3.6. Fuzzy logic

Lutfi Zadeh, a professor in California in 1965, proposed fuzzy logic by noting that computer logic is unable to process data that represents human ideas or ideas that are not clear [9].

Fuzzy means ambiguity and lack of clarity, so the transition between different degrees of membership of the element corresponds to the confused boundaries of the disorganized groups. So the Fuzzy Set X contains all possible outcomes from interval 0 to 1.Thus, Fuzzy logic is a form of many-valued logic, where the truth value can be any real number between 0 and 1. On another words, it adopts the concept of partial truth, where the truth values ranges between absolute false and absolute true.

Fuzzy logic is used to mimic the way a person thinks when making certain decisions to solve a problem. A person searches all available data and thinks about all possibilities between numerical values 0 and 1, or between logical values true and false to solve the problem. Fuzzy logic is a pioneer in managing approximate information and decision-making when a variety of factors need to be taken into consideration [9].

3.6.1. Fuzzy Set:

A mathematical group consisting of elements that have a variable degree of membership is called a fuzzy set, so that an object can be a member of the group and have organic degrees, rather than only full and empty membership.

3.6.2. Membership function:

A fuzzy set X can be described by a membership function mx(x). A membership function defines to what extent a certain element (x) belongs to a (fuzzy) set (X). Thus,

membership functions can have any value between 0 and 1[3].

Crisp Set: also called a classic or clear group, so that all elements which belongs have identical properties. The classic (clear) group is usually defined as a set of $x \in X$ elements or objects that can be limited, countable, or uncountable. Each element can belong to or not belong to group A, $A \subseteq X$. In the first case, the phrase "x belongs to A" is correct, whereas in the latter case this statement is false. Clear combinations based on dual logic are used in computing and formal logic that includes solutions either in two forms such as "yes or no" and "true or false".

The crisp logic of representing knowledge does not provide a suitable method for interpreting inaccurate and non-categorical data, as its functions are based on first-class logic and classic theory of probabilities. In another way, it cannot handle the representation of human intelligence. The Crisp group can perform operations such as union, intersection, flexing, and variation. Characteristics presented in the interchangeably group include distributions, knowledge effectiveness, assistance, identity, transit and circumvention. Although disorganized groups have the same characteristics mentioned above. Table 2.4 summarizes the differences between fragile logic and confusion [3].

Table 3.4: The differences between fuzzy sets and crisp sets

Comparison	**Fuzzy**	**Crisp**
Set definition	Grade of memberships	Divided into members and non-members.
Set assigning	Mathematically	Linear Classification
Possible values	Range from 0 to 1	Crisp values (0 or 1)
Inferring Rules	Mathematically	Manually

3.6.3. Fuzzy logic operations:

The fuzzy logic has three operations which are: complement (negation, NOT), union (disjunction, OR) and Intersection (conjugation, AND. Figure 2.1summarizesand illustrates the fuzzy logic operations [3].

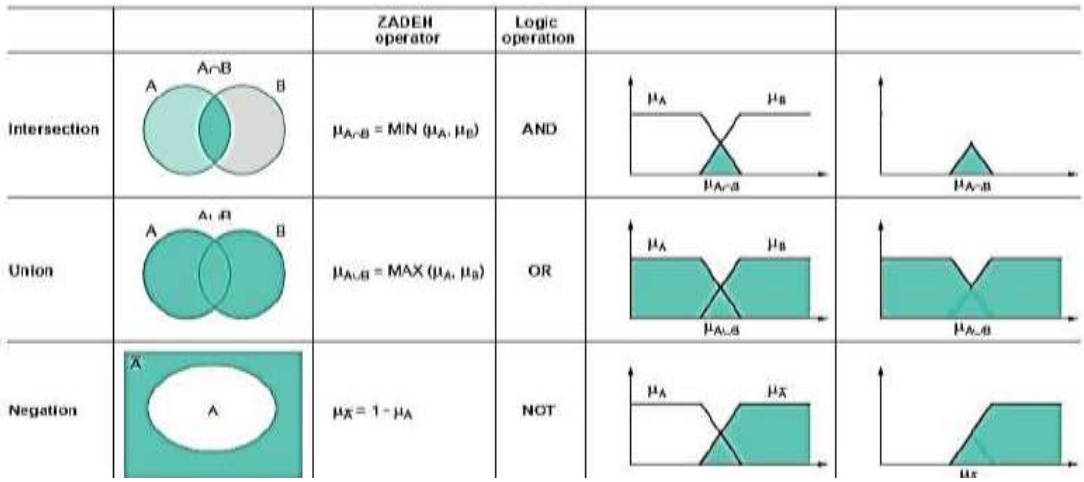


Figure 3.1: Fuzzy logic operations.

3.6.4. Fuzzy Logic Architecture:

Figure 3.2: [9] Illustrates the fuzzy logic architecture which is:

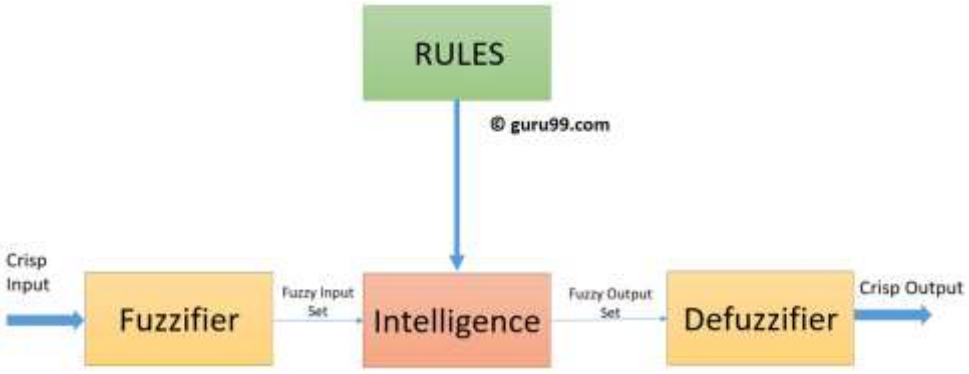


Figure 3.2: fuzzy logic architecture.

1. Fuzzification: first step is converting crisp number (the input) into fuzzy sets.
2. Inference Engine: It helps you to determine the degree of match between fuzzy input and the rules.

3. Defuzzification: finally, the defuzzification process is performed to convert of the fuzzy sets into a crisp value.

3.6.5. Fuzzy Inference System (FIS):

There are two main types of fuzzy inference systems: Mamdani and Sugeno.

A. Mamdani FIS:

Mamdani Inference System process contains three main phases which are fuzzification, rules and operators, and defuzzification [3].

1. Fuzzification: in this phase, fuzzy sets will be defined for inputs and outputs. Triangles will be used for defining the sets ranges with intersection between adjacent ranges; also, each set will be divided between 0 and 1 which is representing 100%. The figure below is illustrating how fuzzy sets are covering an input or an output range.

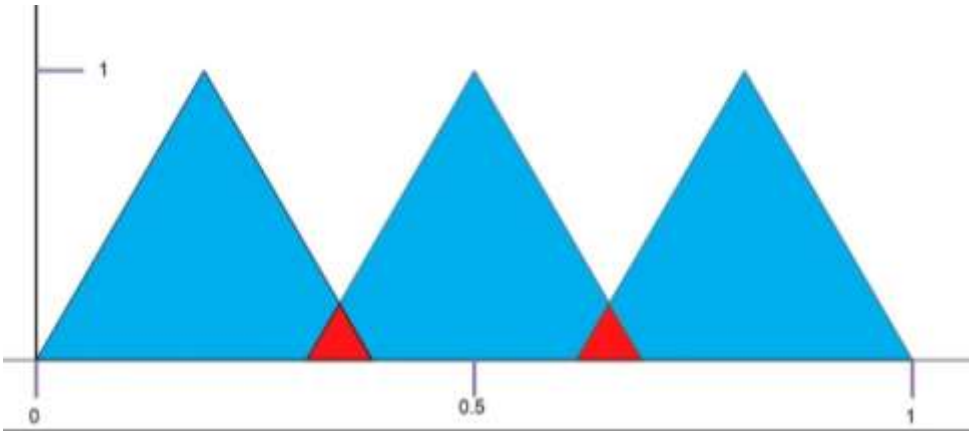


Figure3.3: fuzzification.

The Figure (3.3) above is showing fuzzy sets example; the blue triangles are the ranges while the red ones are the intersection areas between ranges of a set.

2. Rules and operators: Fuzzy logic is using membership values to identify to which ranges the values belong, the idea is illustration Figure (3.4).

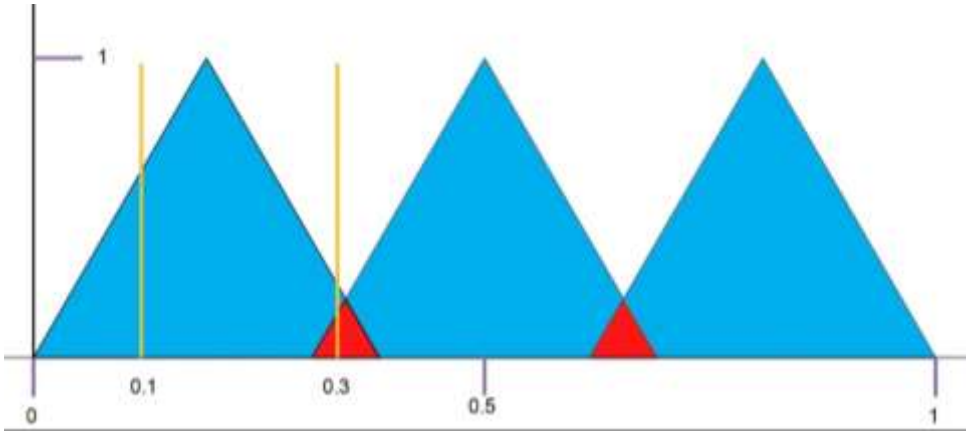


Figure 3.4: membership values.

The first value (0.1) is belonging to one range, so it will get only one membership; the second value (0.3) is belong to two ranges, so it will have two membership values. Fuzzy Logic chooses between the different membership values by MIN and MAX functions depends on the operation between the values, MIN is used for AND operations while MAX is used for OR operations. In our proposed system we use MIN function, which selects the lower membership value.

After obtaining the membership values for each input, the Center of Gravity formula will be used to figure out the result based on the output ranges.

$$\mathrm{COG} = \frac{\sum Memership\ values \times Intersections\ on\ output}{\sum Membership\ values \times Intersection} \qquad \text{E.q 3.1}$$

3. Defuzzification: The objective is to obtain a variable from fuzzy values, the result is a classification or a range name that the COG value is belong to.

As we mentioned before, we use triangles in fuzzy logic. There are two axis x and y, y axis

presented the membership value, which is between 0-1 and the x axis, its range is from 0 to 100, and it represents the match ratio. There are three fuzzy sets, which are education, experiences, and skills as input. Each set has three ranges which are low suitability, medium suitability, and high suitability. Every range has membership value from 0 to 1, and there is one output describe the total match ratio from the three fuzzy sets with one range and one membership.

B. Sugeno FIS:

This fuzzy inference system was proposed to develop a systematic approach for generating fuzzy rules from a given input-output dataset. A typical fuzzy rule in a first-order Sugeno fuzzy model has the form: IF x is A and y is B THEN z = f (x, y), where: A and B are fuzzy sets in the antecedent, z = f(x, y) is a crisp function in the consequent.

3.6.6. Characteristics of fuzzy logic:

1. Mimic the way a human think when making certain decisions to solve a problem.
2. Flexible and easy in implementation with different techniques as machine learning technique.
3. It is the most suitable method to solve uncertain problems or approximate reasoning.
4. Fuzzy logic views inference as a process of propagating elastic constraints
5. Fuzzy logic allows you to build nonlinear functions of arbitrary complexity.

3.6.7. Advantages and disadvantages of Fuzzy Logic System:

A- Advantages of Fuzzy Logic System:

- The fuzzy logic deals with much type of inputs such as imprecise, distorted or noisy input information. It helps you to deal with the uncertainty; It may not offer accurate reasoning, but the only acceptable reasoning.
- It is quite simple, and comes with mathematical concepts of set theory and the reasoning.
- Fuzzy Logic structure is easy and understandable, so it widely used in many commercial and practical purposes.
- It easily modified to improve or alter system performance it keeps the overall system cost and complexity low.

B- Disadvantages of Fuzzy Logic Systems:

- Fuzzy logic is not always accurate, so the results are perceived based on assumption, so it may not be widely accepted.
- Fuzzy systems don't have the capability of machine learning as-well-as neural network type pattern recognition.
- Validation and Verification of a fuzzy knowledge-based system needs extensive testing with hardware.
- Setting exact, fuzzy rules and, membership functions is a difficult task
- Some fuzzy time logic is confused with probability theory and the term.

3.7. BabelNet

BabelNet defined as encyclopedic dictionary, with multilingual semantic network to supply High-quality automatic sense-based translations, Searching and explore the network. BabelNet consist of two large complementary resources, and merge them which are WordNet (lexicographic entries) it is full-fledged taxonomy Wikipedia (encyclopedic)

which is multilingual and continuously updated.

The Multilingual Semantic Network which connects concepts and named entities in a very large network of semantic relations, made up of about 16 million entries, called Babel synsets. Each Babel synset represents a given meaning and contains all the synonyms which express that meaning in a range of different languages [18].

3.7.1. Wikipedia:

Wikipedia is a multilingual Web-based encyclopedia. It is a collaborative open source medium edited by volunteers to provide a very large wide-coverage repository of encyclopedic knowledge. Each article in Wikipedia is represented as a page (Wikipage) and presents information about a specific concept or named entity. The title of a Wikipage is composed of the lemma of the concept defined plus an optional label in parentheses which specifies it's meaning if the lemma is ambiguous. The text in Wikipedia is partially structured. Apart from Wiki pages having tables and infoboxes, various relations exist between the pages themselves. These include:

1. Redirect pages: These pages are used to forward to the Wikipage containing the actual information about a concept of interest. This is used to point alternative expressions for a concept to the same entry, and thus model's synonymy.
2. Disambiguation pages: These pages collect links for a number of possible concepts an arbitrary expression could be referred to. This models homonymy and polysemy.
3. Internal links: Wikipages typically contain hypertext linked to other Wikipages, which refers to related concepts.

4. Inter-language links: Wikipages also provide links to their counterparts contained within wikipedia in other languages.
5. Categories: Wikipages can be assigned to one or more categories, i.e., special pages used to encode topics.

3.7.2. WordNet:

WordNets the most popular lexical knowledge resource in the field of NLP. It is a computational lexicon of the English language based on psycholinguistic principles. A concept in WordNet is represented as a synonym set (called synset), i.e., the set of words that share the same meaning. synsets are related to each other by means of lexical and semantic relations. The inventory of semantic relations varies among parts of speech. bablenet encode knowledge as a labeled directed graph:

1. Each vertex is a Babel synset
2. Each edge is a semantic relation between synsets:
    - is-a relations
    - part-of relations
    - instance-of relations

Also, translating Babel synsets consist of two steps:

1. Exploiting Wikipedia interlanguage links.
2. Filling the lexical translation gaps using a Machine Translation system to translate the English lexicalizations of a concept.

- For each word sense s, Babel translate:
    - Sentences from SemCor(a corpus annotated with WordNet senses) which contain s.

- Sentences from Wikipedia linked to the Wikipage of s.
- The most frequent translation of s is selected for each target language.

BabelNet encodes knowledge as a labeled directed graph G = (V , E) where V is the set of nodes and E ⊆ V × R × V is the set of edges connecting pairs of concepts. Each edge is labeled with a semantic relation from R, i.e., {is-a, part-of , . . . , ∈}, where ∈ denotes an unspecified semantic relation. Importantly, each node v ∈ V contains a set of lexicalizations of the concept for different languages, it call such multilingual lexicalized concepts Babel synsets. Concepts and relations in BabelNet are harvested from WordNet and Wikipedia [19].

BabelNet covers 284 languages and a merger of resources of different kinds:

- WordNet: the most popular computational lexicon of English
- Open Multilingual WordNet: a collection of open wordnets
- WoNeF: a French WordNet
- ItalWordNet: an Italian WordNet
- Wikipedia: the largest collaborative encyclopedia
- Wikidata: the largest collaborative knowledge base
- Wiktionary: the largest collaborative dictionary
- OmegaWiki: a medium-size collaborative multilingual dictionary
- GeoNames: a worldwide geographical database
- Microsoft Terminology: a computer science thesaurus.

3.7.3. BabelNet advantages:

- Multilingualism: the same concept is expressed in tens of languages and it covers 284 languages and 16 million entries!

- Concepts and named entities together: dictionary and encyclopedic knowledge is semantically interconnected.
- Fu/14fledged taxonomy: is-a relations are available for both concepts and named - entities (Wikipedia Bitaxonomy) with labeled relations, pictures, multilingual synsets.
- Easy access: Java and HTTP RESTful APIs; SPARQL endpoint.

3.8. Natural Language process (NLP):

Natural Language Processing or NLP In briefly, the machine ability to understand human language is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages.

3.8.1. Natural Language process model “bag of words” steps:

- Text tokens: is the process of segmenting running text into sentences and words. It is the task of cutting a text into pieces called tokens, and at the same time throwing away certain characters, such as punctuation.
- Stop Words Removal: includes getting rid pronouns and prepositions such as “and”, “the” or “to” in English. In this process some very common words that appear to provide little or no value to the NLP objective are filtered and excluded from the text to be processed, hence removing widespread and frequent terms that are not informative about the corresponding text.

  There is no universal list of stop words. These can be pre-selected or built from scratch.
- Stemming: Refers to the process of slicing the end or the beginning of words with the intention of removing affixes (lexical additions to the root of the

word). Prefixes are affixes that are attached at the beginning of the word for example “inter” in the word “interactive” and the ones attached at the end of the word are called suffixes, for example “ful” in the word “helpful”).

3.8.2. Classification of Stemming Algorithms

Stemming algorithms can be classified in three groups: truncating methods, statistical methods, and mixed methods. Each of these groups has a typical way of finding the stems of the word variants.

- Truncating Methods (Affix Removal): As the name clearly suggests these methods are related to removing the suffixes or prefixes (commonly known as affixes) of a word. Truncating methods incloud Lovins Stemmer, Porters Stemmer, Paice/Husk Stemmer and Dawson Stemmer.
- Statistical Methods. These are the stemmers who are based on statistical analysis and techniques. Most of the methods remove the affixes but after implementing some statistical procedure. Statistical methods include, N-Gram2 Stemmer, HMM3 Stemmer and YASS Stemmer.

3.8.3. Porters Stemmer

Porters stemming algorithm is the most popular stemming methods proposed. It is based on the idea that the suffixes in the English language (approximately 1200) are mostly made up of a combination of smaller and simpler suffixes. It has five steps, and within each step, rules are applied until one of them passes the conditions. If a rule is accepted, the suffix is removed accordingly, and the next step is performed. The resultant stem at the end of the fifth step is returned.

The rule looks like the following: <condition> <suffix> → <new suffix>For example, a rule (m>0) EED → EE means "if the word has at least one vowel and consonant plus EED ending, change the ending to EE". So "agreed" becomes "agree" while "feed" remains unchanged.

- **Porters Stemmer Advantages:**
  1. Produces the best output as compared to other stemmers.
  2. Less error rates.
  3. Compared to Lovins it's a light stemmer.
  4. The Snowball stemmer framework designed by Porter is language independent approach to stemming.

- **Limitations:**
  1. The stems produced are not always real words.
  2. It has at least five steps and sixty rules and hence is time consuming.

