International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.14, No.1 (Jul-23)

http://dx.doi.org/10.12785/ijcds/140103

Towards a New Job Offers Recommendation System Based on the Candidate Resume

Hanae Mgarbi[1], Mohamed Yassin Chkouri[1] and Abderrahim Tahiri[1]

[1]*SIGL Laboratory, ENSATE, Abdelmalek Essaadi University, Tetouan, Morocco*

Received 5 Sep. 2022, Revised 6 May. 2023, Accepted 7 May. 2023, Published 1 Jul. 2023

Abstract: The recruitment web platforms are a solution that has solved several problems, such as eliminating the filing of paper files to organizations, reducing recruitment time, etc. In recent years and especially with the pandemic, the job offers increased across the web with a great diversity of professional platforms and networks offering these offers. This increase in the number of job openings and the number of web platforms presents difficulties for candidates. The candidates must register and create an account on each recruiting platform. This implies a considerable waste of time for candidates to manage their accounts and to find suitable jobs according to the candidate's profile on each platform. Most recommendation systems are oriented towards the jobs platform and not towards the candidate, and depend heavily on the jobs platform. So candidates can apply on job offers specific to their specialties and skills, we propose a recommendation system for the online job offers to allow the classification of job offers to correspond to the profiles (e-portfolio) of graduates or candidates according to their skills. Our research work is carried out from extracting to calculating similarity and ranking passing by specific processing for the preparation of our data set and the classification of job offers and e-portfolios according to the skill specialty.

Keywords: Recommendation system, E- recruitment, Similarity, Classification, NLP, VSM.

1. Introduction

With recent technological advancements, human dependence on the Internet has increased dramatically. Information is now mainly available and shared via the internet using sources such as websites, social networks, and web portals. This advance in internet technology has also had an impact on recruitment. The recruitment web platforms are a solution that has solved several problems such as eliminating the filing of paper files to organizations, reducing recruitment time, etc. In recent years and especially with the pandemic, we have noticed the increase in job offers across the web with a great diversity of professional platforms and networks offering these offers. This increase in the number of job openings and the number of web platforms presents difficulties for graduates. Foremost, to find suitable jobs that are relevant to the candidate's profile. Next, candidates must register and create an account in each recruiting platform. This implies a considerable waste of time for candidates to manage their accounts, fill in all the information concerning the e-portfolio and consult the job offers published on each platform. A study was carried out on the recommendation system for job offers (Teambuilder [1], SAJ [2], etc.). We find that most recommendation systems are oriented towards the company and not towards the candidate and depend heavily on the platform, which means, each recommendation system is linked with a platform offering the job offers. This increases the dependence between the platform and the recommendation system. We also note that some research works only use demonstration data and not real data. Faced with these complexities, and so that graduates can apply to job offers specific to their specialties and skills, we propose a recommendation system for job offers that is independent of the platforms and allows the classification of job offers corresponding to the profiles of graduates according to their skills. Our research work is based on extracting job offers data from websites according to a well-defined data model based on a set of criteria. The extracted job offers will be subject to specific processing for the preparation of the DataSet. Among these treatments, we cite removing stop words, conversion of non-numeric values, text tokenization [3], PostTagging [4], and named entity recognition (NER) [5], The classification of job offers and e-portfolios according to skill category is an important step in reducing the time needed to search for information and recommend jobs in the same skill category. Candidate profiles and job offers are represented as vectors in a vector space of n dimensions (n criteria) according to our data model. Each dimension of the space is a potential instance of the criteria used to describe job offers and profiles. After calculating the similarities between each job offer and candidate profile of the same skill category, the result is ordered in descending order of similarity to display the most

E-mail address:hanae.mgarbi@etu.uae.ac.ma, mychkouri@uae.ac.ma, t.abderrahim@uae.ac.ma

relevant at least relevant job offers for each profile. After all the processing has been carried out, our recommendation system of job offers allows users of the e-portfolio platform to find jobs classified in decreasing order of similarity. E-portfolio platform is a digital environment based on the e-portfolio approach, allowing students to build and enhance academic and extra-academic achievements. It's being part of a lifelong learning approach, to develop their digital visibility by capitalizing on training achievements and skills obtained, and to present themselves digitally. Our case study is based on extracting job offers in Morocco from the web. We have studied all the platforms offering job offers. We have collected over 4000 job postings from the websites that offer job postings. This paper is organized as follows. Section 2 gives an overview of the concepts of the recommendation system. We discuss the related works in the Section 3. Section 4 provides methods for extracting and analyzing and filtering data. Section 5 presents the proposed recommendation system of job offers. Conclusions close the article in Section 6.

