International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I3P213
Volume 69 Issue 3, 69-75, March 2021

A Multi-Criteria Analysis and Advanced Comparative Study of Recommendation Systems

Safia Baali[1], Ibrahim Hamzane[2], Hicham Moutachaouik[3], Abdelaziz Marzak[4]

[1,2,4] *Laboratory of Information Technology and Modelling*
Hassan II university, faculty of Sciences Ben M'sik, Casablanca, Morocco
[3]*Structural Engineering, Intelligent Systems, and Electrical Energy*
Hassan II University, ENSAM, Casablanca, Morocco
safia.baali@gmail.com, hamzane.ibrahim@gmail.com

Abstract *- In order to ensure the performance of delivery, especially in the IT digital services company, we need to affect the right candidate in the right position; in this context, the recruitment process needs to be automatic, subjective, and more accurate. Employers need help to find the right candidate from an of resumes, and many studies have proposed several solutions for recommending a candidate for recruitment and matching between the job offer and cv candidates that exploit text processing and semantics-based techniques. In our research, we aim to present a comparative study between the different approaches used for the matching job and cv candidate; we also proposed a new approach to recommend a potential candidate for a specific work area, our study will be based on an IT service company based in Morocco and aim the automatization of the recruitment process to ensure the assignment of the candidate in the right task and ensure the success of the company, then the customer's satisfaction.*

Keywords — *Matching Job/Resume; recommendation system; Clustering; TFIDF; KMeans; recruitment.*

I. INTRODUCTION

To assure the success of any company and the performance in the delivery, we must take into account the skills of the employees and select adequate profiles to assign them to the right job position.

The objective of our research is to identify the most efficient method of recruitment, especially for Job /profile matching; the HR team aims to build a pool of potential candidates that represents the adequate profile for the job offer position [1], the identification of potential profiles must be in external (new candidate resumes) and internal way (existing employee resumes), to assure the efficiency of the assignment in the adequate job position, HR department uses resumes as a principal input to identify the right profile. The resume is effectively an unstructured document that requires extraction of the relevant information (features) to represent a document in text mining; major terms are considered as features [2] [3]. In this paper, we opt for clustering to group resumes based on the similarity between the terms (features of the candidate profile). The recommended system proposed in our research is defined by the most efficient analytics algorithms that ensure finding adequate candidates for a particular job offer and increase the accuracy of the recommendation, the evaluation of our recommender system will be carried out in an IT digital services company based in morocco.

The structure of this paper is as follows: First, we introduce the background and related work; second, we explain the recommendation system proposed, we present in the third section the experimental work carried out to demonstrate the accuracy of the system, conclude in the final section indicating the perspectives and limitations of our proposition.

II. RELATED WORK

A. *Recommendation System for Human Resources*

The recommender system aims to generate interesting items or products for web users. They offer useful and adapted information to users' profiles based on their preferences and behaviors [4] [5] in many fields, Recruitment (Indeed, CareerBuilder) to recommend jobs, e-commerce (Amazon) to recommend the products for the users, films (Netflix) ...etc. Due to the exponential increase of the available data and resources from the web, Recommendation systems have taken great importance in providing users with suggestions to meet their needs and preferences. Recommendation Systems associate different techniques of information filtering, artificial intelligence, social networks, and human-computer interaction. There are three main approaches in recommendation system, based-content filtering, which makes recommendations by comparing the content of resources with the user's preferences [6], collaborative filtering, which makes recommendations by analyzing the users' opinions and those of other users about the resources they have consulted [7], finally, the hybrid approach which associate the different tow approach in order to improve the accuracy of the recommendation, then increases the user satisfaction.

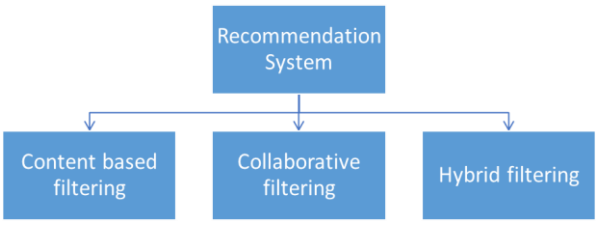

Fig.1. Recommendation System Approach.