Chapter 4

Recommendation Model

4.1. Introduction

The importance of the study reveals in Matching Automatic job with an available job position. There are many techniques and algorithms to help job seekers find the right job. Some are traditional algorithms while others have been found recently, also there are a large number of hybrid algorithms that are a combination of many algorithms. All these algorithms aim to find the best job for the candidate and to make recruitment processes faster, more accurate and transparent.

The aim of our study is to design and develop an e-recruiting system which helps job seekers to find a suitable job for themselves, by giving them the chance to match their CVs with a good job, even though they are not accepted when the education section was not matched with the job application, or getting a low match ratio to be acceptable.

The proposed is a new hybrid model combining the natural language processing (NLP) tasks, expert knowledge database, BabelNet to get bag of words, content base filtering approach to calculate similarity by using TF-IDF, and Fuzzy Logic to select keywords from text or document and obtain the final results.

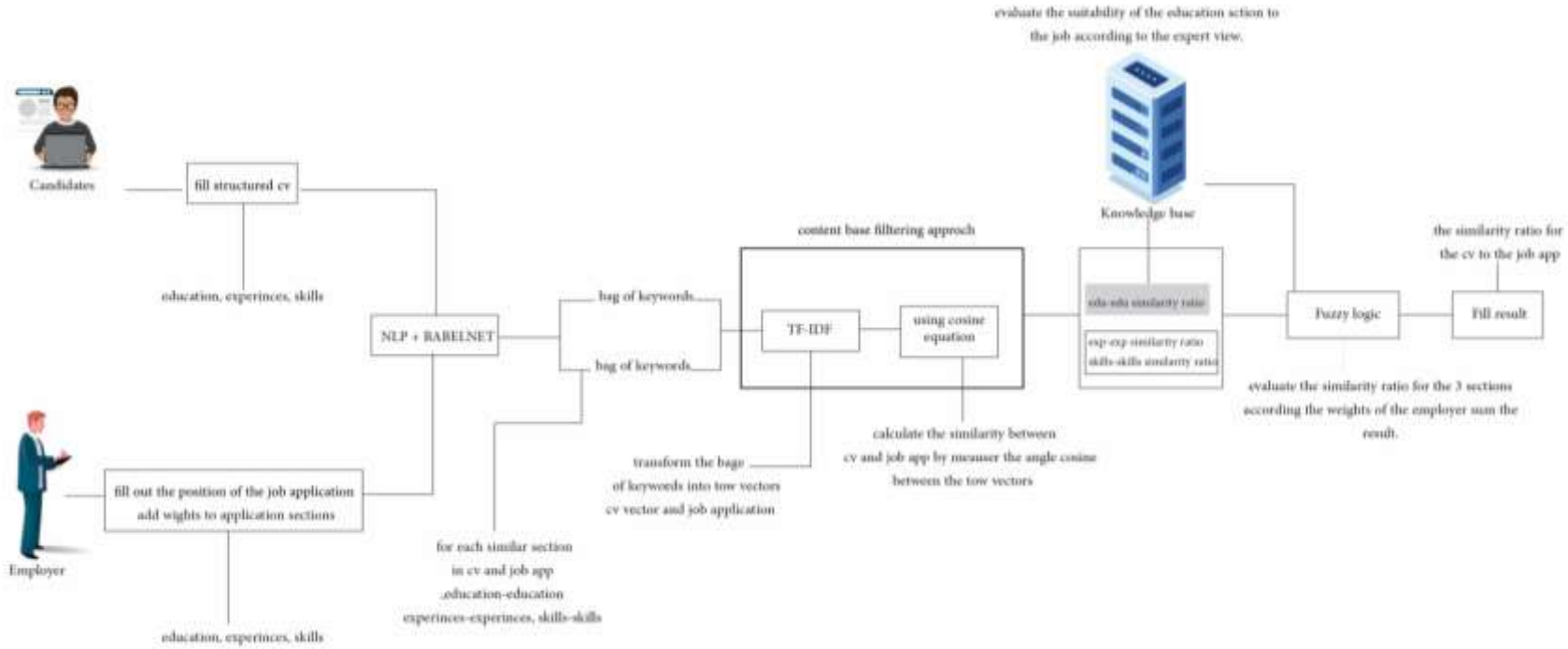


Figure 4.1: The proposed system architecture.

4.2. The Proposed system Architecture:

The system is consisting of two models: User profile and Job application. Also, the system has two main flows, one for the employer and one for the applicants; the flows illustrated the using of job profile and user profile until extract the results.

The system architecture contains a set of algorithms and methods as appear in figure 4.1, it will be used and as the following:

4.2.1. The natural language processing (bag of words model):

In this model the text is represented by bag of words without following the grammatical for and the correct order of words. It also transforms text into a more digestible form so algorithms can perform better. Natural language processing (NLP) is concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. The natural language processing (NLP) efficiency is in improving free text analysis, and describes the main linguistic ideas. It is capacity to learn the structure or relationship of a text that is represented in a numerical format.

NLP (bag of words model), analyzing the text as the following:

- Text tokens: breaking up the texts into individual words.
- Remove stop words: such as (“is”, “the”, “a”, “an”, “in”, etc.) and punctuation marks.
- Stemming: is the process of reducing a word to its roots, by removing the suffixes from the word such as "ing", "tion". This step done by using porter stemming algorithm.

4.2.2. Babel dictionary[1]:

It used to return Synonym words for the words in our system. Synonym words either in user profile model or in job application model.

As we mentioned before in chapter 3, BabelNet defined as encyclopedic dictionary, created by means of the seamless integration and interlinking of Wikipedia, with WordNet, it consist of multilingual encyclopedic dictionary, and semantic network, "which connects concepts and named entities, in a very large network of semantic relations", to supply High-quality automatic sense-based translations, Searching and Explore the network.

BabelNet efficiency comes from its consist which are WordNet, with the most popular computational lexicon of English (lexicographic entries) which is full-fledged taxonomy, and the largest multilingual Web encyclopedia Wikipedia, which is multilingual and continuously updated. and other lexical resources such as Wiktionary, OmegaWiki, Wikidata, Wikipedia infoboxes…etc. Babel synsets is a multilingual semantic network named entities, in a very large network of semantic relations. Each Babel synset represents a given meaning and contains all the synonyms, which express that meaning in a range of different languages. Example for the skill ‘work in a team’:

Info: Babel data before: work in a team
Info: ***** remove stop words + stemming *****
Info: ** analayze result: *******
Info: [work, team]
Info: *****babel start ****
Info: method babel .....

[1] https://babelnet.org/guide

Info: work (6) : study

Info: method babel .....

Info: team (0):

Info: **** babel end ********

Info: result from babel : work study team

Info: *** analyze (stop words + stemming) babel result *****

Info: ***** remove stop words + stemming *****

Info: ** analayze result : *******

Info: [work, studi, team]

Info: result after analyze: [work, studi, team]

Info: final bag of words: work studi team

Table 4.1: Bag of words from bable.net for position skills

Position skills (NPL Process)	**BOW (babelanet synonym words)**
English, language	english establish california compani base irvin comput defunct greater los angel area unit state softwar microcomput technolog microdata corpor languag cultur literatur literari critic theory
work, team	work studi team
Time, management, skill	time clock signal comput program real oper system technolog unix process task manag relat softwar stub sus util applic layer protocol network skill

4.2.3. Content based filtering approach:

Content based filtering approach consists of TF-IDF and cosine similarity.

1. Term Frequency (TF) and Inverse Document Frequency (IDF) TF is simply the frequency of a word in a document. IDF is the inverse of the document frequency among the whole corpus of documents.

Term Frequency-Inverse Document Frequency (TF-IDF) is in the sub-area of Natural Language Processing (NLP). It is used in information retrieval for feature extraction purposes. In short, you are somehow counting the occurrence of each word in a document and weight the importance of each word, and calculate a score for that document, term frequency of the word in the current document to the total number of words in the document. It signifies the occurrence of the word in a document and gives higher weight when the frequency is more so it is divided by document length to normalize.

The efficiency of TF-IDF appears in helped us understand how to calculate the scores for data represented as 1/0. Usually the similarity will be derived from the description of the item. Then each item will be represented by a TF-IDF vector. TF-IDF is very useful in a lot of areas such as content-based filtering, it is crucial to the building of item profiles in the content-based filter recommender systems (user profile and job application in our system).

TF: it used to weigh a term's frequency and its equation is: TF (t) = (Number of times term t appears in a document) / (Total number of terms in the document). IDF: inverse document frequency, it measures the important of a term, and its equation is: IDF (t) = log-e (Total number of documents / Number of documents with term t in it).

The result of TF-IDF = TF * IDF, finally the similarity between any two vectors calculated by the cosine equation [23].

2. Cosine similarity measures the similarity between two vectors It is often used to measure document similarity in text analysis. It is measured by the cosine of the

angle between two vectors and determines whether two vectors are pointing in roughly the same direction.

Cosine is a measure of similarity that can be used to compare documents or, say, give a ranking of documents with respect to a given vector of query words.

A cosine value of 0 means, that the two vectors are at 90 degrees to each other (orthogonal), and have no match. The closer the cosine value to 1, the smaller the angle and the greater the match between vectors [23].

$$cos = \frac{A \,.B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \; B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2} \; \sqrt{\sum_{i=1}^{n} {B_i}^2}} \qquad \text{[25].} \qquad \text{Eq. (4.1)}$$

Where A and B are the vectors which presented the bag of words value, which calculated form TF-IDF for user profile and job application. Also, A_i and B_i are the value of one word from the two vectors A and B, which presented the bag of words for user profile, and job application.

We use cosine similarity to measures the similarity between two vectors, x and y. Cosine similarity is often used to measure document similarity in text analysis, and can be used to compare documents, or ranking of documents with respect to a given vector of query words. Figure 4.2 shows the relationship between similarity and cosine (θ) [25].

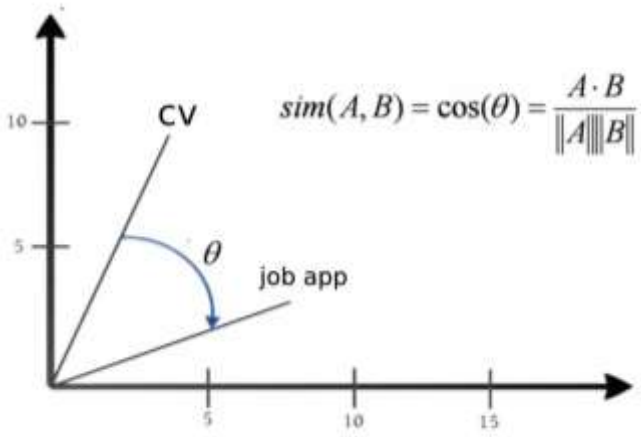


Figure 4.2: the similarity and cosine relationship

Cosine similarity is different from Pearson-based measure in that it is a vector-space model which is based on linear algebra rather that statistical approach. It measures the similarity between two n-dimensional vectors based on the angle between them. Cosine-based measure is widely used in the fields of information retrieval and texts mining to compare two text documents, in this case, documents are represented as vectors of terms.

The following example is a content base similarity:

1. **Bag of Words (BOW) Model:**

The Bag of Words (BOW) model is the simplest form of text representation in numbers. Like the term itself, we can represent a sentence as a bag of words vector (a string of numbers).

We supposed that these words are the bag of word for each user and position:

- Position skills: Web standards for HTTP, HTML, CSS3, and responsive design
- User skills: HTML5 and CSS3 Web Language Design.

We will first build a vocabulary from all the unique words in the above two reviews. The vocabulary consists of these 9 words: 'Web', 'standard', 'HTTP', 'HTML', 'CSS3', 'responsive', 'design', 'HTML5', 'Language', table 4.2 show the words appearance and its repetition in the two reviews.

Table 4.2: the words appearance and repetition in the two reviews

	web	**standards**	**HTTP**	**HTML**	**HTML5**	**CSS3**	**responsive**	**design**	**Language**	**Length of skills (in words)**
Position skills	1	1	1	1	0	1	1	1	0	7
User skills	1	0	0	0	1	1	0	1	1	5

We take each of these words and mark their occurrence in the two skills sections 1s and 0s. This will give us 2 vectors for the two sections:

Vector of position skills: [1 1 1 1 0 1 1 1 0]

Vector of user skills: [1 0 0 0 1 1 0 1 1]

$$\text{Term frequency matrix} = \begin{matrix} \\ \text{Position skills} \\ \text{User skills} \end{matrix} \begin{matrix} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 & w_9 \end{matrix} \begin{vmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{vmatrix}$$

2. Term Frequency-Inverse Document Frequency (TF-IDF)

a. Term Frequency (TF), as we mentioned Term Frequent (TF). It is a measure of how frequently a term, t, appears in a document, d, and it is the number of times a word occurs in a document. Tf (t,d) = f (t,d), T -> term, d -> document.

TF = number of times 'term' appears in document)/ (number of terms in document) [25]. Eq. (4.2)

We will use again, the same vocabulary we had used to build in the bag-of-words model to show how to calculate the TF for user skills section:

User skills: HTML5 and CSS3 Web Language Design.

- Vocabulary**:** 'HTML5', 'CSS3', 'Web', 'Language', 'Design.
- Number of words in user skills section = 5
- TF for the word 'HTML5' = (number of times 'HTML5' appears in user skills section)/ (number of terms in in user skills section) = 1/5. Table 4.3 shows the calculation of IF for User skills, and the position skills

Table 4.3: the calculation of IF for User skills, and the position skills

Term	**Position skills**	**User skills**	**TF Position skills**	**TF User skills**
Web	1	1	1/9 = 0.11	1/5 = 0.2
standard	1	0	1/9 = 0.11	0
HTTP	1	0	1/9 = 0.11	0
HTML	1	0	1/9 = 0.11	0
CSS3	1	1	1/9 = 0.11	1/5 = 0.2
responsive	1	0	1/9 = 0.11	0
design	1	1	1/9 = 0.11	1/5 = 0.2
HTML5	0	1	0/9 = 0	1/5 = 0.2
Language	0	1	0/9 = 0	1/5 = 0.2

b. Inverse Document Frequency (IDF): IDF is a measure of how important a term is.

$$IDF_t = \log \frac{\text{total number of documents}}{\text{number of documents with term "t"}}$$ [25]. Eq. (4.3)

We can calculate the IDF values for the all the words in user skills:

IDF (web) = log (number of documents/number of documents containing the word 'this')

IDF (web) = log (2/2) = log (2) = 0

We can calculate the IDF values for each word as table 4.4 shows.

Table 4.4: IDF values for each word.

Term	Position skills	User skills	IFD
Web	1	1	Log 2/2 = 0
standard	1	0	Log 2/1 = 0.301
HTTP	1	0	Log 2/1 = 0.301
HTML	1	0	Log 2/1 = 0.301
CSS3	1	1	Log 2/2 = 0
responsive	1	0	Log 2/1 = 0.301
design	1	1	Log 2/2 = 0
HTML5	0	1	Log 2/1 = 0.301
Language	0	1	Log 2/1 = 0.301

We can now compute the TF-IDF score for each word in the corpus. Words with a higher score are more important, and those with a lower score are less important [25]:

$$TF\text{-}IDF_{t,d} = TF_{t,d} * IDF_t \quad \text{Eq. (4.4)}$$

We can now calculate the TF-IDF score for every word in user skills section:

TF-IDF (‘web’, user skills section) = TF (‘web’, user skills section) * IDF (‘web’)

= 1/5 * 0 = 0

Similarly, we can calculate the TF-IDF scores for all the words with respect to all the sections, as in table 4.5:

Table 4.5: the TF-IDF values

Term	Position skills	User skills	TF Position skills	TF User skills	IDF	TF-IDF Position skills	TF-IDF User skills
Web	1	1	0.11	0.2	0	0.11* 0 = 0	0.2 * 0 = 0
Standard	1	0	0.11	0	0.301	0.11* 0.301 = 0.033	0
HTTP	1	0	0.11	0	0.301	0.033	0
HTML	1	0	0.11	0	0.301	0.033	0
CSS3	1	1	0.11	0.2	0	0	0
Responsive	1	0	0.11	0	0.301	0.033	0
Design	1	1	0.11	0.2	0	0	0
HTML5	0	1	0	0.2	0.301	0	0.2 * 0.301 = 0.0602
Language	0	1	0	0.2	0.301	0	0.0602

We have now obtained the TF-IDF scores for our vocabulary. TF-IDF also gives larger values for less frequent words and is high when both IDF and TF values are high i.e the word is rare in all the documents combined but frequent in a single document.

3. Cosine similarity:

$$COS = \frac{A \,.B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \; B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2} \; \sqrt{\sum_{i=1}^{n} {B_i}^2}} \qquad \text{Eq. (4.1)}$$

(A): Position skills TF-IDF array [0 0.033 0.033 0.033 0 0.033 0 0 0]

(B): User skills TF-IDF array [0 0 0 0 0 0 0 0.06.2 0.0602]

To calculate cosine similarity, we need to calculate:

1. $$\sum_{i=1}^{n} A_i\ B_{i=} (0*0)+(0.033*0)+(0.033*0)+(0.033*0)+(0*0) \\ +(0.033*0)+(0*0)+(0*0.2)+(0*0.2)=0$$

2. $$\| A \| = \sqrt{\sum_{i=1}^{n} {A_i}^2} \qquad \text{[25].} \qquad \text{Eq. (4.5)}$$

$$= \sqrt{(0)^2+(0.033)^2+(0.033)^2+(0.033)^2+(0)^2+(0.033)^2+(0)^2 \\ +(0)^2+(0)^2}$$

$$= 0.066$$

3. $$\| B \| = \sqrt{\sum_{i=1}^{n} {B_i}^2} \qquad \text{[25].} \qquad \text{Eq. (4.5)}$$

$$= \sqrt{(0)^2+(0)^2+(0)^2+(0)^2+(0)^2+(0)^2+(0)^2+(0.062)^2+(0.062)^2}$$

$$= 0.042284$$

$$\therefore COS = \frac{A.B}{\|A\|\|B\|} = \frac{0*0}{\|0.066\|*\|0.042284\|} = 0$$

In this random case the similarity equal 0, a cosine value of 0 means that the two vectors are at 90 degrees to each other, and have no match. The closer the cosine value to 1, the smaller the angle and the greater the match between vectors.

4.2.4. Mamdani Fuzzy Inference System

We build our system using mamdani Fuzzy Inference System, the usefulness of mamdani systems for modeling, simulation and decision making. mamdani systems are most often classified as a form of Approximate Reasoning, which has been defined as the process or

processes by which a possible imprecise conclusion is deduced from a collection of imprecise premises, they are designed to incorporate expert knowledge in the form of IF-THEN rules expressed in natural language. IF-THEN rules express some causality or relation between two (or more) variables.

Mamdani Fuzzy Inference System has three processes: fuzzy process**,** Inference Engine: which contains IF-THEN rules and defuzziation.

The efficiency of mamdani systems is to explore the logical consequences of a model based on IF-THEN rules via simulation. In the best-case scenario a mamdani system provides a function that complies with its generating set of IF-THEN rules, which is a different exercise from that of finding the relation or consequences implied by those rules.

We used Mamdani fuzzy inference system to build a control system by synthesizing a list of linguistic rules obtained from experienced human operators. We used it because of some reasons, first of all its suitable for human input, and it’s easier to understand when the rule is created by human expert knowledge.

Example of mamdani fuzzy logic process:

To evaluate the effectiveness and the reliability of the mamdani fuzzy logic, we will use a job post and a candidate CV as show in figures 4.3 and 4.4.

Personal Information		
Mohammad Abdalfattah Yasin	21/04/1989	
Jenin – Aneen – Othamn Bin Affan St.		
m.y.aauj@gmail.com	0568098360	
Experience		
No Experience		
Education		
Arab American University	B.S. Computer Science	2007 - 2017
Skills		
Oracle	Beginner	-
Other Information		
Yes	0	Unknown

Figure 4.3: the candidate CV.

Company Information		
Ramalla - Almasyoon	00970-2-2987792	
Position Information		
Experience		
Development	0+	
Education: B.S Computer Science		
Skills		
Java	JEE	OOA
OOD	Oracle	MySQL
Object Relational Mapping (ORM)	JavaScript (JS)	HTML
JQuery	AJAX	CSS
Other Information Personal Skills: 1. Team Player 2. Good presentation and communication skills 3. Time management skills 4. Good English language (written and verbal)		

Figure 4.4: the job post.

In the above job post, the education weight was 80%, experience weight was 20%, and other inputs were weightless for the acceptance process.

Evaluate education and experience: assume that the inputs to the system are the education, skills and experience. While the output variable which represents the final result is suitability.

As we mentioned before to obtain the fuzzy logic values, we need to follow up the common steps in mamdani fuzzy logic:

a. Fuzzification:

We have three inputs to fuzzified them to three fuzzy sets, which are education, experience, and skills.

For education, we have three sets which are low suitability, medium suitability, and high suitability as the following:

Fuzzify input variable edu': {'poor', 'good' , 'excellent'}

TERM lowEdu = (0,1) (33, 1) (66, 0) ;

TERM medEdu = (33, 0) (66,1) (100,0);

TERM highEdu= (66 0) (100,1);

Also skills and experience fuzzified as:

Fuzzify input variable 'skill': {'poor', 'good' , 'excellent'}

TERM lowEdu = (0,1) (33, 1) (66, 0) ;

TERM medEdu:= (33, 0) (66,1) (100,0);

TERM highEdu = (66 0) (100,1);

Fuzzify input variable 'experience': {'poor', 'good' , 'excellent'}

TERM lowEdu = (0,1) (33, 1) (66, 0) ;

TERM medEdu= (33, 0) (66,1) (100,0);

TERM highEdu= (66 0) (100,1);

The following figure 4.4 is illustrating the values of fuzzy sets.

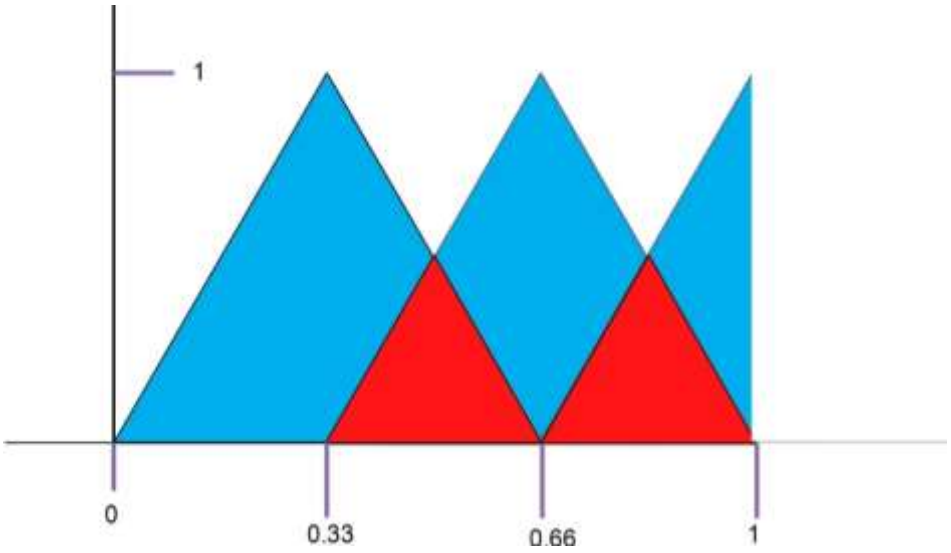


Figure 4.5: the values of fuzzy sets.

Suppose there is a job post, we have 80% percent for the education, and suppose the B.S of computer science takes 100% for the development position. On the other hand, we have the experience that weights 20%, for this applicant he gains 0% for the lack of experience. The values from 0 to 1 are the suitability, based on the membership function.

b. If-then rules:

There are many rules for our proposed system such as:

RULE 1: IF edu IS lowEdu and skill IS lowSkill and ex is lowEx THEN tip IS low;

RULE 2: IF edu IS lowEdu and skill IS lowSkill and ex is medEx THEN tip IS low;

RULE 3: IF edu IS lowEdu and skill IS lowSkill and ex is highEx THEN tip IS med;

RULE 4: IF edu IS lowEdu and skill IS medSkill and ex is lowEx THEN tip IS low;

RULE 5: IF edu IS lowEdu and skill IS medSkill and ex is medEx THEN tip IS low;

RULE 6: IF edu IS lowEdu and skill IS medSkill and ex is highEx THEN tip IS med;

RULE 7: IF edu IS lowEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 8: IF edu IS lowEdu and skill IS highSkill and ex is medEx THEN tip IS med;

RULE 9: IF edu IS lowEdu and skill IS highSkill and ex is highEx THEN tip IS med;

RULE 10: IF edu IS medEdu and skill IS lowSkill and ex is lowEx THEN tip IS low;

RULE 11: IF edu IS medEdu and skill IS lowSkill and ex is medEx THEN tip IS low;

RULE 12: IF edu IS medEdu and skill IS lowSkill and ex is highEx THEN tip IS med;

RULE 13: IF edu IS medEdu and skill IS medSkill and ex is lowEx THEN tip IS low;

RULE 14: IF edu IS medEdu and skill IS medSkill and ex is medEx THEN tip IS med;

RULE 15: IF edu IS medEdu and skill IS medSkill and ex is highEx THEN tip IS med;

RULE 16: IF edu IS medEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 17: IF edu IS medEdu and skill IS highSkill and ex is medEx THEN tip IS med;

RULE 18: IF edu IS medEdu and skill IS highSkill and ex is highEx THEN tip IS high;

RULE 19: IF edu IS highEdu and skill IS lowSkill and ex is lowEx THEN tip IS med;

RULE 20: IF edu IS highEdu and skill IS lowSkill and ex is medEx THEN tip IS med;

RULE 21: IF edu IS highEdu and skill IS lowSkill and ex ishighEx THEN tip IS med;

RULE 22: IF edu IS highEdu and skill IS medSkill and ex is lowEx THEN tip IS med;

RULE 23: IF edu IS highEdu and skill IS medSkill and ex is medEx THEN tip IS med;

RULE 24: IF edu IS highEdu and skill IS medSkill and ex is highEx THEN tip IS high;

RULE 25: IF edu IS highEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 26: IF edu IS highEdu and skill IS highSkill and ex is medEx THEN tip IS high ;

RULE 27: IF edu IS highEdu and skill IS highSkill and ex is highEx THEN tip IS high ;

Therefore, the rules will be fired to the first CV is:

RULE 19: IF edu IS highEdu and skill IS lowSkill and ex is lowEx THEN tip IS med;

The CV will take 1 as a membership value for the education.

For experience, we have three sets which are low experience, medium experience, and high experience the following figure is illustrating the values.

The values from 0 to 1 are the suitability based on the required experience by the job post, while the experience is optional in this job post, based on the membership function, this CV will take 0 as a membership value for the education.

For skills, there is no need to measure them because the weight is zero for them.

c. Center of Gravity

By applying the COG equation, we will have the following:

$$\text{COG} = \frac{\sum Memership\ values \times Intersections\ on\ output}{\sum Intersection} \qquad \text{Eq. (4.7)}$$

$$\text{COG} = \frac{\begin{matrix} Education\ Memership\ value \times Education\ Intersections \times Education\ Weight + \\ Experience\ Membership\ Values \times Expericne\ Intersection \times Experience\ Weight \end{matrix}}{Education\ Intersections + Experience\ Intersection}$$

Eq. (4.8)

$$\text{COG} = \frac{1 \times 1 \times 0.8 + 0 \times 0 \times 0.2}{1 + 0}$$

COG = 0.8

d. Defuzziation:

The CV suitable for the position 80% , that mean the medium, according to Defuzziation of the output variableas:

DEFUZZIFY suitability // Defzzzify output variable ' suitability ': {'poor', 'good', 'excellent'}

TERM lowEdu = (0,1) (33, 1) (66, 0) ;

TERM medEdu = (33, 0) (66,1) (100,0);

TERM highEdu = (66 0) (100,1);

So the candidate CV valued as good to the position.

4.3. Conclusion:

Our hybrid system consists of several effective and efficient algorithms, all of them work with documents, words and text, whether in the analysis texts such as NLP process, or extracting synonymous words to produce a bag of keywords using BableNet, or from the ability of calculate the repetition of words in text and convert words into values 0 / 1, and represent them in vectors as well as the ability, to calculate the angle between the two vectors by using the cosine equation, the result represent the similarity value between them. At last using fuzzy logic used to calculate the final results.

All these successive algorithms working and complementing each other, to make the system highly efficient in dealing with text, and make the system more accurate and more realistic.

Chapter 5

Implementation:

5.1. Introduction

As we mentioned earlier, at the present time, there is a gap between the requirements of the labor market and the scientific specializations available in universities. In universities, the presence of similar majors with different names leads to have a large number of majors for a limited number of jobs. This has made it difficult to find the suitable major for available/referred jobs in the job market.

Therefore, we proposed a model to manage the recruitment process by assessing the CVs of applicants according to the needs of the employer. The employer can determine these needs accurately by dividing the job application form into three sections: education, experience and skills. The employer can give each section weighting percentage according to its importance for him. This can be accomplished by give the weighting percentage (ranking) according to an approach that uses a method of content-based filtering approach, Fuzzy logic and BabelNet dictionary to match the vacancy with applicants, to extract the best results.

Also, we created a database represents the different opinions of the lecturers in the Palestinian universities, about the relationship between the different specializations and the vacancies in the labor market by conducting (carrying out) a survey. The main objective of this survey is to measure the appropriateness of specialization for the job according to the

educational outcomes.

The system allows the employer to evaluate the employees after a period of employment, as a kind of feed back to the database. Then the newly added data to the database improves the resulting arithmetic means which obtained initially/originally from the targeted survey to the academics in the universities, to evaluate the matching of jobs in the information technology market and the existing relevant specialization.

5.2. Discussion

5.2.1. CV and Resume:

Most jobs require the user to submit a resume or CV to introduce his/her education, experiences and skills in a professional format to a potential employer; its purpose is to convince the employer his qualifications necessary to fulfill the needs of the posted job.

There are differences between a curriculum vitae and a resume, but they are very similar in formatting shape, resumes written in one page, but CVs are more comprehensive because they written into several pages.

There are two types of resume and CV, European CV and the U.S. resume and CV, In the US, a CV is distinct from a resume in the sense that it is used primarily in academic and research circles or in medical careers. It is much more comprehensive than a U.S. resume, which is almost always one page long. The U.S. resume is used by most American job seekers [8].

Using the term “European CV” is being for simplicity’s sake to present a snapshot of what CV expectations generally are like Western European countries.

The differences between the European CV and. U.S. resume. Or U.S. CV:

1. Length: for a European CV are 2 to 3 pages. A U.S. CV starts at 3 pages-20 pages.
2. Personal information: European CVs start off with the following types of information: Marital Status, Age, Number of children (ages optional), Personal Interests, Nationality, and gender are also commonly mentioned on a European CV. On the other hand, sharing any type of personal information on a resume, CV, or cover letter is considered very unprofessional in the US.
3. High school information: A European CV is always expected to contain some secondary school information; On the other hand, a U.S. CV does not contain this information.
4. Photo: A European CV will, in some countries, requires a photo, A U.S. resume or U.S. CV will almost never include a photo.
5. Paper size for printed versions: The European CV should always be printed on ISO A4 paper, The U.S. resume and U.S. CV, should always be printed on American "letter size" paper.

5.2.2. Structured CV:

We will use a structured CV template based on the U.S. resume and CV type. The structure of CV is an important factor in job search. A well-structured CV that looks professional and presents the information clearly will be much more effective than a messy CV that is difficult to read. The proposed template will facilitate the information retrieval and inference from the applicants' CVs. The following Figure 5. 1 shows structured CV template that will be used.

Personal Information		
Name Mr./Mrs. (option)	Date of Birth	
Address		
Email	Phone Number	
Experience (new to old)		
Work Place	Position	Duration
Education		
Institute	Degree	Duration
Skills		
Skill	Level	Certificate
Other Information		
Available to change address	Current Salary	Expected Salary

Figure 5.1: The structured CV template

The system allows the user to fill, download and update the structured CV form.

5.2.3. Structured Job Application:

We also used a structured job application form to filled in a specific sections, which is the same of the CV sections (education- experience- skills) the proposed format will facilitate the information retrieval and inference from the applicants' CVs.

We also reviewed the (ISO 9001 - the Job Description) [7]to realize our goal of the proposed structure form for the job application, which is finding the most suitable employee. We try to cover all the job description rules in our job application format such as: the title of the role or function, the specifications of responsibilities and the qualifications

required for the job. The following Figure 5.2 shows structured job template that will be used.

Company Information		
Contact Person		Company Name
Company Address		Company Phone
Company Email		Business Scope
Position Information		
Position Title		Department
Start Date		Salary
Personal Information		
Name		
Address		
Email		Phone Number
Experience		
Position		Duration
Education		
Degree		
Skills		
Skill	Level	Certificate
Other Information		
Posting Date		End Date

Figure 5.2: The structured job template

5.3. The implementation of the proposed system

The system is based on the analysis of the jobs within the criteria and scales, that determined by the employer to provide the largest number of candidates or jobs offers in the suggested lists. This makes the recruiting processes more transparent, accurate, realistic, and faster and save time and effort in finding suitable jobs.

Each section in job application is given a certain percentage of 100% depending on the importance of the section to the employer. Then we compare and match these sections with similar sections in the user profile, then give them suitable marks according to the weights given from employer. These ratios are collected for all sections together to be the final result to suit the job applicant for the job.

The system is divided into two models: user profile and Job profile. Also the system has two main flows, one for the employer and one for the applicants; the flows illustrated the using of job profile and user profile until extract the results.

5.3.1. User profile model:

Figure 5.3 depicts the applicant's CV processing flow, which can be summarized as follows:

1. The applicant signs in or register for an account to start finding job.
2. Fill the CV via a structured template according to the U.S. resume and CV type.
3. NLP analyzing and BabelNet dictionary will used to get bag of words.

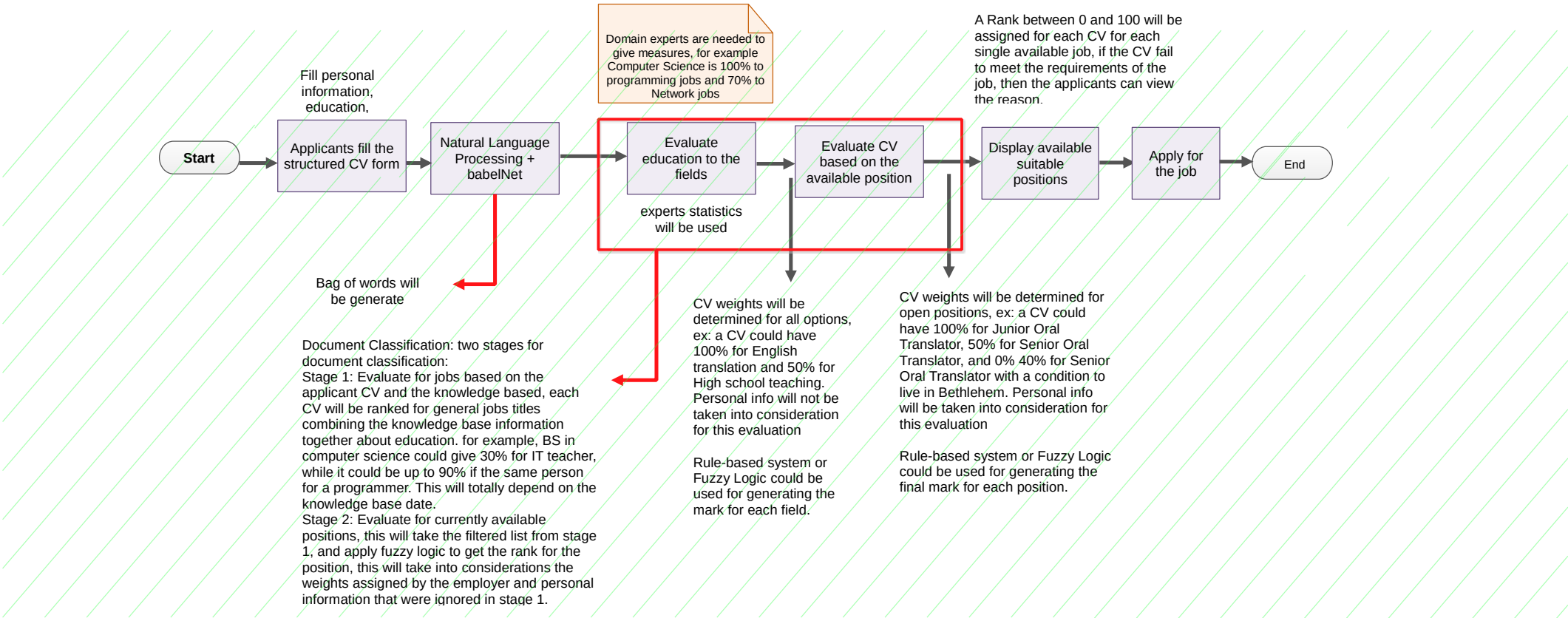


Figure 5.3: CV (Applicant) flow.

4. Evaluate the CV sections, first expert statistic used to evaluate the education section, then CV section evaluate by using content base algorithm according to the available position.
5. Fuzzy Logic will be used to generate results for each CV sections according to the job post sections.
6. Fuzzy Logic will be used to generate the final marks for each CV and available position.
7. Personal information section can be conditions for the employer as address or age.
8. The output of the user profile model is a list of applicants ranking according to the final matching ratio result.

5.3.2. Job profile model:

The job profile model created for business owners who are looking for employees, the employer creates a profile and add the vacant jobs, by selecting the field of job Category, and then adding the work requirements within specific groups (e.g. skills, education, experiences, personal information's, etc.), the employer show his actual needs by giving each section (skills, education, experiences)in the job form a percentage express, its importance according to the employer, the total of the percentages for three section must be100%.

After that the system returns to employer a list containing the names of the most suitable persons for the job with an appropriate ratio for each person. The employer can access the files of the candidates and access to their CVs.

To get the final list of candidates, the job profile model goes through several steps similar as the steps of the user profile model, using the same algorithms and methods.

NLP process will be used to analyze the text for each section in the job form, generating bag of words using babelnet dictionary, content base algorithm to find the similarity between job profile and user profile, expert knowledge database to evaluate the education section, and fuzzy logic mamadni type to calculate the similarity between the same sections in the job and the user profiles according to employer needs. Then get the final mark of hole cv (match) according to the available job position with similarity ratio. Figure 5.4 illustrates the Employer flow (Job flow).