Despite the growing popularity of the recommendation Systems, there are many limitations and problems from which we quote: [8], [9].

- **Critical Mass problem:** This issue illustrates the difficulty of dealing with the fact that there are few items evaluated and few users who conduct these evaluations.
- **Cold Start problem:** we often find ourselves confronted with the problem that a user is compared with no other user. This problem is because few or no users have evaluated a given item or a given user has rated few or no items.
- **Induction problem:** Recommender systems are based on the principle that a user who has exhibited behavior in the past will tend to exhibit similar behavior in the future. However, this principle is not necessarily valid in the real context.

In Human resources, and especially in e-recruitment, there are many works related to the recommendation system for the base of the adequate profile on the matching between the job offer and candidate profile, using different techniques to improve the accuracy of the recommendation.

B. Semantic similarity in the resumes

The recommendation system allows the search of the adequate profiles for the job offer, and resumes contain the principal feature of information for the candidate in various domains, many keywords specify the field of expertise, experience, skills of the candidate, Educational attainment, it contains keywords that help recruiters to match the experience of the candidate to the job offer. The cv contains unstructured data, which is mandatory to be converted into a vector presented by term-document matrix; the rows contain the resumes, columns contain the most important features extracted from resumes [10]. A resume contains all the information related to the candidate, the summary of experiences, technical and soft skills; in different resumes, we can find various words used in order to describe the same context; these words are normally related semantically. There are many methods used to process the textual data, and we can measure the semantic similarity using the classification of the terms into synsets provided by WordNet. WordNet is a large lexical database of English. Synsets are interlinked using conceptual-semantic and lexical relations [11], so the columns in the term-document matrix are presented by synsets in order the selection of major terms; we also quote other methods, the use of word bag to compose a word vector of the same dimension, and then use the TF-IDF as a numerical statistic that is intended to reflect how important a word is to a document, it assigns weights to the word to measure its relevance in the document[12].

C. Clustering

Cluster analysis is an unsupervised machine learning algorithm that allows involving the discovery of natural grouping in data automatically, and it analyses the input data to find groups in feature space; there are many methods related to clustering; we quote two principals and popular methods.

D. Hierarchical clustering (HCA)

It is a method of cluster analysis that aims to define a hierarchy of clusters [13]. It is divided generally into two principal types:

- **Agglomerative:** called the "bottom-up" method: each observation starts in its cluster, and pairs of clusters are merged as one moves up the hierarchy; this method is used by [14] to define the cluster related to the work area, for the recommending profile resume to the appropriate job offer.
- **Divisive:** called "top-down" method: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

E. K-means Clustering

K-means algorithm [15] is a typical clustering algorithm based mainly on distance. It represents the evaluation index of similarity, and the similar objects are the closest ones, then the similarity is the greatest. The selection of k initial clustering center points impacts the clustering result because, in the first step of this algorithm, any k objects are randomly selected as the initial clustering center, initially representing a cluster.

The steps of the K-means are the following:

- Choose the number K of clusters.
- Select at random K points the centroids.
- Assign each data point to the closest centroid (that forms K clusters).
- Compute and place the new centroid of each cluster.
- Reassign each data point to the new closest centroid.

III. MULTI-CRITERIA COMPARATIVE STUDY

A. SWOT Analysis

The table below presents a minimal SWOT analysis to summarize the strengths and weaknesses of each approach:

TABLE I. SWOT ANALYSIS MINIMAL

Models	Type of the Model	Positives	Negatives
M1	Prospect [16]: Mining Tool, rank the candidates by matching to the job description and use of the Filters, Resume Segmentation is based on Lexicon and Visual Feature	Decision support tool to help these screeners shortlist resumes efficiently. Ranking the candidate	The Presence of skill in the sections describing projects is not weighted higher than other parts of the resume. The Presence of skill in more recent projects is not weighted higher than those in older projects
M2	Domain-WordNet	Solution for semi-automatically	The data are only related to the