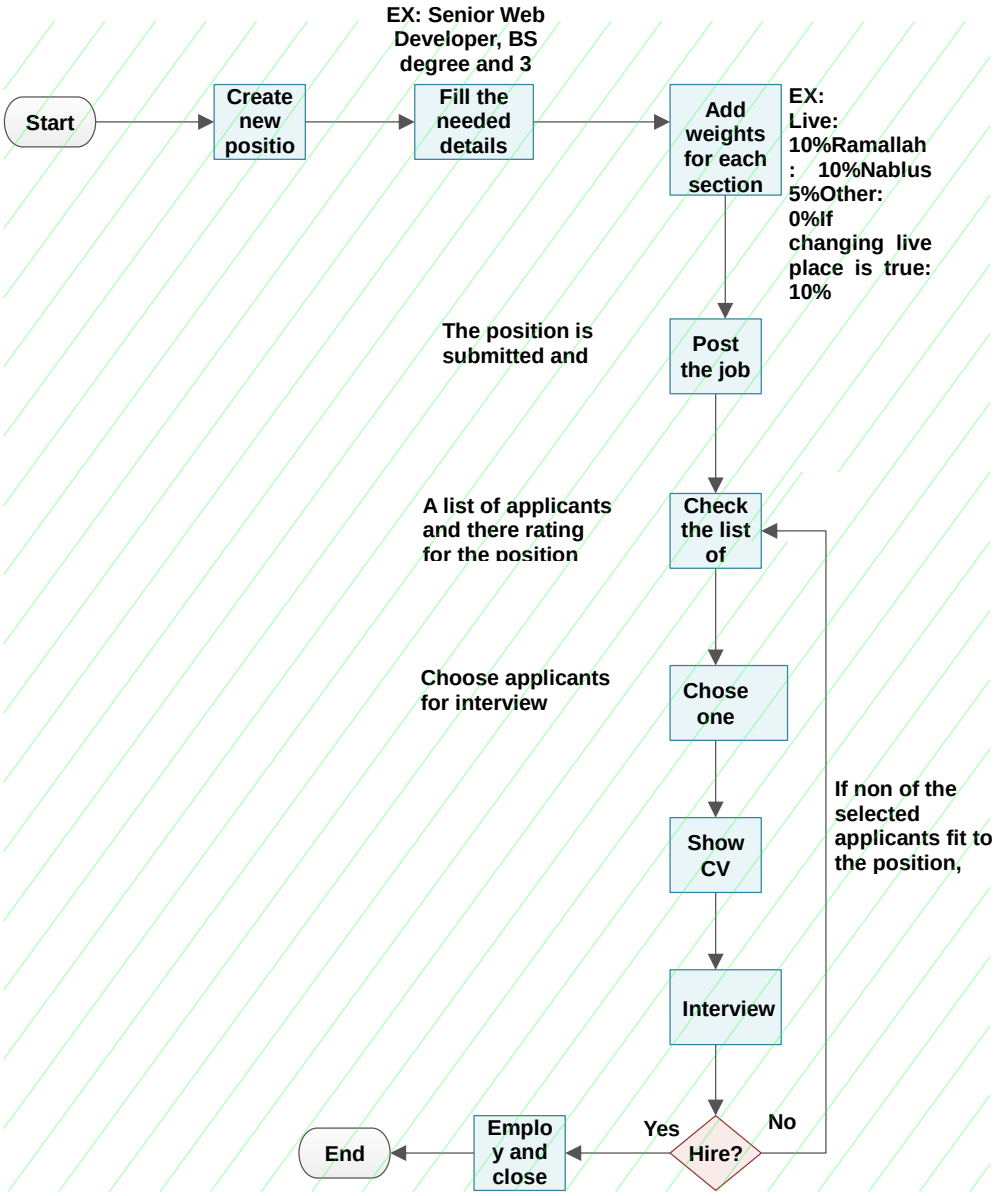


Figure 5.4: Employer flow (Job flow).

The flow will be as follow:

1. Start: Create account or sing in
2. Create new position or update the position
3. Fill the job position using job template.

4. Adding weights to each department: We allow the employer to add weights to each section in the job template while filling the job position (education - experiences - and skills). These weights represent the actual and real needs that the employer wants to have in the employee; the sum of these weights is 100%.
5. Submitted the job position to it by applicants.
6. A list of ranking applicants and there rating for the position will be displayed for the employer.
7. The employer can choose applicants for interview
8. Also, the employer can see their CVS, and make the interview and decide hire the applicant or choose another one from the list.

5.3.3. CVs Positions Matching:

After the applicant fills the required fields, the system will calculate the matching for the available positions, the matching process will move in the following phases:

1. Complete Profile: the system will ensure that the applicant fill all of the fields, if any field is missed then the CV will be considered as incomplete and it will not continue to the next phase.
2. Experts evaluation: the system will evaluate the CVs based on education, experiences, and skills, as we mentioned, the education will be evaluated using experts' knowledge base; this will shrink the possible positions that each CV could be a candidate for.
3. Positions matching without personal conditions: the system will match the available positions for each CV referring to the position domain, and the CV domain; the matching will take into considerations the experience level, and all the other

professional factors. The matching will take the weights of each part into consideration and each part will be validated according to its weight.

4. Position Matching with personal conditions: the system will filter the results from step 3 according to the personal information, all the positions will be presented to the applicant in case they can edit their personal info or if they can improve them in the future to match the needed personal info, for example changing the living place or having a driving license.

5.3.4. CVs – Positions Ranking:

After the system filters the applicants' CVs according to the positions, the system will rank the positions for each CV, the Ranking process will move in the following phases:

1. Ranking the matched position with unmatched personal information: any position that match the CV but did not match the personal info will be ranked with zero, the applicant will be able to view the reasons to have a zero rank.
2. Ranking the matched position with matched personal information: the matching level is up to 100% based on the fuzzy logic result, the ranking will be presented form the most matching positions to the lowest ones.
3. Presenting results for employers: the employers will have a view for each position to check all the matched CVs with matched personal information only with their ranks.

5.4. Conclusion

We implemented our hybrid system, which based on the babel dictionary, content base algorithm and the fuzzy logic to obtain the best accuracy and most efficient results in e-recruitments systems.

And we use BabelNet dictionary to generate a bag of words and the content -based algorithm to get match results close to reality. Also, we used two specific and clear templates for the CVs and job applications, the forms divided into specific sections, so that both the user and the company can arrange their data in an understandable and unambiguous way. We allowed the employer to define his/her job needs accurately by dividing the job into categories (education, experience and skills), each category has a certain percentage in which the employer expresses the importance of these categories with regard to the vacant job in reality. Also, the process of matching between the user and the request for employment, with the percentage of user suitability for the job greatly reduces the lack of understanding or confusion, that may befall the user when applying for the vacant position.

The fuzzy Logic inputs are the outputs of the content base algorithm, after finding the similarity between user profile and job application by calculating the cosine between the two vectors, A and B. The two vectors A and B represent the tf-idf result of the bag of words for the same section in user profile, and job application for example (skills – skills).

The fuzzy logic implemented in two stages, as Mamdani fuzzy inference type to find a similarity between the job sections and the CV sections (education, experience, skills) according to the percentages set by the employer, that express his/her real needs in the

employee required to fill the job, and in calculating the center of Gravity (COG), to get the final result of the whole CV from 100%.

The system is suitable for any job, whether in the education sector, industry, or any of the different economic sectors. Also, it is not limited to university graduates, but it allows everyone to use the site easily, searching for a job in our proposed system is automatically done. when the user sign in his account, the offered job list appears with the suitability of the user for the jobs in percentage, according to his CV, So the user can apply for the jobs which matches with him.

Chapter 6:

Experiment Results and Analysis

6.1 . Introduction

To make sure that our proposed recruitment system will improve performance in making decision in appointment, we have carried out a number of different scenarios for testing the proposed system, using web application simulation. In these scenarios, we applied the same real data of candidates "real CVs" used in other recruitment systems for certain job, and tested it in our system using the same job position, to measure the effect of our proposed system in the final result, such as in effectiveness, reliability and accuracy.

6.2. Data Collection

The study needs different data to test and evaluate the proposed research hypothesis. We collected data in different ways and methods, such as interview and survey (questionnaire), all the data that we obtained are real data, and the data was collected as follows:

1. CVs: We have obtained from the General Personal Council all the resumes of actual applicants, who applied for java programmer job. The numbers of applicants are 207, so we reserved 207 CVs, and built accounts for every user, and submitted the CVs.
2. Job application: the chosen job is programmer, which was previously announced by the General Personal Council, the job contains the following information: educational, qualification, experience, tasks and responsibilities and skills, Figure 6.1 shows the job application appears in our system as

programmer

Palestinian general personnel council

Required Education: [computer science, computer engineering, computer information system, information and communication technology]

Required Experience: [Object Relational Mapping (ORM), mainly Hibernate Framework. - 0 years, Java Struts Model View Controller (MVC) Framework, Spring, XML, and XML Schemas. - 0 years, OOA/OOD and knowledge of design patterns. - 0 years, JavaScript, jQuery ,AJAX. - 0 years, SQL & Relational DBMS (Oracle / MySQL). - 0 years, Web standards, HTTP, HTTPS, HTML/HTML5,CSS3, and responsive design. - 0 years]

Required Skills: [English language -Very Good, work in a team-Very Good, work under pressure-Very Good, communication skill-Very Good, Time management skill-Very Good]

Figure 6.1: The job application

3. The survey: we create online survey to build an expert knowledge base. We used Google forms, and also, we used Microsoft word in a hard copy, so the survey could reach a largest target numbers, and as quickly as possible. The survey targets the lecturers in the Palestinian universities in the West Bank, to benefit from their experience, and suggestions in finding the best possible job matching for each of the proposed specializations in the computer field, which was chosen to be studied in the survey.
4. The survey was created to give appropriate rate (0 - 10), to express the suitability of specific specialization for a list of jobs. There are 7 specializations in information technology field that accredited by the ministry of Palestinian higher education, and taught at different Palestinian universities. These specializations are computer science, Computer Engineering, Software Engineering, Computer Information Systems, Information and Communication Technology, Networking and Information Security and finally Web & Multimedia Technology.

The jobs list was chosen from labor market. These jobs are the most requested at the present time, according to the needs of the Palestinian labor market with great progress in the field of technology and communications. These jobs are programs developer, Mobile App Developer, Web developer, Project manager, Computer programmer, Computer and

information systems managers, Computer systems analyst, E-learning engineer, Computer Networks Engineer, System Engineer, Systems Architect, Data Engineer, Database Administrator, and finally Information Security Analyst.

The scale from (0 -10) was chosen to determine the suitability between the specialization and the list of the jobs, the number (10) express that the job with the (10) rate is the most suitable job for the specialization (extremely matching), and the number (0) told us that the specialization is not suitable for the job and the applicant cannot fill the position completely, (not related).

The lecturers also can determinate the degree of each specialization, which are Master degree, Bachelor degree and diploma degree, before rating the suitability of job.

We analyze the data and extract the result using Google forms after entered all the data in the hard copy into Google forms. The results of the survey used to build the expert database.

6.2.1. Building the Expert knowledge base:

As we mentioned, building an expert knowledge base is our main contribution in our proposed system. At the beginning, we defined the expert as the person who have a master degree or PHD in computer science, or other approaches in computer fields and have an experience in teaching computer courses, or certificates in Palestine universities.

We collected all the computer specialization name and degree in all the universities and colleges in West Bank and Gaza at first, then we divided the computer specializations into three groups according the specialization degree: diploma, bachelors and master.

The study includes ten universities in west bank: (Birzeit University, Alquds University, Alquds Open University, An-Najah National University, Palestine Technical University

Kadoorie, Arab American University, Hebron University, Bethlehem University, Palestine Polytechnic University and Palestine Ahliya University).

We collected the number of the academic staff all over these universities in west bank, which is equal approximately 140 lecturers who have a master or PHD in computer science approaches. Then we started preparing the survey after determinate the computer specializations and the jobs list, by selection a stratified randomized sampling as the following:

1. Divide the community into categories or homogeneous groups or layers, in our study the layers in the university.
2. Select total number of layers.
3. Select size of each layer separately.
4. Determine numbers of individuals per stratum in the selected sample. The proportional distribution is used when each layer is represented according to its relative weight in the study population.

Population Study and sample:

The population size = 140, table 6. 1 summarized how we distributed the population in groups, according to each university academic staff number in the computer domain.

Table 6.1: The population distributing

University Name	Number of academic staff
Birzeit University	**24**
Alquds University	**13**
Alquds open University	**9**
An-Najah National University	**24**
Palestine Technical University Kadoorie	**7**
Arab American University	**18**
Hebron University	**8**
Bethlehem University	**6**
Palestine Polytechnic University	**23**
Palestine Ahliya University	**8**
	The sum = 140

We suppose that Margin of error (or confidence intervals) $\alpha = 0.05$.

1. Confidence level = $1 - \alpha$

   $= 1 - 0.05 = 95\%$

2. Z-Score = the number of standard deviations a given proportion is away from the mean. To calculate Z-Score we need to divide $\frac{\alpha}{2}$ to get the alpha level $= \frac{0.05}{2} = 0.025$, then we subtract $0.5 - 0.025 = 0.475$, finally used standard normal distribution/ Z-table [29] to find the value of Z-Score = 1.96 for the desired confidence level (95%).

3. We used Steven Thompson equation to calculate the sample size X from the population size as in the next formula:

$$X = \frac{N*P(1-p)}{[[N-1*(\alpha^2 \div z^2)]+p(1-p)]} \qquad [32] \text{ Eq. (6.1)}$$

$$X = \frac{140*0.5(1-0.5)}{\left[[140-1*\left(0.05^2 \div 1.96^2\right)]+0.5(1-0.5)\right]} = 103$$

Where

X: The sample size (103).

N: Population size (140).

Z: Confidence level at 95% (1.96).

α: Error of proportion (0.05).

P: standard of deviation (0.5).

4. We generate a simple random sample for each layer by using distribution method proportional. Sample size for each layer = $\frac{n}{N}$X. [22] Eq. (6.2)

The equation variable n: The sample size in table 3.4.

N: the population size = 140

X: the sample size of the population= 103

e.g.: the sample size of Alquds University = $\frac{13}{140}$x 103 = 9.56 ≅ 10

Table 6:2 summarized the results of sample size for each university.

Table 6.2: The sample size for each university

The university	The sample size distribution
Birzeit University	18
Alquds University	10
Alquds open University	6
An-Najah National University	18
Palestine Technical University Kadoorie	5
Arab American University	13
Hebron University	6
Bethlehem University	4
Palestine Polytechnic University	17
Palestine Ahliya University	6
	The sum = 103

Some universities refused to fill out the survey, so we turned to the labor market to cover the shortage of the sample. Table 6.3 the results of sample size for each university.

Table 6.3: The sample size for each group

The groups	The sample size
Birzeit University	18
Alquds University	6
Alquds Open University	6
An-Najah National University	18
Palestine Technical University Kadoorie	5
Arab American University	0
Hebron University	6
Bethlehem University	4
Palestine Polytechnic University	0
Palestine Ahliya University	6
Labor market	**34**
	The sum = 103

5. To find the Confidence Interval for Computer Programmer and specializations, we used this formula:

$\bar{X} - t_{1-\alpha/2\,;\,n-1}\, s/\sqrt{n}\,,\quad \bar{X} + t_{1-\alpha/2\,;\,n-1}\, s/\sqrt{n}\,,$ Eq. (6.3)

To find the value of t, we needs to calculate the degrees of freedom = Subtract 1 from sample size. 103 – 1 = 102

$T = 1 - \alpha/2$
1- 0.05/2 = 1- 0.025
t = 0.975

The critical points from t table $t_{0.975,102} = 1.984$

We also need to find s which equal, $S^2 = = \frac{\sum(x-\bar{x})^2}{n-1}$ Eq. (6.4)

The confidence interval for computer programmer and computer science

$\bar{X} - t_{1-\alpha/2\,;\,n-1}\, s/\sqrt{n}\,,\quad \bar{X} + t_{1-\alpha/2\,;\,n-1}\, s/\sqrt{n}\,,$

8.77 - 1.984 x 1.6073103/ √103 , 8.77 + 1.984 x 1.6073103/ √103

8.77 -1.984 x 1.6073 / 10.14889 , 8.77 + 1.984 x 1.6073103/10.14889

Table 6.4: calculate the $\bar{X}$ and S.

Table 6.4: calculate the $\bar{X}$ and S

job		$\bar{X}$	**S**
Computer programmer	Computer science	8.77	1.6073103
	Computer Engineering	7.98	2.04236257
	Software Engineering	8.44	1.59757938
	Computer Information Systems	7.85	2.11365382
	Information and Communication Technology	6.67	2.27778658
	Networking and Information Security	6.71	2.52117304

Knowledge base will be implemented, and will be used in the proposed system. The knowledge base will contain information gathered from the survey at first, then we will updated knowledge base by getting feedback from labor market, We will ask the employer to give his/her opinion about the employee' specialization after a while from his/her employment for the position. This feedback will affect the expert's knowledge database, and make it more reliability. Figure 6.2 illustrated the generation of the expert knowledge base. The knowledge database contains data extracted from the survey, and the result of the knowledge database is the average of the lecturers rating, which describes the suitability for specialization to job position.

We use the knowledge experts' database in two cases: needs revision

1. Before employment: to evaluate the specialization according to the job title.
2. After employment: employers can evaluate or review their employee an according to their specialization also, to link the knowledge database with the Labor market, it's a very important step for universities, to develop the courses and create new plans that meet the real needs of the labor market.

we will use the employer's evaluation to update the knowledge database to be more related to the real recruitments process, this update and feedback will have affective the knowledge database as 10% only in the total result, which generate from the survey, and in both positive or negative cases.

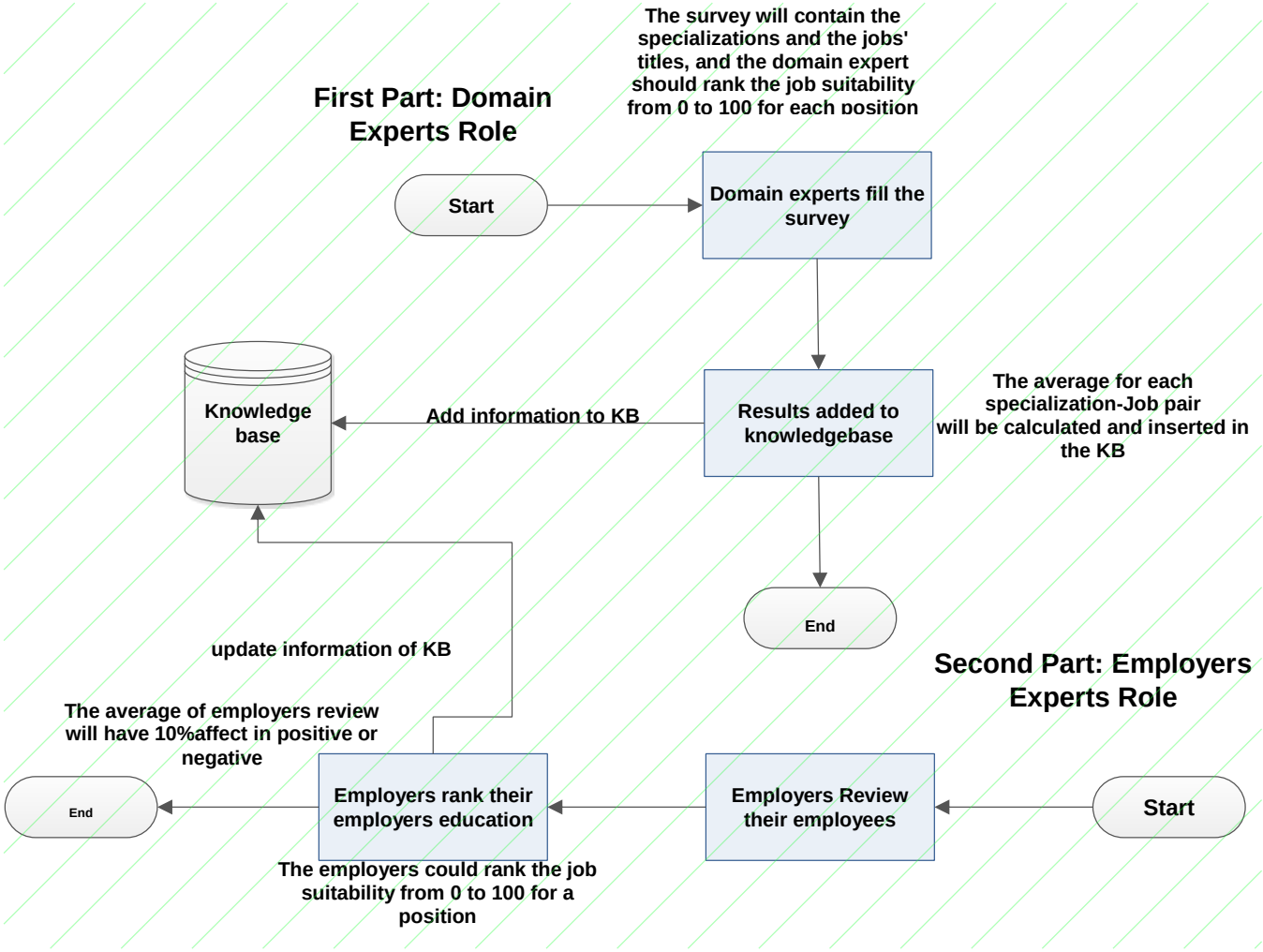


Figure 6.2: Expert knowledge base generation from the survey result and employer review.

Figure (6.3) illustrating the building of the knowledge base ER diagram.

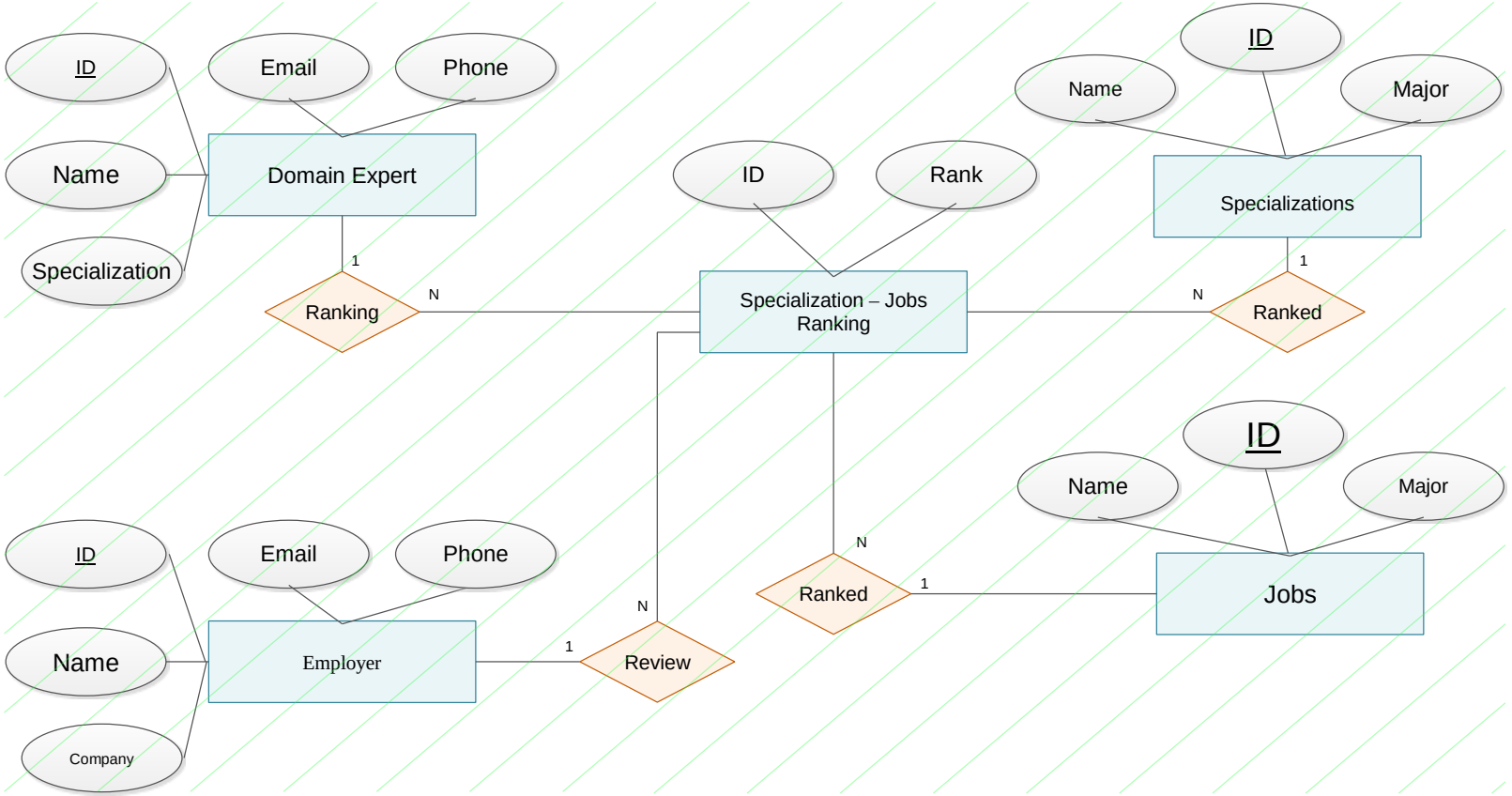


Figure 6.3: ER diagram for the expert knowledge base.

6.3. The proposed system experiments

6.3.1. The home page:

The Home page: you can register in as user, or company, Figure 6.4 shows the proposed system home page.

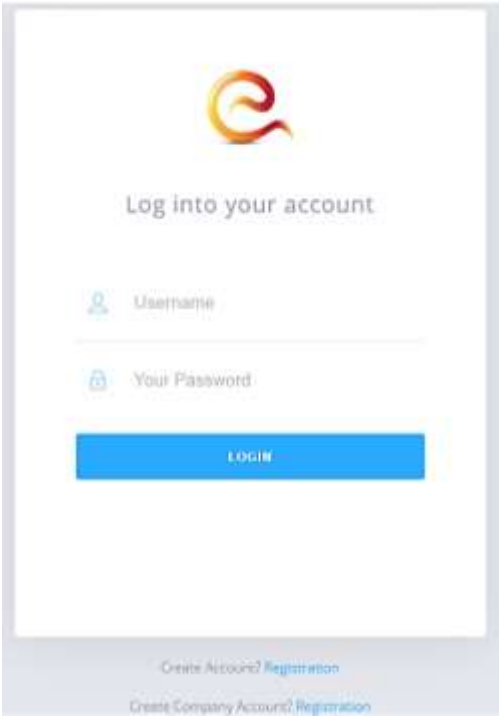


Figure 6.4: The proposed system home page

6.3.2. The user profile:

- After singing we can fill and submit the structure CV, or log in and update it. Figure 6.5 shows the CV structure form.

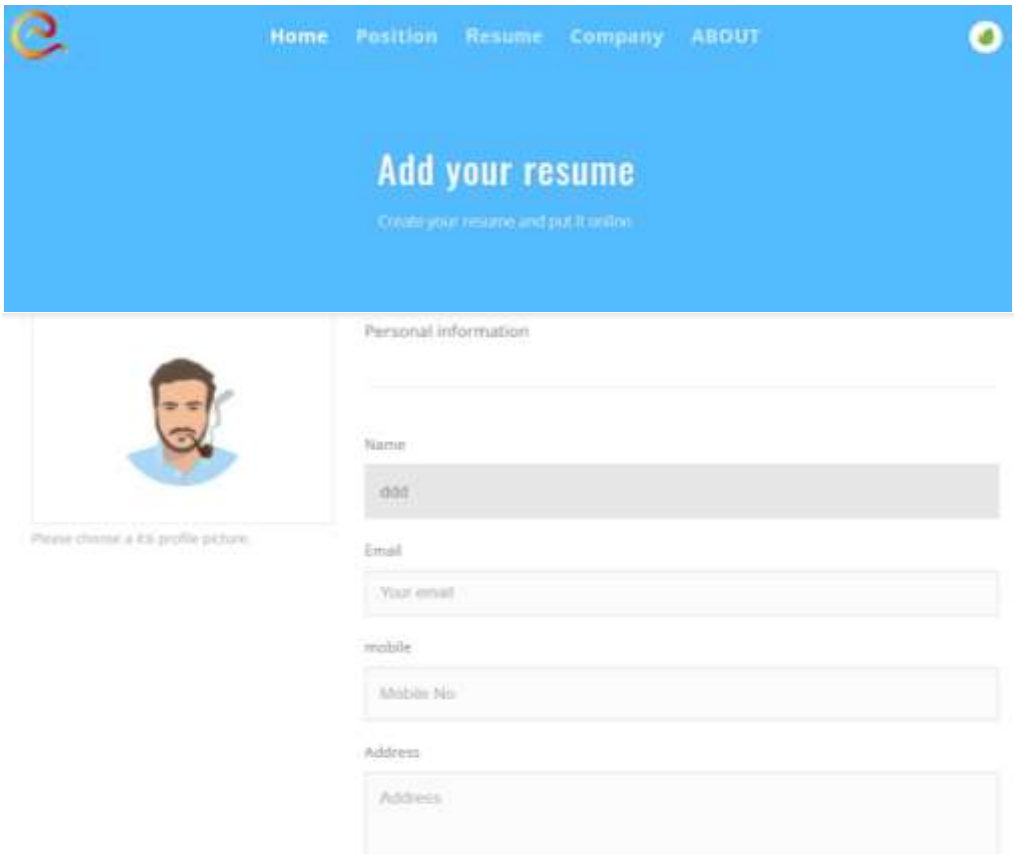


Figure 6.5: CV structure form.

- After filling the CV template, a list of jobs will appear to the user with his/her suitability to the jobs that posted by the companies which created accounts on the web site. The match ratio calculated according to user CV, in this step the sytem uses the expert database, which contains the survey result to evaluate the education section.Table 6.5 and table 6:6 summarize the results of the survey for a programmer job, and computer science specialization as an example of the survey results as

1. The suitability of different specialization for computer programmer job. Table 6.5 shows the survey results (expert's opinion average) for a programmer job.

Table 6.5 the survey results for a programmer job

Job	Specialization	Degree	Rate	Average
Computer programmer	Computer science	BA	10	8.77
	Computer Engineering	BA	10	7.98
	Software Engineering	BA	8	8.44
	Computer Information Systems	BA	6	7.85
	Information and Communication Technology	BA	7	6.67
	Networking and Information Security	BA	6	6.71

2. The suitability of computer science specialization for different jobs. Table 6.6 the survey results for a Computer science specialization.

Table 6.6: The survey results for a Computer science specialization

Specialization	Job Title	Degree	Rate	Average
Computer Science	Program developer	BA	10	8.89
	Mobile App Developer	BA	10	8.57
	Web developer	BA	10	8.47
	Project manager	MS	8	8.02
	Computer programmer	BA	10	8.77
	Computer and information systems manager	MS	8	8.32
	Computer systems analyst	BA	7	7.85
	Computer Networks Engineer	BA	6	6.97
	System Engineer	BA	8	6.77
	Systems Architect	BA	6	6.92
	Data Engineer	BA	8	7.29
	Database Administrator	BA	8	8.40
	Information Security Analyst	BA	8	7.58

- The user can apply for the jobs he/she matches with in good ratio. Figure 6.6 shows the list vacancies are arranged according to the most suitable job for candidate CV.

Figure 6.6: The list of jobs according to the user CV

6.3.3. Job profile:

- The employer will be able to post a job in the structured form, and fill the sections (education, experiences and skills) with the suitable ratio express, the actual needs of the employer, and the required qualification of the applicant according to the new employee.

Figure 6.7shows job application form, and how the employer can determine his/her most important section in job application, by adding suitable percentage.

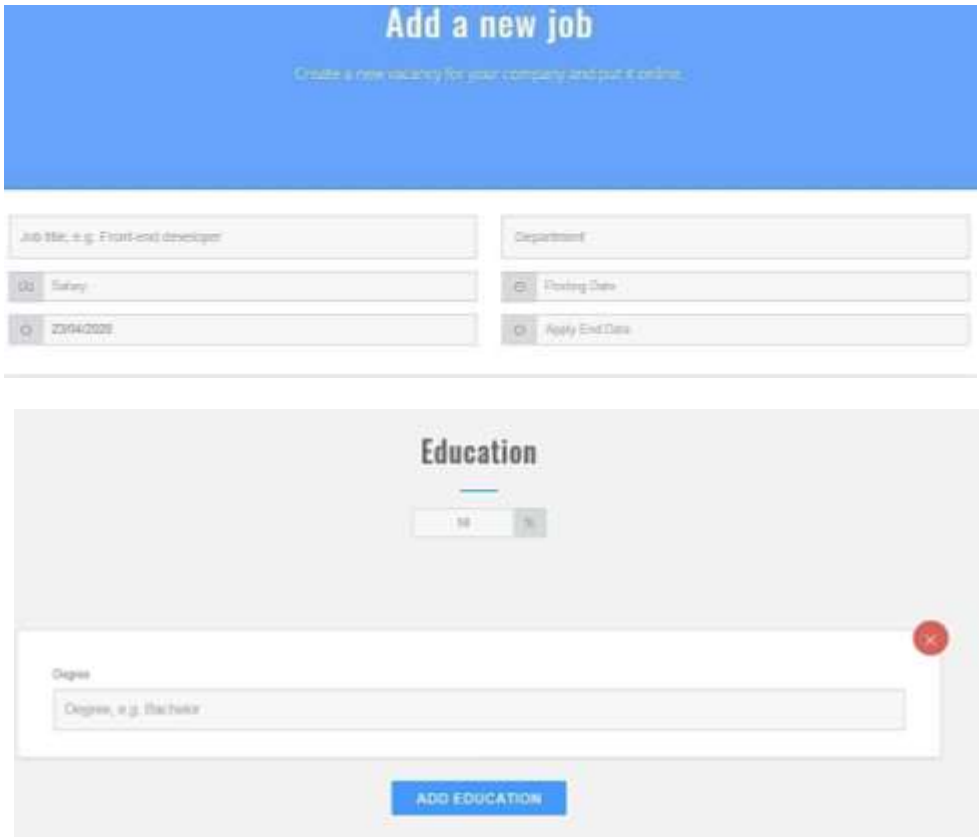


Figure 6.7: job application

- Manage jobs: the employer can manage his/her job after posting it, by add, edit or delete the job. Figure 6.8 shows manage operations the employer can use it.

Figure 6.8: the manage job operations.

- View candidate: a list of candidates with matching ratio will appear for the employer, also we can find another matching ratio for each section of the CV

individually, and the employer can view the candidate CV and connect with him/her, if he/she matches with the job.

- The system let the employer evaluate the employee during his/her work after the employment. The evaluation scale rating the employee from 0-10, according to his/her specialization. Figure 6.9 shows the candidate list with evaluation rate.

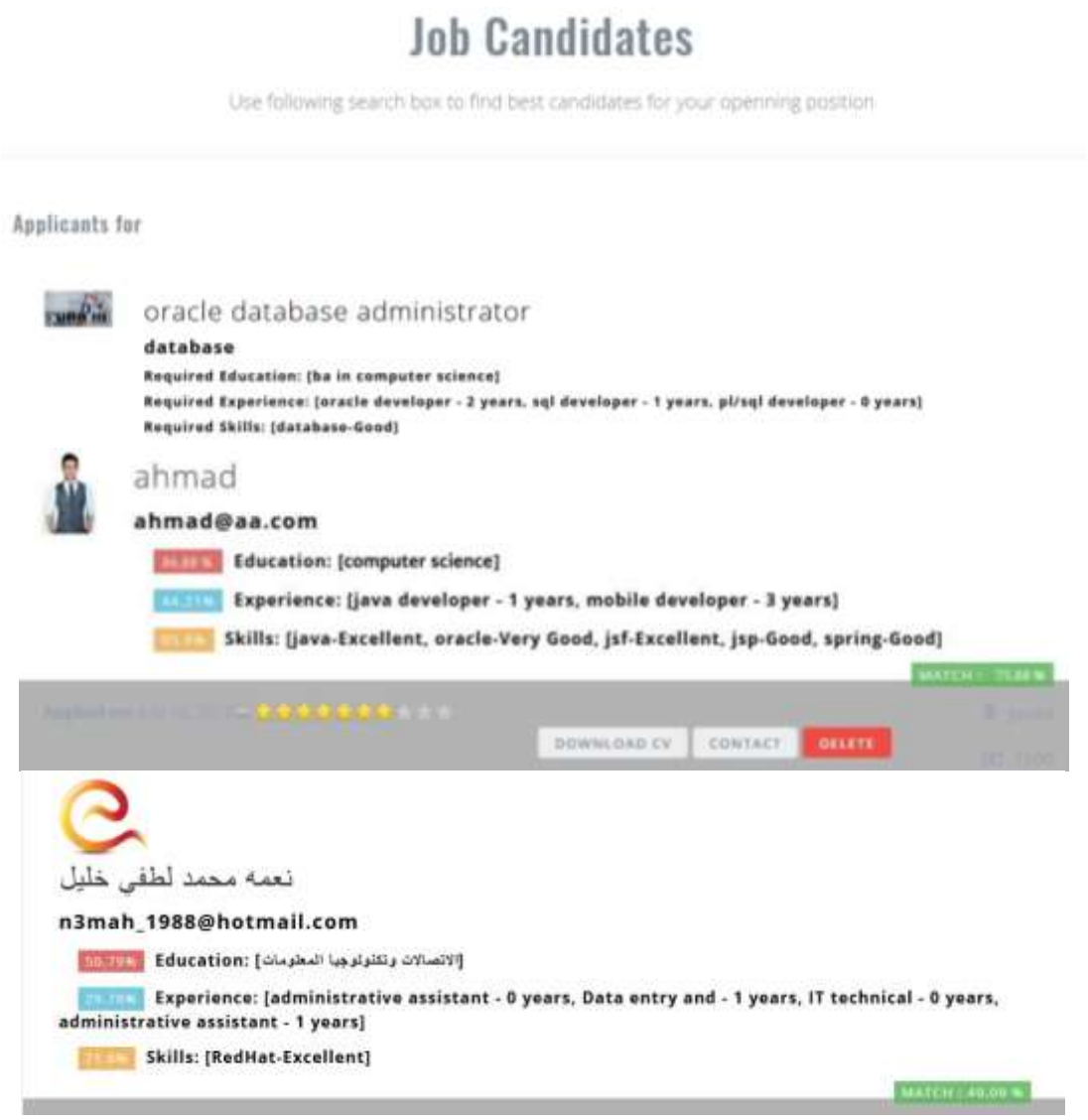


Figure 6.9: the candidate list and the evaluation rate.

6.4. Performance Evaluation

To evaluate our proposed system, we will compare it with General Personal Council recruitment system components, and the matching process between CVs and the job applicant. Also, we will discuss the results, and the difference between the two systems, and will make a comparison between our results and the actual results of the job.

The chosen job to be practically applied was a programmer job, which was previously announced by the General Personal Council, the job contains the following information: education, qualification, experience, tasks, responsibilities and skills.

The CVs we obtained are all real CVs of applicants who applied for the programmer job, the number of applicants was 207, including the entire candidates who Accepted or rejected out of the job. So, we tested our system using all these CVS.

6.4.1. General personnel council recruitment system:

General personnel council recruitment system divides into two parts:

1. Job application:

- Post a job: Each job has a specific number in the system, and to post a job, an agreement between the general personnel council, and the job advertiser made to determine the contents of the employment announcement, including education, qualifications, experience and other conditions.
- The needed education for the position is determined by agreement, with considering all similar specialization to be within the list of specializations that are allowed to apply the job, the similar specialization list provided from the Ministry of Higher Education and Scientific Research, if this list was founded.
- A table was created with names and numbers for all similar and acceptable specializations related to a certain position using ORACLE database.