Models	Type of the Model	Positives	Negatives
	[17] is automatically generating a domain-specific semantic lexicon.	enriching domain-specific ontologies Provide qualitative "good" enough ontologies to be comparable to standard ontologies	specific domain for the society Epiqo. Unknown knowledge of the domain experts. Here is a rigid pre-set structure to the ontologies
M3	Matching using Lucene Engine [18,19], Scoring [20] using the similarity.	The recommendation of potential candidates using the extraction of competence. The use of the classification using supervised machine learning to detect the activity area and improve the accuracy of the system	The system doesn't consider the detection of the recent experience related to the need for the job offer. The automatic construction of ontologies of skills is not taken into account.
M4	Measure of similarity (Jaccard [21,22], Levenshtein [23], Hamming [24])	Recommendation of the potential profiles (scoring, automatic annotation, pseudonymization)	The matching of job and resume doses do not take into account the entire content of documents, but only the content found via ontologies
M5	The proposition of machine learning-based adaptive approach [25]. The objective is to compute the cost of transforming a profile into a job offer	learning how human experts (solved cases in the past in order to predict the behavior in the future situation) The approach suggests representing job offer and profile using shared terminologies in order to overcome the limitations of dealing with heterogeneous representations of the skill	There is no use of automatic ontologies for competencies. The model proposed must be evaluated using a real and large volume of data.
M6	The matching algorithm [26] will be able to select relevant clusters	The clustering allows the identification of clusters and calculates the similarity between the new job offer and profiles	Small training data The data set contains the resumes and job offers; for a new entry, it will be difficult to ensure the accurate recommendation

We aim to make a difference between approaches by facilitating the choice of the best model to be used according to desired criteria and their importance.

The score of an approach is calculated based on several criteria. So far, we have identified X criteria; indeed, based on SWOT analysis:

- **C1:** To treat the automation of the recruitment, we must consider the document of the resume for building a semantic space, preprocessing the document text, and use TF-IDF for the construction of the feature's vector.

Models	Type of the Model	Positives	Negatives
	and only match against all vacancies contained within these clusters. The data is analyzed using data mining with the analytic software WEKA	containing in clusters.	
M7	Data preprocessing [12] Text mining using TFIDF. Cosine Similarity between job offers. clustering using Kmeans++	Identification of cluster by field of activity.	The matching between a new resume and a job offer is not taken into account.
M8	Use of Synset-based document matrix construction method (WordNet) for text mining, agglomerative hierarchical clustering for clustering [14].	Accurate Identification of the cluster by field of activity. Recommendation of new resume for a specific field of activity	The matching between a new resume and a job with the offer is not taken into account. For the use of agglomerative hierarchical clustering, it is mandatory to specify the distance metric and the linkage criteria

B. Multi-criteria analysis

After seeing the advantages and disadvantages of each model, we will now develop a multi-criteria analysis between these frameworks. A Multi-Criteria Decision Analysis Criteria definition (MCDA) is a valuable tool that can be applied to many complex decisions. It can solve complex problems that include qualitative and/or quantitative aspects in the decision-making process.

- **C2:** For analyzing the data and building of the specific cluster related to the work area, the use of the clustering that is adapted to various changes in data can also produce higher clusters.
- **C3:** The best way to measure the similarity between documents and clusters is the cosine similarity; however, the similarity must also be measured between a new job offer and a Resume to increase the accuracy of the recommendation.
- **C4:** To consider the automatic construction and progression of the ontologies of skills.

C. Multi-criteria analysis method

There are several possible methods to make a comparison between the frameworks using several criteria. These methods can be divided into three main families.

- **Complete aggregation (top-down approach)**: Aggregating the n criteria to reduce them to a single criterion.
- **Partial aggregation (bottom-up approach)**: Comparing potential actions or rankings to each other and establishing between them outranking relations.
- **Local and iterative aggregation:** Looking primarily for a starting solution, then we proceed to an iterative search to find a better solution.

D. Weighted Sum Method (WSM)

We chose the Weight Sum Method (WSM) for our analysis. Indeed, this method allows us to find the best possible approach by assigning a weight to each comparison criterion; it allows considering all the criteria according to their value and without a criterion penalizing the other criteria [27].

We presented the four comparison criteria cited on which the comparative study will be based. We notice that these criteria are based on the characteristics of each of the approaches presented in the comparative study and the SWOT analysis presented above; we summarized all the characteristics (strengths and weaknesses) in four global criteria to ensure better analysis and optimize the comparison [28].

These criteria have the same importance; therefore, the WSM weight will be the same for each criterion and equal to “1”. However, as we will see further, the weight of each criterion can change depending on each company.

E. Multi-criteria choice matrix

The WSM method starts with filling the multi-criteria choice matrix. The columns contain the frameworks to be compared, and its lines contain criteria with the weight assigned to each criterion which we agree "1" as all the criteria have the same importance, and in cells, there is the score given to each framework based on the detailed comparative study of each framework [2, 3, 5].

About the score, we will then use the maturity model, which consists of five levels of maturity, to weight the criterion on each framework; each level will give a score; for example, "level 1" will leave a score of "1".

We recall the definition of the five levels by modifying the definitions to apply it to our case [29].

- **Level 1:** The criteria are not applied.
- **Level 2:** The criteria are not applied completely.
- **Level 3:** The criteria are fully applied.

The table below represents the resulting multi-criteria choice matrix according to the score of each criterion.

TABLE II. TABLE TYPE STYLES

Models / Criteria	*C1*	*C2*	*C3*	*C4*	*average*	*%*
M1	1	1	2	2	1.5	50
M2	2	1	1	3	1.75	58.3
M3	2	1	1	1	1.25	41.7
M4	2	1	1	2	1.5	50
M5	1	1	2	2	1.5	50
M6	2	3	3	2	2.5	83.3
M7	3	3	3	2	2.75	91.7
M8	2	3	3	2	2.5	83.3

We convert the table into a spider chart for the visual purpose; we notice that there is no complete model; however, Model 3 is the most complete according to our investigations, see figure below:

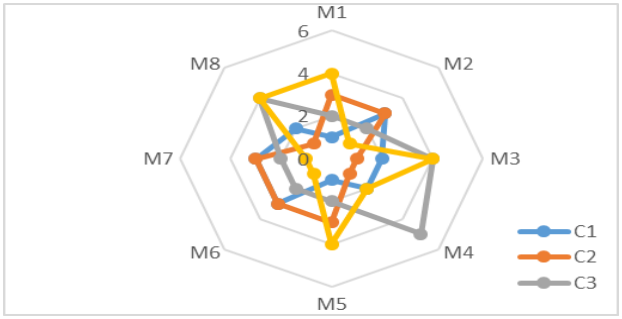

Fig.2. Spider chart multi-criteria decision.

F. Discussion

As we analyze the results, we conclude that the model M7 presents 91.7% of the use of the four criteria, then the model 6 and 8 with 83.3% of the use, these models have an accurate recommendation of the resume to the job, otherwise many improvements must be applied to these models to improve the accuracy of the recommendation.

IV. PROPOSED APPROACH

The main objective of our research is to develop a recommendation system for an IT digital services company that can manage the knowledge that is distributed among many unstructured documents (CV profiles, Job offers).

The principal parts of our system are:

- Recommend an adequate resume for the job offer in the specific skill.
- Matching job offer/cv, resume using the measure of similarity.

A. Recommendation system approach

To assign the new CV to the new job offer, we need to define a specific group that is related to the specific work area, in which we can recommend the new CV candidate or the existing employees in the company. According to the methodology followed in the works [14], the

recommendation of the new CV is based on the extraction of skills related to the CV and the measure of the cosine similarity between CV and the cluster related to the different work area (sales, account, Purchase, Customer service), this approach aims to match the new CV with a specific domain but not with a specific job position, to give a solution for this limitation, we need to approve the accuracy of the recommendation through the addition of another step to measure the cosine similarity between a new job offer and CV candidates contained in the specific cluster related to the specific work area, based on other features aside from the skills like age, educational Attainment, etc.