2. User profile**:** you need to following steps:

- Create a new account to submit employment applications
- Confirm the identification number: you will be directed to the ID entry page, to confirm if the ID number is already used and you have a previous account. If the ID number is not available, it will be transferred to the data fill page. If the number is available, a message will appear to modify the password.
- Submitted the CV, you will receive a SMS on your phone, contains the user name and the password.
- Sign in: to verify the user name and the password
- Fill the CV template, and attaching education and experience certificate as files in PDF or jpg format.
- To apply for a job application through the general personnel council, whether it happens through a website or other means of advertisement that the applicant believes that the job is suitable for him, sign in the account and press the button "submit employment application".
- The candidates choose their specialization from drop-down list.

General personnel council recruitment system gives the results of the applicants rejected or accepted, according just to specialization. The system depends on the academic specialization in the selection process of candidates for vacancies jobs as the only condition. If the specialization of a candidate match with the specialization required on the advertising, the candidate move to the next step to the interview, and the employers look into the experiences and skills sections, and if his/her specialization does not match, he/she will be rejected.

6.4.2. General personnel council recruitment system result vs. our result for the programmer job:

The general personnel council recruitment system results for the programmer job was a list of acceptable candidates' number, according to their specialization which was 139 candidates. These candidates will be allowed to enter a practical exam and interview. The pass mark is 65% and above in the practical exam, and less than that does not enough transfer candidate to reach to the interview. The last list of candidates contains 3 candidates only who pass the practical exam.

Testing our proposed system, which relies on evaluating the specialties of applicants using the expert database, in order to measure the appropriateness of similar specializations for the different jobs? After using the content base algorithmic fuzzy logic, we used the same job issued application of programmer by the general personnel council, which contained the following specialties to fill the position: Computer science, computer engineering, information technology and communications and computer information systems.

We tested all the CVs (207 CVs) after assigning the 100% percentage to the Education section in the job form, and neglecting the rest of the CV sections by giving it 0%, thus the percentage of matching CVs to the job depends on the percentage of specialization and only equal it, as the general personnel council works. Accordingly, we got a list of acceptable candidates for the job, which contained 123 candidates, with different percentage for each candidate CV according to his/her specialization. The acceptable list includes the CVs that matched with the job percentage 100% to 50%. 77 candidates of 123 are already on general personnel council list.

Since the general personnel council sets the success rate at 65%, this reduces the number of those who were accepted to the position in the list to only 45 names.

The table 6.7 summarizes the results of matching the specializations in the job application and the specializations of the applicants in descending order, from most suitable to least suitable.

Table 6.7: The results of matching specializations
in job application with candidates CV.

Job advertising Specialization	**Matching Ratio**
علم حاسوب	86.88 %
هندسة حاسوب	86.88 %
تكنولوجيا معلومات واتصالات	63.66%
انظمة معلومات حاسوبية	54.95%

Table 6.8 also summarized the result to the same job posted (programmer), but showed the result in details, for the first 10 candidate just according to the education section, so the education section percentage will be 100%.

Table 6.8: the of candidate list for programmer job according to the education percentage.

Candidate Name	Candidate Education	Education percentage = 100%	Final match ratio
انسام نعيم ربيع	هندسة حاسوب	86.88%	86.88%
يوسف طوافشة	علم حاسوب	86.88%	86.88%
نضال عاصي	علم حاسوب	86.88%	86.88%
ميس نفاع	علم حاسوب	86.88%	86.88%
ماهر ميناوي	علم حاسوب	86.88%	86.88%
فاطمة جرار	علم حاسوب	86.88%	86.88%
علاء الخباص	علم حاسوب	86.88%	86.88%
طريف اشتيه	علم حاسوب	86.88%	86.88%
ضياء كوازبه	علم حاسوب	86.88%	86.88%
شعبان عبدالعال	علم حاسوب	86.88%	86.88%

6.5. Result discussion

All the specializations in the job advertising have the same importance, and they are equal, although each specialization has different name and courses. This is the fundamental difference between general personnel council recruitment system, and our proposed system which shows the difference of suitability degree of different specializations for the same job, according to a survey that we carried out, and we targeted the lectured in universities and some companies in the labor market.

The candidates were listed in alphabetical order, and in random order according the accepted in the job, and specialization. While our candidates list orders candidates according to their suitability of job, the candidate list returns to the employer with match

ratio, depends on conditions he/she chooses as percentage for each section in job application, to achieve his/her real needs to choose the best candidate to his/her work.

When we compared the results of our system with the general personnel council results, we found that our list contains fewer candidates than the general personnel council list , our list contained 123 accepted candidates, 77 of them are already on the general personnel council list, while the general personnel council list contains 139 for the same job, also our list ranked the candidates in descending order, based on their specialties with proportions Percentage from most suitable specialty for the job to least suitable one, these results emerged based on a comparison of the applicants' specializations with the specializations in the expert knowledge database that we generate in our proposed system. Each specialty in the database has a set of jobs and percentages that express the suitability of the specialization with the proposed jobs. If we consider the percentage 65% to be the pass mark as the general personnel council, then the number of accepted candidates will be 45which make it a very specific list.

Our system provides a list of accepted candidates to the employer in a new form, where the new list divided into 3 sections, which are (education, experiences and skills), each section contains a percentage of its match with the corresponds section in the applicant's CV, with a brief information about each section. This gives the employer a comprehensive clear brief overview of the applicant, In addition to the total percentage of CV matching with the applicant.

Also, our proposed system deals with all specialization, that are not mentioned in the job advertisement, and gives them a similarity rate that may be high, or allow their holders to access the list of practical exam or interview above (50% - 65%).

Table 6.9 list the specialization that are not included in the job application, and they got match ratio within 50%, which are:

Table 6.9: Another specializations not in advertisement,
and have match ratio within 50%.

Candidate specialization	Matching Ratio
حاسوب	86.5%
هندسة البرمجيات	78.14%
هندسة وتكنولوجيا الاتصالات	67.09%
تكنولوجيا معلومات حاسوبية	61.06%
هندسة نظم حوسبة وشبكات	56.25%
نظم معلومات حاسوبية	56.86%
هندسة انظمة الحاسوب	54.71%
الحوسبة التطبيقية	50.06%

In order to make the differences between our system and the recruitment systems mentioned in related works and the general personnel council system clear, and To talk more about our system, we retest the same job (the programmer) and the same CVs, by distributing the percentage 100% to all sections of the job application, according to the employer needs. Figure 6.10 shows the job application with the required qualifications.


programmer

Palestinian general personnel council

Required Education: [computer science, computer engineering, computer information system, information and communication technology]

Required Experience: [Object Relational Mapping (ORM), mainly Hibernate Framework. - 0 years, Java Struts Model View Controller (MVC) Framework, Spring, XML, and XML Schemas. - 0 years, OOA/OOD and knowledge of design patterns. - 0 years, JavaScript, jQuery ,AJAX. - 0 years, SQL & Relational DBMS (Oracle / MySQL). - 0 years, Web standards, HTTP, HTTPS, HTML/HTML5,CSS3, and responsive design. - 0 years]

Required Skills: [English language -Very Good, work in a team-Very Good, work under pressure-Very Good, communication skill-Very Good, Time management skill-Very Good]

Figure 6.10: The job application.

Table 6.10 shows the job application sections with their percentages.

Table 6.10: The sections percentages

Education percentage	Experience percentage	Skills percentage	The CV Percentage
50%	30%	20%	100%

We represented the results as they appeared through our proposed system. Table 6.11shows the results for the first ten candidates, and shows the clear difference between the previous results and the current results after changing the matching ratios for the job application sections.

Table 6.11: The match percentages results and the final match ratio

Candidate Name	Education	Education percentage = 50%	experiences percentage = 30%	Skills percentage = 20%	Final match ratio
احمد ابو عرة	علم حاسوب	87.81	89.26	86.36	87.95
مراد الدقة	أنظمة معلومات حاسوبية	81.09	71.73	90.46	80.10
معتصم حمامرة	نظم معلومات حاسوبية	71.44	49.74	93.13	69.26
اسيد عاروري	علم حاسوب	87.81	26.73	84.6	68.84%
روان ابو بكر	علم حاسوب	87.81	19	91.15	67.83%

The results shows that our proposed system provides an opportunity for applicants who were not accepted in the programmer job, when the education section was evaluated from 100%, or had a low match ratio to be accepted or had a good chance to employ , after we distribute the percentage of the job application to gather all the sections, each percentage represents the importance of the section Affiliate it, and determines the importance of the section in relation to other sections and expresses employment needs, the accepted candidates list also changed, the results appear with match ratio of each section, and the final result of the CV matching ratio with the job.

As for the hypotheses mentioned at the beginning of the study, we compared the results of our proposed system that we achieved with them, as follow:

The first hypothesis that we approved is the effect of the expert Knowledge database on the recruitment process. It plays an important role in organizing and arranging university specializations in the domain of computers, according to the appropriate jobs for them, and opens the opportunity for other specializations in the same domain to compete, and not excluding them. These results are also included in the survey.

We approved the reflecting of employer's view in job application, and employee qualifications, and their impact in recruitment process, through the system results which appear in the applicants' recommendation list and the feedback of the export knowledge database from the employee evaluating process from the labor market.

Our proposed system classifies and arranges the applicants for a job according to employer needs, by dividing the job application into sections, and the employer can give these sections a percentage.

We used a structured CV with a structured job application to facilitate the ranking and matching processes. The results are lists of applicants according to the weighting of applicants 'CVs based on the employer needs, which enhance the employment process results, and that are our hypotheses.

6.6. Precision, Recall and F measure

1. Precision = $\frac{|\{\text{relevant dcuments} \cap \text{retrieved dcuments}\}|}{|\{\text{retrieved dcuments}\}|}$ Eq.(6.5)

In General personnel council recruitment system results, relevant documents = relevant CV which is 207 CV, and retrieved documents = retrieved CV which is 139. Precision = $\frac{|\mathbf{207} \cap \mathbf{139}|}{|\mathbf{139}|} = 1$

In our system Precision $= \frac{|139 \cap 123|}{|123|} = \frac{|77|}{|123|} = 0.62601$

2. Recall = $\frac{|\{\text{relevant dcuments} \cap \text{retrieved dcuments}\}|}{|\{\text{retrieved dcuments}\}|}$ Eq.(6.6)

For General personnel council recruitment system = $\frac{|207 \cap 139|}{|207|}$

$= \frac{|139|}{|207|} = 0.6714$

In our system Recall = $\frac{|139 \cap 123|}{|139|} = \frac{|77|}{|139|} = 0.55395$

3. F mausare: 2 x $\frac{|\text{Precision x Recall}|}{|\text{Precision} + \text{Recall}|}$ Eq.(6.7)

For General personnel council recruitment system F measure = 2 x $\frac{|1 \text{ x } 0.6714|}{|1+0.6714|}$

$= 2\,x\ \frac{|0.6714|}{|1.6714|} = 2\,x\ 0.4016 = 0.8032$

In our system F measure $= 2 \text{ x } \frac{|\text{Precision x Recall}|}{|\text{Precision} + \text{Recall}|}$

$$= 2 \text{ x } \frac{|0.62601 \text{ x } 0.55395|}{|0.62601+0.55395|} = 2 \text{ x } \frac{|0.34677|}{|1.1796|}$$

$$= 2 \text{ x } 0.29464 = 0.58929$$

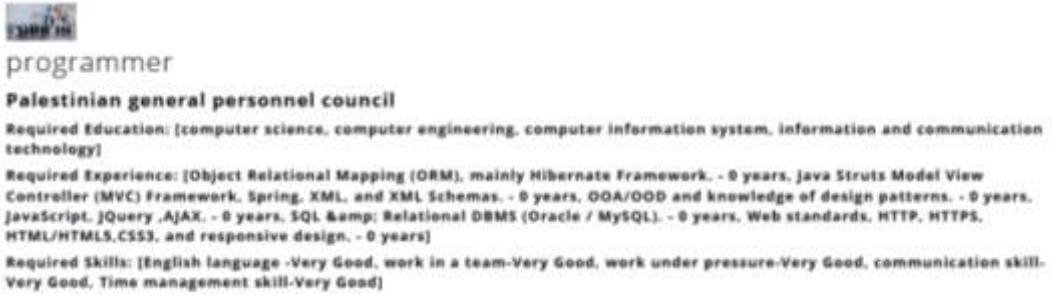
6.7. Case Study:

6.7.1. Candidate CV: we choose cv candidate as the Figure 6.11 shows:

Personal Information		
Ahmad abu arrah	2/8/1992	
Address: tobas		
A.a.abuarrah@gmail.com	0597552887	
Education		
QOU	Information and Communication Technology	2015-2000
Experience		
IREX	administrator OCA	2017
Tobas sport club	Web design, HTML5 and CSS3	2017
Tobas sport club	Advanced Excel trainer	2017
Taawon for Conflict Resolution	Conflict Resolution trainer	2018
AMIDEAST	scratch programing language trainer	2018
Skills		
Young Leaders in Civic Engagement	Excellent	
General Training Skills Workshop.	Excellent	
ICT Skills for Everyday Training-Of-Trainers-Workshop	Excellent	
Tamheed Advisor	Excellent	
Most significant change (MSC)	Excellent	
Other Information		

Figure 6.11: Candidate CV

6.7.2. Job application: which presented in Figure 6.12.


programmer

Palestinian general personnel council

Required Education: [computer science, computer engineering, computer information system, information and communication technology]

Required Experience: [Object Relational Mapping (ORM), mainly Hibernate Framework. - 0 years, Java Struts Model View Controller (MVC) Framework, Spring, XML, and XML Schemas. - 0 years, OOA/OOD and knowledge of design patterns. - 0 years, JavaScript, jQuery ,AJAX. - 0 years, SQL & Relational DBMS (Oracle / MySQL). - 0 years, Web standards, HTTP, HTTPS, HTML/HTML5,CSS3, and responsive design. - 0 years]

Required Skills: [English language -Very Good, work in a team-Very Good, work under pressure-Very Good, communication skill-Very Good, Time management skill-Very Good]

Figure 6.12: programmer job application

1. **NLP**

   Using bag of word model to analyzing the text in each section in cv and job

2. **BableNet**

   Using bableNet to return Synonym words for the words in each section in the CV and job application. Table 6.13 and table 6.14 present bag of words for User experience section in the cv and job app.

Table 6.12: bag of words for User experience

User experience	**BOW**
Web design, HTML5 and CSS3	sheaf theori topo websit site comput occup manag webmast html markup languag open format technic communic world wide web consortium standard css cascad style sheet stylesheet typeset program design broken stream cipher compact disc dvd copi protect digit right histori cryptographi televis content scrambl system
scratch programing language trainer	educ train occup posit author superintend facil school comput program scratch
administrator OCA	comput occup system administr oca

Table 6.13: bag of words for position experience

Position experience	BOW
Java Struts Model View Controller (MVC) Framework, Spring, XML, and XML Schemas.	java comput platform cross softwar class base program languag concurr famili specif request jvm object orient creat static type sun microsystem brand name product stub introduc soviet cigarett strut model graphic imag visual effect architectur pattern scienc design mvc view control peripher element user interfac placehold technolog widget establish unit state missouri valley confer sport midwestern organis framework fabric dos offic suit microsoft applic develop tool free net spring microkernel oper system proprietari aspect enterpris web xml layer protocol bibliographi file format data serial markup open present technic communic world wide consortium standard schema manag databas
Web standards, HTTP, HTTPS, HTML/HTML5 ,CSS3, and Responsive design.	web softwar free program browser gnome core applic os posix base webkit fork use meson gtk document generat liter engin stub tex introduct british invent cern english human comput interact inform age world wide standard http layer protocol hypertext transfer internet network wikipedia pend chang protect page consortium https cryptograph secur communic transport uri scheme html markup languag open format technic css cascad style sheet stylesheet typeset design broken stream cipher compact disc dvd copi digit right manag histori cryptographi televis content scrambl system respon

3. Use TF- IDF to transform the bag of words into vectors:

```
public double tfCalculator(List<String> totalterms, String termToCheck) {

double count = 0;    //to count the overall occurrence of the term termToCheck
for (String s : totalterms) {
if (s.equalsIgnoreCase(termToCheck)) {
                count++;
            }
```

```
        }
        return count / totalterms.size();
    }

    public double idfCalculator(List<List<String>> allTerms, String termToCheck) {
    double count = 0;
    for (List<String> ss : allTerms) {
    for (String s : ss) {
    if (s.equalsIgnoreCase(termToCheck)) {
                            count++;
                            break;

                    }
                }
        }

    return 1+ Math.log(allTerms.size() / count);
    }
```

4. Calculate the similarity using cosine equation:

```
public double cosineSimilarity(double[] docVector1, double[] docVector2)
   {
     double dotProduct = 0.0;
     double magnitude1 = 0.0;
     double magnitude2 = 0.0;
     double cosineSimilarity = 0.0;
// System.out.println("vectors : " +docVector1.length + " - " +
docVector2.length);
for (int i = 0; i < docVector1.length; i++) //docVector1 and docVector2 must be of
same length
          {
dotProduct += docVector1[i] * docVector2[i];    //a.b
                magnitude1 += Math.pow(docVector1[i], 2);    //(a^2)
```

```
            magnitude2 += Math.pow(docVector2[i], 2); //(b^2)
        }

        magnitude1 = Math.sqrt(magnitude1);//sqrt(a^2)
        magnitude2 = Math.sqrt(magnitude2);//sqrt(b^2)

        if (magnitude1 != 0.0 | magnitude2 != 0.0)
        {
            cosineSimilarity = dotProduct / (magnitude1 * magnitude2);
        }
        else
        {
            return 0.0;
        }
        return cosineSimilarity;
    }
```

5. Mamdani fuzzy logic:

a.

```
FUZZIFY edu                // Fuzzify input variable edu': {'poor', 'good' , 'excellent'}
TERM lowEdu := (0,1) (25, 1 ) (50, 0) ;
TERM medEdu := (40, 0) (72,1) (85,0);
TERM highEdu := (80, 0) (92,1) (100,1);
END_FUZZIFY


FUZZIFY skill    // Fuzzify input variable 'skill': {'poor', 'good' , 'excellent'}
    TERM lowSkill := (0,1) (25, 1 ) (50, 0) ;
    TERM medSkill := (40, 0) (72,1) (85,0);
TERM highSkill := (80, 0) (92,1) (100,1);
END_FUZZIFY
```

FUZZIFY exper // Fuzzify input variable 'exper': {'poor', 'good' , 'excellent'}

TERM lowEx := (0,1) (25, 1) (50, 0) ;

TERM medEx := (40, 0) (72,1) (85,0);

TERM highEx := (80, 0) (92,1) (100,1);

END_FUZZIFY

b. If-Then Rolse:

AND : MIN; // Use 'min' for 'and' (also implicit use 'max' for 'or' to fulfill DeMorgan's Law)

ACT : MIN; // Use 'min' activation method

ACCU : MAX; // Use 'max' accumulation method

RULE 1 : IF edu IS lowEdu and skill IS lowSkill and ex is lowEx THEN tip IS low;

RULE 2 : IF edu IS lowEdu and skill IS lowSkill and ex is medEx THEN tip IS low;

RULE 3 : IF edu IS lowEdu and skill IS lowSkill and ex is highEx THEN tip IS med;

RULE 4 : IF edu IS lowEdu and skill IS medSkill and ex is lowEx THEN tip IS low;

RULE 5 : IF edu IS lowEdu and skill IS medSkill and ex is medEx THEN tip IS low;

RULE 6 : IF edu IS lowEdu and skill IS medSkill and ex is highEx THEN tip IS med;

RULE 7 : IF edu IS lowEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 8 : IF edu IS lowEdu and skill IS highSkill and ex is medEx THEN tip IS med;

RULE 9 : IF edu IS lowEdu and skill IS highSkill and ex is highEx THEN tip IS med;

RULE 10 : IF edu IS medEdu and skill IS lowSkill and ex is lowEx THEN tip IS low;

RULE 11 : IF edu IS medEdu and skill IS lowSkill and ex is medEx THEN tip IS low;

RULE 12 : IF edu IS medEdu and skill IS lowSkill and ex is highEx THEN tip IS med;

RULE 13 : IF edu IS medEdu and skill IS medSkill and ex is lowEx THEN tip IS low;

RULE 14 : IF edu IS medEdu and skill IS medSkill and ex is medEx THEN tip IS med;

RULE 15 : IF edu IS medEdu and skill IS medSkill and ex is highEx THEN tip IS med;

RULE 16 : IF edu IS medEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 17 : IF edu IS medEdu and skill IS highSkill and ex is medEx THEN tip IS med;

RULE 18 : IF edu IS medEdu and skill IS highSkill and ex is highEx THEN tip IS high;

RULE 19 : IF edu IS highEdu and skill IS lowSkill and ex is lowEx THEN tip IS med;

RULE 20 : IF edu IS highEdu and skill IS lowSkill and ex is medEx THEN tip IS med;

RULE 21 : IF edu IS highEdu and skill IS lowSkill and ex is highEx THEN tip IS med;

RULE 22 : IF edu IS highEdu and skill IS medSkill and ex is lowEx THEN tip IS med;

RULE 23 : IF edu IS highEdu and skill IS medSkill and ex is medEx THEN tip IS med;

RULE 24 : IF edu IS highEdu and skill IS medSkill and ex is highEx THEN tip IS high;

RULE 25 : IF edu IS highEdu and skill IS highSkill and ex is lowEx THEN tip IS med;

RULE 26 : IF edu IS highEdu and skill IS highSkill and ex is medEx THEN tip IS high ;

RULE 27 : IF edu IS highEdu and skill IS highSkill and ex is highEx THEN tip IS high ;

c. DEFUZZIFY:

DEFUZZIFY: tip: tip // Defzzzify output variable 'tip' : {'poor', 'good' , 'excellent'

TERM low := (0,1) (25, 1) (50, 0) ;

TERM med := (40, 0) (72,1) (85,0);

TERM high := (80, 0) (92,1) (100,1);

METHOD : COG; // Use 'Center Of Gravity' defuzzification method

END_RULEBLOCK

6.8. Summary:

It is clear from the results we obtained in this chapter, that our proposed system is easy to use through the simple user interface, because of the few steps needed to be used on our system, also it is characterized by flexibility in dealing with by consistent with the requirements of the employer expressed by certain percentage ratio.