The methodology of our research is as follows. We present in figure 3 a diagram of our proposed Framework.

In the sections below, the explanation of each step:

a) Data collection

Data are the unstructured documents of CV resumes related to the IT services, for various work areas (Business Analyst, Developers, data scientist, Testers …), our CV is collected from the website Indeed.

b) Data preprocessing

The resumes (CV) present unstructured data; in this step, we will remove stop words, tokenization and lemmatization, normalization, then we process the data vectorization of the word to a vocabulary-text matrix using TF-IDF [30] that is considered as a numerical statistic that assigns weights to the world to measure its relevance in the document.

$$TFIDF(x,y) = TF(x,y) \times \log(N/DF(x))$$ (Term x within document y)

$TF(x,y)$ = frequency of x in y

$DF(x)$ = number of documents containing y

N = Total number of documents

c) Clustering

Grouping data into the different work areas (Business analysis, development, Data Science, Testing, etc.) using clustering with KMeans. The reason behind adopting the KMeans clustering is that it is adapted to various changes in data, it can also produce higher clusters, the algorithm used makes it possible to partition the large datasets, the script used for the clustering is processed with python (Scikitlearn).

Assignment of the new cv candidate to the adequate cluster: using the cosine similarity between the new cv and the cluster.

d) Matching

The extracted features of the job offer for a specific work area are mapped with the CV in the same work area, based on the measure of similarity of the features technologies skills, age, Educational Attainment. The measure of similarity will be applied using the cosine similarity; the formula for measuring cosine similarity is the following:

$$\sum_{i=1}^{n} A_i \times B_i \Big/ \sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}$$

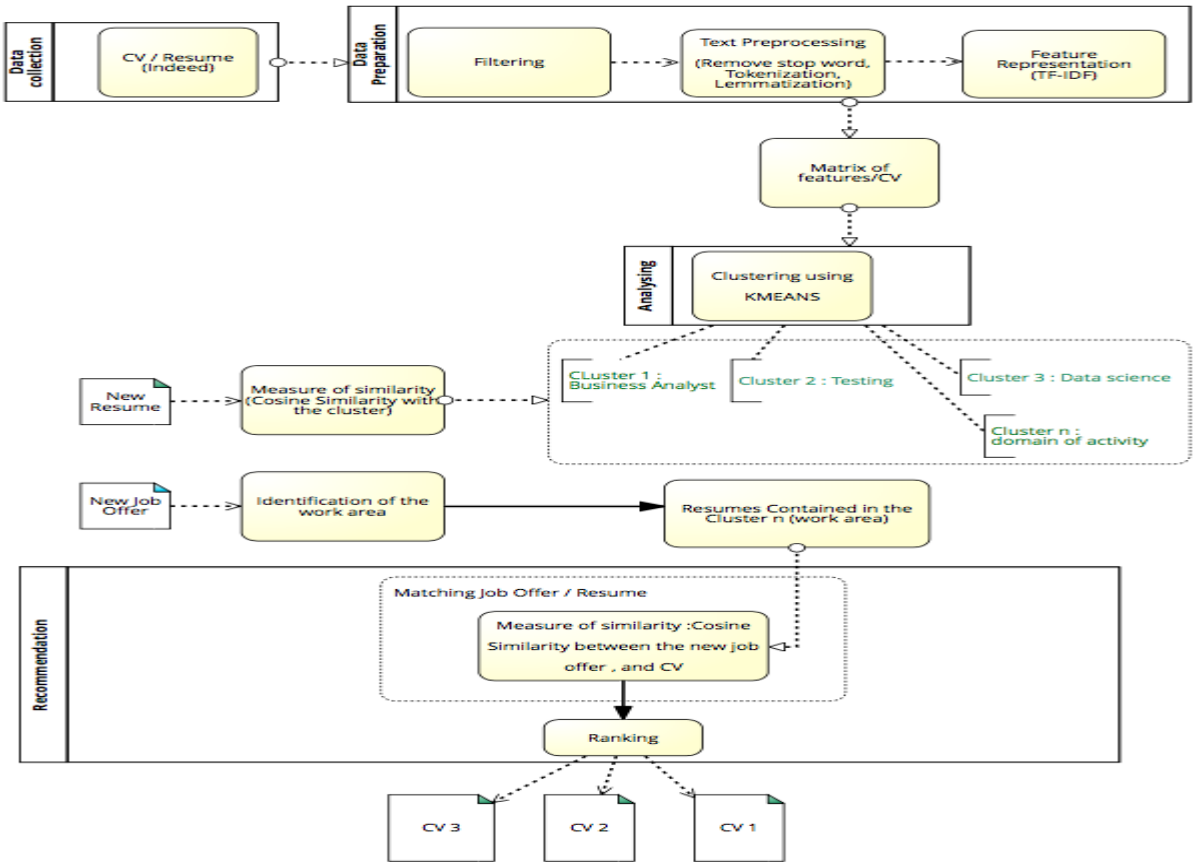

Fig.3. Recommendation System Framework

e) Recommendation

The decision of recommendation: for a new job offer, the matching result will be sorted by the matching score result from the measure of cosine similarity, the highest matching score involves the best matching, then the CV is recommended for the job offer in the input.

B. Evaluation of the recommendation System

To evaluate the accuracy of our recommendation system, we will take a manual evaluation for almost 60 employees who are recruited in the IT services company. Strategy: We have 60 employees who are recruited 6 months ago; we will measure the accuracy of our system as follows:

- We will upload the 200 CVs of the employees in our system.
- The CV is assigned to the adequate cluster (work area) (business analysis, testing, development, etc.).
- We upload a new job offer for the specific work area.
- The recommended resumes for each job offer are obtained.
- We compare the result of the recommendation with the actual situation.

V. CONCLUSION

In this paper, we have proposed a new approach to recommend the potential resumes for the recruitment; the proposed recommender system is developed for IT services company, in order to accelerate the recruitment process and improve its subjectivity. The implementation of the recommender system is in progress, the proposed approach uses different methods as TF-IDF for the preprocessing, KMeans clustering to group the dataset of resumes in a different cluster of the work area, and cosine similarity to measure the similarity of the new job offer with the resumes contained in the work area.

As a perspective of our research, we will take into account the progression of the competence ontology for matching job offers and resumes. Also, we will use many training data to test the accuracy of our system and introduce the results. In the future, an Interface integrating the proposed approach will be designed and implemented, we validate it with the live data sets of resumes.

References

[1] Derous, E. and Ryan, A.M., By any other name: discrimination in resume screening, The Oxford Handbook of Job Loss and Job Search. (2016).

[2] Bafna, P.B., Shirwaikar, S. and Pramod, D., Multi-step iterative algorithm for feature selection on dynamic documents," International Journal of Information Retrieval Research (Research), 6(2)(2016) 24-40.

[3] Sun, T. and Vasarhelyi, M.A., Embracing textual data analytics in auditing with deep learning The International Journal of Digital Accounting Research., (2018).

[4] Kawtar, N BDIoT'19: Proceedings of the 4th International Conference on Big Data and Internet of ThingsOctober (3) (2019) 1–5.

[5] Farhin Mansur, Vibha Patel, Mihir Patel, A Review on Recommender Systems, International Conference on Innovations in information Embedded and Communication Systems (ICIIECS) (2017).

[6] A. Brun, A. Hamad, O. Buffet and A. Boyer . Vers l'utilisation de relations de préférence pour le filtrage collaboratif, Actes du dixseptième congrés francophone AFRIF-AFIA sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'10), Caen, France, (2010).

[7] Z. zaier,these : modèle multi-agents pour le filtrage collaboratif de l'information,JANVIER (2010).

[8] Poonam B. Thorat, R. M. Goudar, Sunita Barve, Survey on Collaborative Filtering, Content-based Filtering, and Hybrid Recommendation System, in International Journal of Computer Applications (0975 – 8887) 110(4)(2015)

[9] Lops P., de Gemmis M., Semeraro G. Content-based Recommender Systems: State of the Art and Trends. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) Recommender Systems Handbook. Springer, Boston, MA, (2011).