As appears from candidates lists, the system is well effective in extracting lists, according to the proportion of matching required for each sections in the employment announcement, and in terms of accuracy, the results are satisfactory, as there are very close matching rates in some cases, and spaced out in other cases, according to seekers qualifications.

Chapter 7

Conclusion and Future Work

The recruitment process is an important and effective process in society, it is a sensitive and accurate process that depends on more than one factor, and more than one participates. Also, in order to make the recruitment process based on a successful scientific basis and achieves its goals by using systematic, efficient and better matching between candidates and offered jobs. We have included universities in the recruitment process and linked them to the labor market.

The expert domain objective is to link/match learning outcomes of different disciplines of universities with the labor market requirements. This is to determine as much as possible the most appropriate. As a case study, this work targeted specializations related to information technology; namely Computer science, Computer Engineering, Computer Engineering, Software Engineering, Computer Information Systems, Information And Communication Technology, Networking and Information Security and Web & Multimedia Technology).

In this research, we propose a new system passed on content base filtering approach and Fuzzy Logic, that find the appropriate job for job seekers, and provided a new method for the vacancy announcement process, which includes identifying the most important requirements in the applicant according to the opinion of the employer, to meet his/her needs accurately, as he/she controls the posted job which divided into specific categories,

that similar to the categories in the CV which are (qualification - experience - skills), each category has a specific percentage that reflects the importance of that part to the employer.

In our system, there is a type of interaction between the employer and the applicant, the employer can access the list of applicants for the job, and download their CVs, and then communicate with those he/she deems qualified for the job. It also gives users an idea of their suitability for the available jobs posted on the system, so they can apply for the appropriate job and can update the CV, which will also affect the results positively or negatively.

The proposed system approved that the change of ratios in the posted job each time is important; it could change the list of applicants and the match ratios for the same job. The system also gives the employer the ability to evaluate the employee after a while of work, as a feedback to the expert domain, so the data which measures the appropriateness of specialization for the required job become the mean of the data in survey, with the evaluation of the labor market for specialization.

Up to our knowledge, there is no previous study dealt with the subject of studying the lists of candidates for the jobs produced by the different e-recruitment systems, in different aspects, such as the importance, effectiveness and whether it is actually used, or whether it mimics the reality or not. In addition, the role of the employer in the online recruitment process was limited to the publication of a job application only. Consequently, we faced difficulty in finding a reliable (real -life) database regarding recruitment process to use in our work.

Here are the most important results we got through all over our study. First of all, our proposed system shows the difference of stability degree of different specializations for the same job, which is not like that in other systems. Another important thing is that our

candidate list orders candidates to their stability of job that makes it more effective that system that order them in an alphabetical order. Another result that the number of candidates us fewer than other systems because our list ranked the candidates in descending order based on their specializations from most suitable to the job to the least. In addition, our system provides a list of accepted candidates to the employer in a new form which gives the employer a comprehensive clear brief overview of the applicant. Also, our proposed system deals with all specializations, even though they are not mentioned in the job advertisement, and give them the similarity rate that may be high. Moreover, our proposed system gives the opportunity for applicants who were not accepted in a certain job because as we mentioned before we have three sections education, experience and skills each of them will be evaluated out of 100% so if you get a high rate on one section it will affect the other two sections positively. And expert knowledge database plays an important role in organization university specializations in a certain domain and give these specializations a chance to compete.

Moreover, our proposed system classifies the applicants for a job according to employer needs by the three sections we mentioned before. Finally, we used a structured CV with a structured job application to facilitate the ranking, so the result becomes according to the weighting of applicant's CVs based on the employer needs

As a future work, we intend to expand the expert knowledge database to include training and courses certificate, not only the specializations. Furthermore, we will study the effect of adding the obtained relevant training to the given specialization along with the specialization content for employment matching. Obviously, this will increase the accuracy of matching. On the other hand, we expect an overhead in system updates, but we believe that including such new information to the content will make the recruitment process more efficient and realistic.

We will study the impact of training and candidate specialization and the relationship between then to hiring a job, and the possibility to increase job seekers opportunities to employment, in terms of the training content and its relationship to specialization. Obviously, this will increase the accuracy, thus increase the overhead of system updates, but we believe that having this new information in hand will make the recruitment process more efficient and realistic.

References:

[1] Sun, A. Lim, E. and Ng, W. (2002): “Personalized Classification for Keyword-based Category Profiles. Proceedings of 6th European Conference on Research and Advanced Technology for Digital Libraries”.

[2] Mansouri, B. ZahediRi, M. Campos, C. and Farhoodi, M. (2018): “Online Job Search: Study of Users' Search Behavior using Search Engine Query Logs. International ACM SIGIR Conference on Research & Development in Information Retrieval”. pp: 1185–1188.

[3] Moraga, C. (2005): “Introduction to Fuzzy Logic Article. University of Niš”.

[4] Musale, D. Nagpure, M. Patil, K. and Sayyed, R. (2016): “Job Recommendation System Using Profile Matching and Web-Crawling. International Journal Of Advance Scientific Research and Engineering”. Vol. 1, no. 2, pp: 29-43.

[5] Mansur, F. Patel, V. and Patel, M. (2017): “A review on recommender systems. 4th International Conference on Innovations in Information, Embedded and Communication Systems”.

[6] Isinkaye, F. Folajimi, Y. and Ojokoh, B. (2015): “Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal”. Vol. 16, no. 3, pp: 261-273.

[7] ISO 9001. (2015):. Organizational Structure and the Job Description Requirements. (http://9001quality.com/iso-9001-organizational-structure-job-description-requirements.html, 03.03.2019).

[8] Anand, J. and Devi, S. (2016): “Literature review on e-recruitment and its perceived benefits: A walk towards paperless HR. International Journal of Applied Research”, pp: 528-531.

[9] BIH, J. (2006): “An Introduction to Fuzzy Logic. IEEE”. pp: 6 - 21.

[10] Gil, J. Paoletti, A. and Pichler, M. (2016): “A Novel Approach for Learning How to Automatically Match Job Offers and Candidate Profiles. Information Systems Frontiers”.

[11] Abial, M. Brown, I. (2020): “Conceptualizations of E-recruitment: A Literature Review and Analysis. IFIP International Federation for Information Processing”.

[12] Varmazyar, M. and Nouri, B. (2014): “A fuzzy AHP approach for employee recruitment. Research Institute of Petroleum Industry”. vol. 3, pp: 27 - 36.

[13] Daramola, O. Oladipupo, O. and Musa, A. (2010): “A Fuzzy Expert System Tool for on - line Personnel Recruitment. *Journal of Business Information Systems*”.

[14] Aggarwal, P. Tomar, V. and Kathuria, A. (2017): “Comparing Content Based and Collaborative Filtering in Recommender Systems. International Journal of New Technology and Research”. Vol. 3, no. 4, pp: 65-67.

[15] Manjare, P. kumbhar, J. Ovha, S. and Munde, R. (2017): “An Effective Job Recruitment System Using Content-based Filtering. International Research Journal of Engineering and Technology”. Vol. 04, pp: 556 – 559.

[16] Pati, P. (2016): “Study on Recommendation System. International Journal of Research in Computer Applications and Robotics”. Vol. 4, no. 10, pp: 5-8.

[17] R. Half, “22 Best Technology Job Boards You Need to Know About”, 2019. (https://www.roberthalf.com/blog/job-market/22-best-technology-job-boards-you-need-toknow-about.htm, 10.05.2020).

[18] Navigli, R. (2013): “A Quick Tour of BabelNet 1.1. Conference: Proceedings of the 14th international conference on Computational Linguistics and Intelligent Text Processing”. Vol. I, pp: 25 – 37.

[19] Navigli, R. and Ponzetto, S. (2010): “BabelNet: Building a Very Large Multilingual Semantic Network. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics”. pp: 216–225.

[20] Al-Otaibi, S. and Ykhlef, M. (2012): “A survey of job recommender systems. International Journal of the Physical Sciences”. Vol. 7, pp: 5127 - 5142.

[21] Bouraga, S. Jureta, I. Faulkner, S. and Herssens, C. (2014): “Knowledge-based recommendation systems. International Journal of Intelligent Information Technologies”.

[22] Statistics How To. (2013): ”Stratified Random Sample: Definition, Examples. Florida”.(https://www.statisticshowto.com/stratified-random-sample.htm, 02.08.2020).

[23] Statistics How To. (2014): “Z Alpha/2 (za/2): What is it, How to Find it. Florida”. (https://www.statisticshowto.com/z-alpha2-za2, 02.08.2020).

[24] SABHA, S. (2018): “Impact of Online recruitment on recruitment. International Journal of Education and Research”. Vol. 6, no. 4, pp: 47 – 52.

[25] Qaiser, S. and Ali, R. (2018): “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications”. Vol. 181, no. 1.

[26] Thompson, S. (2012): “Sampling, Third Edition. Wiley Series in Probability and Statistics”. pp.59-60.

[27] Desai, V. Bahl, D. Vibhandik, S. and Fatma, I. (2017): “Implementation of an Automated Job Recommendation System Based on Candidate Profiles”, International Research Journal of Engineering and Technology, Vol. 04.

[28] Gupta, Y. (2016): “Literature review on E-Recruitment: A step towards paperless HR. International Journal of Advance Research in Computer Science and Management Studies”. Vol. 4, no. 1.

Appendices:

Appendix A: The Survey hardcopy Form

ALQUDS UNIVERSITY
FACULTY OF SCIENCE & TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

The attached questionnaire is part of a study conducted by the Al-Quds University to obtain a master's degree in computer science.

The thesis Employment Recommendation System using Content- Based and Fuzzy Logic for studying the effect of domain experts who working in higher education in Palestinian universities in the West Bank and have degree in computer science or other approaches in computer fields and have an experience in teaching computer courses or certificates and employer view in the recruitment process.

By filling the questionnaire, your contribution, and your thoughtful suggestions will help improve our study.

تأتي هذه الاستبانة استكمالاً لرسالة ماجستير في تخصص علم حاسوب من جامعة القدس بعنوان:

Employment Recommendation System using Content-Based and Fuzzy Logic

تستهدف هذه الدراسة العاملين في مجال التعليم العالي في الجامعات الفلسطينية في الضفة الغربية والحاصلين على شهادات عليا في تخصصات الحاسوب المختلفة، للاستفادة من خبرتهم ومقترحاتهم في ايجاد أفضل وظيفة ممكنة لكل تخصص من التخصصات المقترحة في مجال الحاسوب، آملين أن تثري ارائكم ومقترحاتك هذه الدراسة.

Researcher
Amani Habboub

Personal information:

Position: Education

Years of experience:

Please choose the appropriate degree to get the most suitable job for the given specialization and then rate it from 10 (the best job) – to 0 (the job that does not apply to the specialization at all).

Appendix table of different disciplines in the branches of computer science in the Palestinian Universities to be used in the case of the proposal of other disciplines that are not in the table and more suitable for the functions listed below.

Please Rate the following specialization to each position from 0 (not related) to 10 (extremely matching)					
Specialization	Jobs	degree			Rate
		MS	BA	diploma	
Computer Science	Programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems manager				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Computer Engineering	programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems managers				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Software Engineering	Programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems manager				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Computer Information Systems	programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information				

	systems managers				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Information and Communication Technology	programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems managers				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Networking and Information Security	programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems managers				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				
Web & Multimedia Technology	programs developer				
	Mobile App Developer				
	Web developer				
	Project manager				
	Computer programmer				
	Computer and information systems managers				
	Computer systems analyst				
	Automatic learning engineer				
	Computer Networks Engineer				
	System Engineer				
	Systems Architect				
	Data Engineer				
	Database Administrator				
	Information Security Analyst				
Notes	----------				

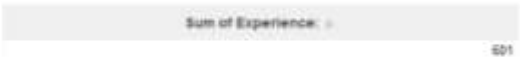
Appendix B: The survey results: the rate column shows what the most expert rate is.

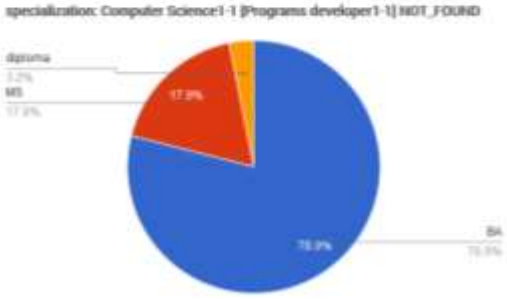
Specialization	Degree	Job	Rate	The result average
Computer science	BA	Program developer	10	8.89
	BA	Mobile App Developer	10	8.57
	BA	Web developer	10	8.47
	MS	Project manager	8	8.02
		Computer programmer	10	8.77
	MS	Computer and information systems manager	8	8.32
	BA	Computer systems analyst	6	7.85
	BA	Computer Networks Engineer	7	6.97
	BA	System Engineer	8	6.77
	BA	Systems Architect	6	6.92
	BA	Data Engineer	8	7.29
	BA	Database Administrator	8	8.40
	BA	Information Security Analyst	8	7.58
Computer Engineering	BA	Program developer	10	
	BA	Mobile App Developer	9	
	BA	Web developer	8	
	MS	Project manager	10	
		Computer programmer	10	
	MS/BA	Computer and information systems manager	10	
	BA	Computer systems analyst	8	
	BA	Automatic learning engineer	10	
	BA	Computer Networks Engineer	10	
	BA	System Engineer	10	
	BA	Systems Architect	10	
	BA	Data Engineer	8	
	BA	Database Administrator	7	
	BA	Information Security Analyst	8	
Software Engineering	BA	Program developer	10	
	BA	Mobile App Developer	8	
	BA	Web developer	10	
	MS	Project manager	10	
		Computer programmer	8	
	MS	Computer and information systems manager	8	
	BA	Computer systems analyst	8	
	BA	Computer Networks Engineer	7	
	BA	System Engineer	9	
	BA	Systems Architect	7	
	BA	Data Engineer	7	
	BA	Database Administrator	10	
	BA	Information Security Analyst	7	
Computer Information Systems	BA	Program developer	8	
	BA	Mobile App Developer	7	
	BA	Web developer	8	
	MS	Project manager	8	
		Computer programmer	6	

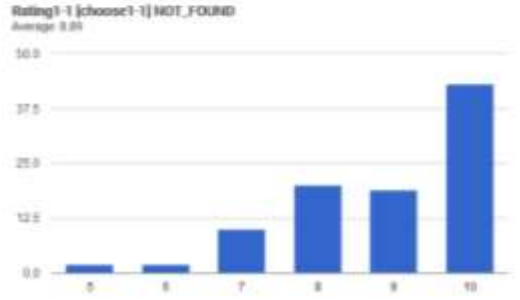
	MS	Computer and information systems manager	8	
	BA	Computer systems analyst	8	
	BA	Computer Networks Engineer	5	
	BA	System Engineer	5	
	BA	Systems Architect	7	
	BA	Data Engineer	8	
	BA	Database Administrator	8	
	BA	Information Security Analyst	6	
Information and Communication Technology	BA	Program developer	7	
	BA	Mobile App Developer	7	
	BA	Web developer	7	
	MS	Project manager	8	
		Computer programmer	7	
	MS	Computer and information systems manager	8	
	BA	Computer systems analyst	7	
	BA	Computer Networks Engineer	8	
	BA	System Engineer	8	
	BA	Systems Architect	8	
	BA	Data Engineer	8	
	BA	Database Administrator	8	
	BA	Information Security Analyst	8	
Networking and Information Security	BA	Program developer	6	
	BA	Mobile App Developer	6	
	BA	Web developer	5	
	BA	Project manager	8	
	BA	Computer programmer	6	
	MS/BA	Computer and information systems manager	8	
	BA	Automatic learning engineer	8	
	BA	Computer Networks Engineer	10	
	BA	System Engineer	8	
	BA	Systems Architect	8	
	BA	Data Engineer	8	
	BA	Database Administrator	8	
	BA	Information Security Analyst	10	
Web & Multimedia Technology		Program developer	8	
	BA	Mobile App Developer	10	
	BA	Web developer	10	
	BA	Project manager	8	
		Computer programmer	7	
	BA	Computer and information systems manager	8	
	BA	Computer systems analyst	7	
	BA	Computer Networks Engineer	0	
	BA	System Engineer	0	
	BA	Systems Architect	0	
	BA	Data Engineer	6	
	BA	Database Administrator	8	
	BA	Information Security Analyst	0	

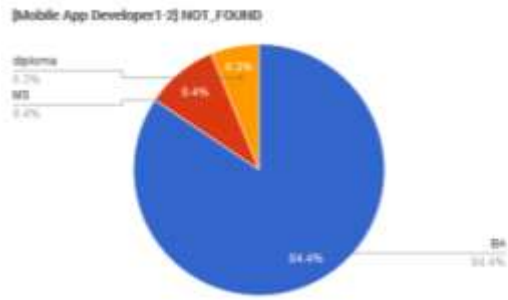
Appendix C: The survey results example in charts for the computer science specialization.

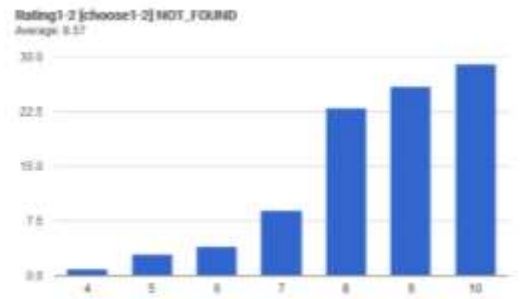
The graph axis: y presents the rate of the experts, and the x axis presents the average of the expert rate.

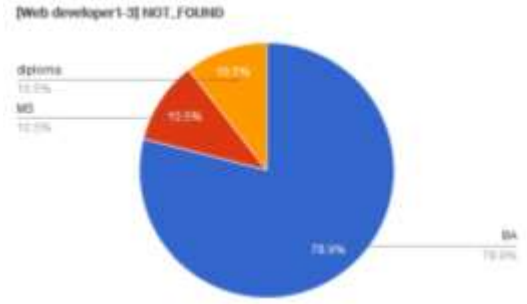


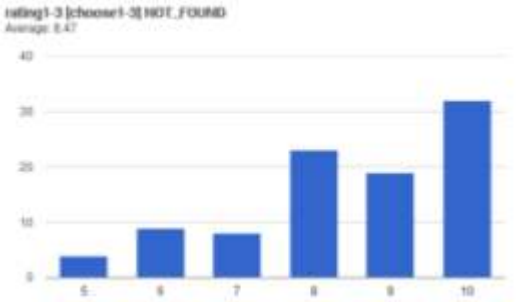


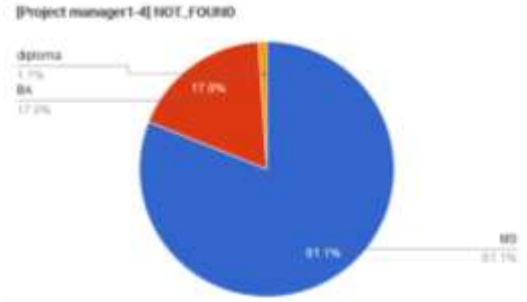


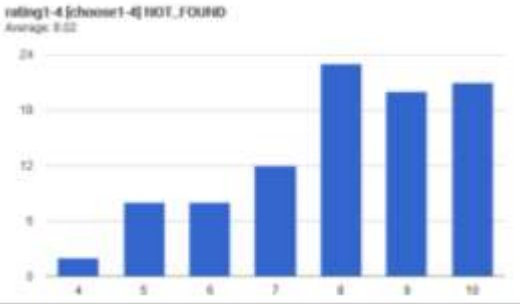


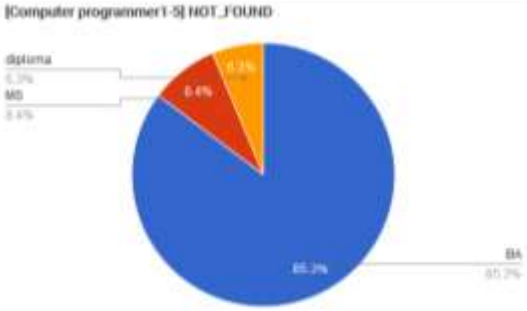


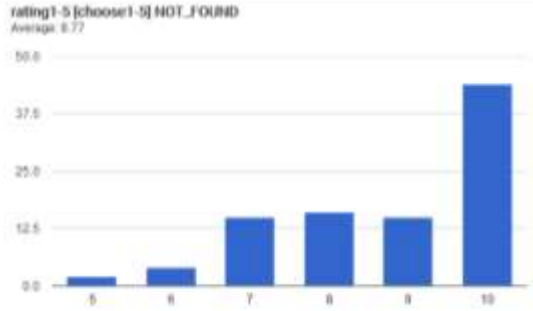

[Web developer1-3] NOT_FOUND
diploma
MS
BA

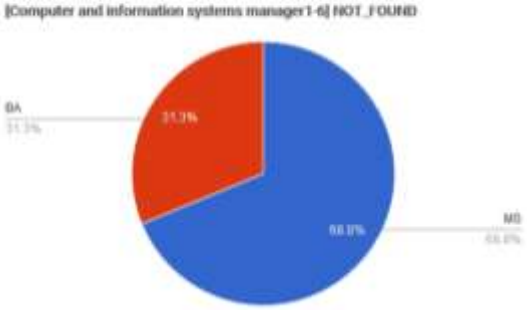

rating1-3 [choose1-3] NOT_FOUND
Average: 8.47

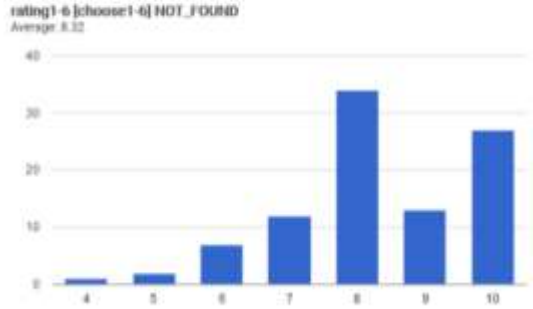

[Project manager1-4] NOT_FOUND
diploma
BA
MS

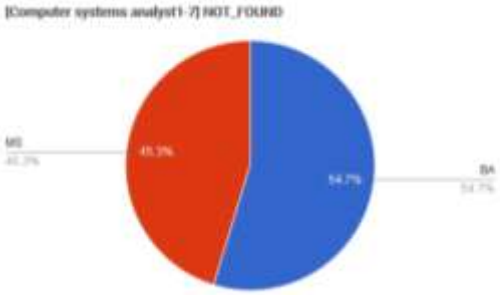

rating1-4 [choose1-4] NOT_FOUND
Average: 8.02

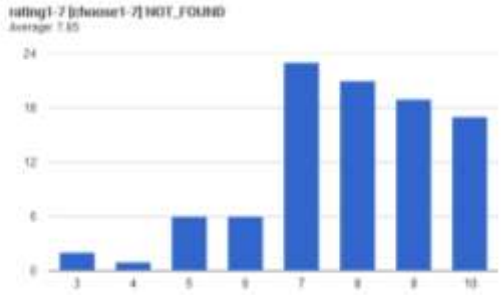

[Computer programmer1-5] NOT_FOUND
diploma
MS
BA

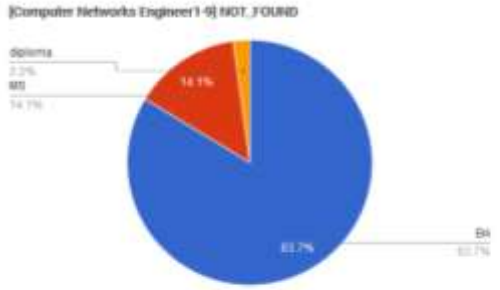

rating1-5 [choose1-5] NOT_FOUND
Average: 8.77

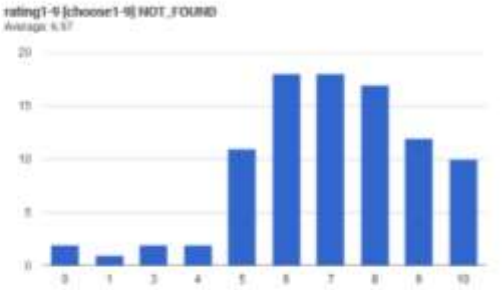

[Computer and information systems manager1-6] NOT_FOUND
BA
MS

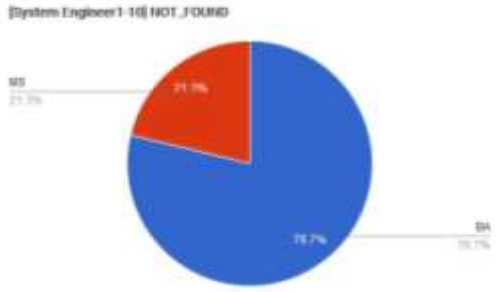

rating1-6 [choose1-6] NOT_FOUND
Average: 8.32

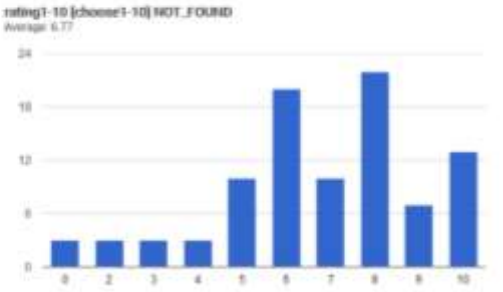

[Computer systems analyst1-7] NOT_FOUND
MS
45.3%
54.7%
BA

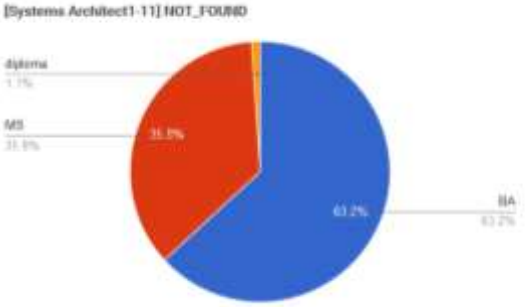

rating1-7 [choose1-7] NOT_FOUND
Average: 7.85
24
18
12
6
0
3
4
5
6
7
8
9
10

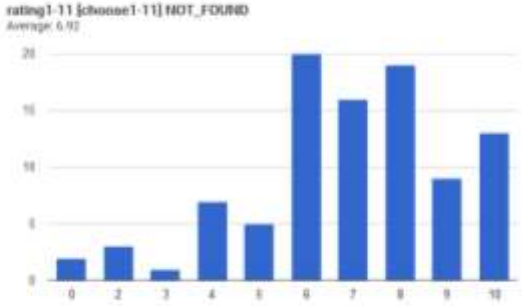

[Computer Networks Engineer1-9] NOT_FOUND
diploma
2.2%
MS
14.1%
83.7%
BA

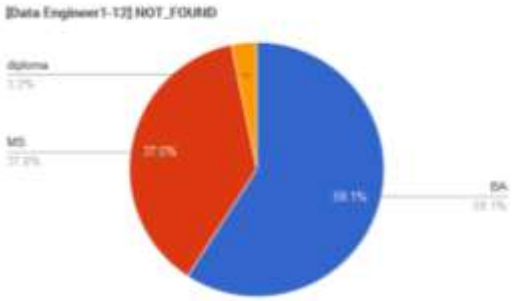

rating1-9 [choose1-9] NOT_FOUND
Average: 6.97
20
15
10
5
0
0
1
3
4
5
6
7
8
9
10

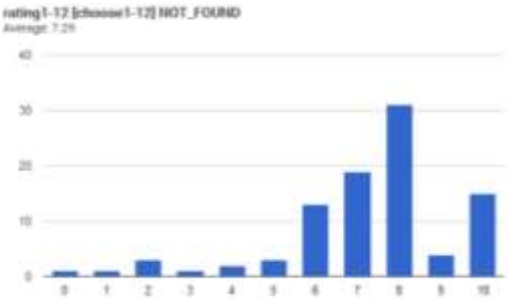

[System Engineer1-10] NOT_FOUND
MS
21.3%
78.7%
BA


rating1-10 [choose1-10] NOT_FOUND
Average: 6.77
24
18
12
6
0
0
2
3
4
5
6
7
8
9
10


[Systems Architect1-11] NOT_FOUND
diploma
1.7%
MS
35.8%
63.2%
BA

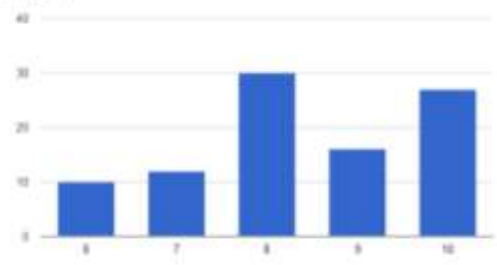

rating1-11 [choose1-11] NOT_FOUND
Average: 6.92
20
15
10
5
0
0
2
3
4
5
6
7
8
9
10


[Data Engineer1-12] NOT_FOUND
diploma
3.2%
MS
37.0%
59.1%
BA

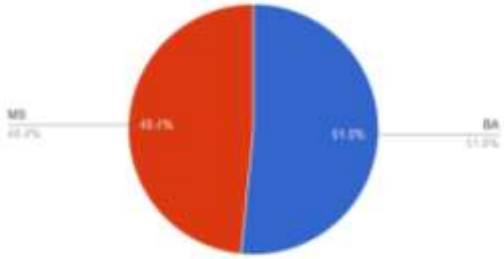

rating1-12 [choose1-12] NOT_FOUND
Average: 7.29
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10

[Database Administrator1-13] NOT_FOUND

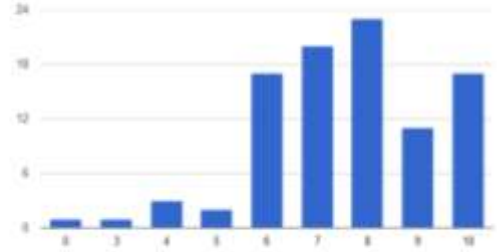
rating1-13 [choose1-13] NOT_FOUND



[Information Security Analyst1-14] NOT_FOUND



rating1-14 [choose1-14] NOT_FOUND



Notes1	Notes1
BA is most appropriate for most of these jobs!	1

specialization: Computer Engineering2-1 [Programs developer2-1] NOT_FOUND

Appendix D: Specializations of Computer domain in Palestinian Universities in the West Bank.

degree	Specialization
Master	**Software Engineering**
	Computer science
	Computer science / software engineering
	Advanced Computing
	Computing
	Computing and Information Systems
	Master of Information Technology
	Information Technology and Systems Management
	Master of Informatics
Bachelor degree	**Computer Science**
	Computer Engineering
	Computer Systems Engineering
	Networks and Communication
	Management Information Systems
	Information and Communication Technology
	Sub-specialties (Web Technologies, Information Security
	Computer Information Systems
	Networking and Information Security
	Computer Science in partnership with the labor market
	Communication Engineering
	Network Technology and Mobile Phone
	Applied Information Technology
	Information Technology
	Multimedia Section and Digital Production
	Information Systems
	Electronic Marketing
	Multimedia / Graphics
	Computer and Communication Engineering
	Software Engineering
	Software Development

	Multimedia & Web Development
	Mobile Computing and Smart Appliance Applications
	Web & Multimedia Technology
	Computer Network Security & Protection
	Multimedia
	E-Learning
	Software and databases
	Information Security Engineering
	Computer Systems Development
	Computer and Digital Systems
Diploma	**Software and Database**
	Telecommunication
	Electronic Maintenance
	Multimedia and Animation
	Web Design & Development
	Information Technology
	Graphic Design
	Mobile Phone Technology
	Computer Science
	Information Technology
	Multimedia
	Computer Maintenance
	Electronics Instrument Technology
	Internet and computer networks
	Networks
	E - Marketing
	Networking & Computer Maintenance
	Computer Engineering
	Computer Networks and Technical Support
	Management and development of web pages

Appendix E: Job application and job descriptions

ISO 9001 – The Job Description

Any person in the organization, who is related to the realization of the product, must have a job description (J.D.). A documented J.D. should include:

1. The title of the role or function. The actual name of the role. It is important to identify every role: Operational Manager, Administrative Secretary, purchase manager etc.
2. Subjection. To which another role higher in the hierarchy, the function is subjected to. To whom must it report at the end of the day?
3. Specifications of responsibilities. The activities and tasks that combine the function's working day. What must he does every day: typing information to the computer, answering telephone calls, fixing malfunctions, surfing the internet … etc.
4. Specifications of authorities. The authorities are the points along the processes where the function is authorized to decide: submitting price quotations, approving payments, approving credits etc.
5. External qualification. The external qualifications required for the job: engineer, electronic technician, certified fireman, certified pilot etc.
6. Internal qualification. The internal qualifications initiated by the organization required for the job.

نظام توصية لعملية التوظيف باستخدام نهج التصفية القائم على المحتوى (CBF) والمنطق الضبابي (Fuzzy Logic)

اعداد: أماني باسم محمد حبوب.

المشرف: د. بديع سرطاوي.

الملخص

في الوقت الحاضر، هناك فجوة بين متطلبات سوق العمل والتخصصات العلمية المتاحة في الجامعات. وقد أدى وجود تخصصات مماثلة بأسماء مختلفة في الجامعات، إلى وجود عدد كبير من التخصصات لعدد محدد من الوظائف. وذلك جعل من الصعب العثور على التخصص الأنسب للوظائف المتاحة في سوق العمل.

تعد المطابقة التلقائية للوظائف والمرشحين، مشكلة كبيرة للشركات والمتقدمين للعمل معا. هناك العديد من التقنيات والخوارزميات لمساعدة الباحثين عن عمل في العثور على الوظيفة المناسبة، تهدف كل هذه الخوارزميات إلى إيجاد أفضل وظيفة للمرشح الانسب، وجعل عمليات التوظيف أسرع وأكثر دقة وشفافية. في هذه الأطروحة، نقترح نهجًا يستخدم طريقة التوصية القائمة على المحتوى content-based filtering approach جنبًا إلى جنب مع المنطق الضبابيFuzzy logic وقاموس BabelNet لمطابقة الوظيفة الشاغرة مع المتقدمين لها، لاستخراج أفضل النتائج.

نقترح أيضًا إنشاء قاعدة بيانات، تحتوي على رأي الأكاديميين والمحاضرين في الجامعات الفلسطينية بالتخصصات المتاحة بهذه الجامعات، والوظائف الانسب لتلك التخصصات بناءً على مخرجات

التعليم لكل تخصص، كما يتيح النظام المقترح لصاحب العمل القدرة على تقييم الموظف بعد فترة من العمل، كنوع من التغذية الراجعة لقاعدة البيانات، بحيث تصبح البيانات المخزنة عبارة عن الوسط الحسابي للبيانات التي تشمل رأي المحاضرين بالتخصصات، مع تقييم سوق العمل للتخصص. وجدير بالذكر ان هذا البحث استهدف تخصصات ذات علاقة بتكنولوجيا المعلومات على وجه الخصوص مثل علم الحاسوب وهندسة الحاسوب وهندسة البرمجيات، كحالة للدراسة.

وبالتالي، يتم تلخيص مساهمتنا الرئيسية على النحو التالي: نقترح نموذجًا لإدارة عملية التوظيف عن طريق تقييم السير الذاتية للمتقدمين، وفقًا لاحتياجات صاحب العمل، الذي يمكنه تحيد احتياجاته بدقة عن طريق تقسيم اعلان الوظيفة الى ثلاث فئات: (المؤهل العلمي – الخبرات – المهارات)، بحيث يستطيع صاحب العمل اعطاء كل فئة نسبة مئوية حسب اهميتها بالنسبة له. مع مراعاة رأي الأكاديميين العاملين في الجامعات الفلسطينية في مختلف التخصصات، وعلاقتهم بسوق العمل، وذلك لجعل العملية أقرب إلى الواقع وأكثر شفافية، وبالتالي توفير الوقت والجهد المبذول للباحثين عن وظائف وأصحاب العمل على حد سواء.

في النظام المقترح، يوجد نوع من التفاعل بين صاحب العمل والمتقدم للوظيفة، ويستطيع صاحب العمل الاطلاع على قائمة المتقدمين للوظيفة وتحميل سيرهم الذاتية، والتواصل مع من يراهم مؤهلين للعمل. كما تعطي المستخدمين فكرة عن مدى ملائمتهم للوظائف المتاحة والموجودة على النظام، فيستطيع طالب الوظيفة التقدم للوظيفة الملائمة كما يستطيع تحديث السيرة الذاتية، والذي سيؤثر ايضا على النتائج بشكل ايجابي او سلبي.

ولتطوير النظام في المستقبل، نرى أن تطوير قاعدة البيانات لتشمل رأي الأكاديميين والسوق العمل بالدورات والتدريبات التي يلتحق بها الباحثون عن عمل، والموظفون الراغبون في تطوير أنفسهم، ومدى تأثيرها على التخصص الجامعي، أي كيفية تأثيرها على ايجاد فرصة عمل، سلبا ام ايجاباً، سواء كانت بمجال التخصص او بمجالات جديدة لا علاقة لها بالتخصص الجامعي.