[10] Jayabharathy, J. and Kanmani, S., Correlated concept based dynamic document clustering., (2014).

[11] Algorithms for newsgroups and scientific literature, Decision Analytics, 1(1)(2002) 199-206. Pantel, P. and Lin, D. "Document clustering with committees," Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM,.

[12] Jian Chen; Keke Li; Zhiheng Liu: Data Analysis and Knowledge Discovery in Web Recruitment—Based on Big Data Related Jobs, International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (2019).

[13] Ravindar Mogili, G. Narsimha Hierarchical Agglomerative Based Iterative Fuzzy Clustering to Impute Missing Values in Health Datasets, Intelligent System Design -Springer pp 605-613 (2020).

[14] Prafulla Bafna, Shailaja Shirwaikar, Dhanya PramodTask recommender system using semantic clustering to identify the right personnel, VINE Journal of Information and Knowledge Management Systems (2019).

[15] Santosh Kumar Majh, Shubhra Biswal: A Hybrid Clustering Algorithm Based on Kmeans and Ant Lion Optimization, Advances in Intelligent Systems and Computing book series AISC, 813(2018).

[16] Singh, A., Rose, C., Visweswariah, K., Chenthamarakshan, V. and Kambhatla, N., PROSPECT: a system for screening candidates for recruitment, Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, (2010) 659-668.

[17] Wolfswinkel, J.F., Semi-automatically enriching ontologies: a case study in the e-recruiting domain masters thesis, University of Twente., (2012).

[18] A Casagrande, F Gotti, G Lapalme :Cerebra, un système de recommandation de candidats pour l'e-recrutement . hal.archives-ouvertes.fr (2017).

[19] L.Azzopardi, YMoshfeghi: Lucene4IR: Developing information retrieval evaluation resources using Lucene.ACM SIGIR Forum (2017).

[20] A Białecki, R Muir: Apache Lucene 4 information retrieval. academia.edu (2012).

[21] P Darmon, R Mazouzi, O Manad :TeamBuilder: D'un moteur de recommandation de CV notés et ordonnés à l'analyse sémantique du patrimoine informationnel d'une société . hal.archivesouvertes.fr (2018).

[22] Suphakit Niwattanakul, Jatsada Singthongchai: Using of Jaccard Coefficient for Keywords Similarity, Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, Hong Kong (2013).

[23] Bin Cao, Ying Li, and Jianwei Yin., Measuring Similarity between Graphs Based on the Levenshtein Distance., Appl. Math. Inf. Sci. 7(1)(2013) L, 169-175.

[24] Abraham Bookstein, Vladimir A. Kulyukin: Generalized Hamming Distance. Information Retrieval 5, 353–375 (2002). https://doi.org/10.1023/A:1020499411651.

[25] Jorge Martinez-Gil, Alejandra Lorena Paoletti: A Novel Approach for Learning How to Automatically Match Job Offers and Candidate Profiles, Information Systems Frontiers, 2019 – Springer.

[26] Enrico P. Chavez: Feature Selection for Job Matching Application using Profile Matching Model, IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (2019)

[27] HAMZANE Ibrahim and Belangour Abdessamad, A Built-in Criteria Analysis for Best IT Governance Framework International Journal of Advanced Computer Science and Applications(IJACSA), 10(10)(2019). http://dx.doi.org/10.14569/IJACSA.2019.0101026

[28] HAMZANE Ibrahim and Belangour Abdessamad, Project Management Metamodel Construction Regarding IT Departments, International Journal of Advanced Computer Science and Applications(IJACSA), 10(10)(2019). http://dx.doi.org/10.14569/IJACSA.2019.0101029

[29] HAMZANE Ibrahim and Belangour Abdessamad, MDA Transformation Process from predictive project management methodologies to agile project management methodologies, International Journal of Emerging Trends in Engineering Research, 8(9)(2020).

[30] Annalisa Wahyu Romadon; Kemas M Lhaksmana: Analyzing TF-IDF and Word Embedding for Implementing Automation in Job Interview Grading, 2020 8th International Conference on Information and Communication Technology (ICoICT).