2. Recommendation system

The recommendation system [6] is based on the measurement of the similarity between the user and the elements which are information objects and can be images, documents, web pages, etc. Classical information retrieval methods are based on vector modeling to obtain the relevance between the vector which represents the user and the other one representing the item.

A. Features of the recommendation system

The functionality of a recommendation system [7] relates to the nature of the field of application. We detail the following three features: Prediction: the system predicts the opinion that will have given a user to an item that has not yet been rated by this user. Prediction is often used in the case of collaborative recommendation systems. Recommendation: offers a list of items ordered according to the user's preferences. If the prediction must be done before the recommendation, a list of items with a higher match value will be recommended. However, several recommendation systems use the recommendation without going through a prediction step. This is the case with the majority of content-based recommendation systems, but also with some collaborative systems. Recommendation with a constraint: the user defines the constraints on the candidate elements for the recommendation, with limiting just on the elements satisfying the constraints defined to apply the functionality of the recommendation.

B. Recommendation systems techniques

We will present afterwards three techniques of recommendation system on content-based, collaborative and hybrid filtering.

1) Content-based system

A content-based recommendation system [8] recommends items to users with similar characteristics to items they have chosen in the past. The approaches used in content recommendation systems are generally inspired by the fields of documentary research and information filtering. The architecture of a content-based recommendation system consists of three components: the content analyzer, user profile learning, and item filtering (or recommendation).

2) Collaborative filtering system

Collective filtering [8] is a procedure that depends on the transmission of information between users. Although the term has been introduced for less than twenty years to rate an item through the exchange of opinions. With innovation, it becomes possible for a user to share their ratings with thousands or even numerous users, rather than just ten individuals. This is what collaborative filtering does by prescribing to the current customer items valued by e-service customers with whom they share similar preferences. Collaborative filtering strategies [9] base their suggestions on the historical context of users' inclinations for items, disregarding qualifiers describing clients and things. To give a group of things to the reviewed customer, the preferences of similar users (the people who assigned scores close to those named by the reviewed customer) are used to determine a measure of scores that the executive would relegate. Some calculations are used to check the similarities between clients / things: Pearson connection [10], cosine proximity [11], Spearman relation [12], etc.

3) hybrid filtering System

Content-based proposals and community suggestions have regularly been seen as integral. The content-based proposal makes it conceivable to suggest new things not yet appraised by any client, while community separating can possibly suggest a thing on the off chance that it has effectively been evaluated by a specific number of clients. The hybridization [8] of these two methods has been the subject of a few explorations that concentrate to address the weaknesses of every strategy and exploit their qualities. In this framework, for a thing to be prescribed for the current client two rules should be met: (a) its substance profile should be like the user's substance profile, and (b) the thing should be appraised by the neighbors nearest to the current user.

C. Approaches of content-based systems

As part of our research, we are relying on content-based recommendation systems. To do this, we will subsequently explain the techniques used for profile training and content-based recommendation techniques.

1) Learning the user's profile

There are a few methods utilized for client profile preparing in a content-based proposal framework, like the Vector Space Model (VSM) [13], Weighted Semantic Network [14], etc. We present in the following area the vector space model procedure. The vector space model is a model dependent on a vector space, which addresses users in a similar space as the one demonstrating items like terms, pictures, reports, questions, and so on by vectors in a vector space. Each component of a vector space addresses an attribute of an item compared to a fundamental component of a vector space of the VSM. Every part of an item vector mirrors the importance of the corresponding characteristic for an item vector. Vector spaces assume a significant part in intellectual science, design grouping and data recovery, where data recovery identifies with the techniques and methods of recovering and acquiring the necessary data from an asset or a collection of data. The VSM has an approach to gauge the likeness between an inquiry vector and a data object (among user and item) and to rank the items according to the similarity value by descending order.

2) Recommendation

As indicated by content-based proposal methods [15], the level of correspondence between item i to the user u ought to be estimated. Let item i which models the depiction of item i and let user u address the profile and preferences of this user u for the substance of the item. The profile is built by dissecting the substance of the items previously appraised by the user utilizing content investigation methods got from data research. Among these strategies, we refer to similarity [13], near neighbors [16], and classification [16]. Examination procedures of data or suggestion depend on vector displaying, which straightforwardly gives the significance of the proportion of similitude between the vector addressing the user and the one addressing the item, or for the most part between two data objects. The likeness is acquired by estimating the cosine of the point framed by the two vectors demonstrated in a similar vector space of the VSM, and each value is related to a term (A term is either a word or an idea (for unstructured things) or the value of a characteristic (for organizing things)) showing the significance of the term in the item. The value can be a boolean demonstrating the presence of the term in the item or a genuine showing the recurrence or the significance of the term in the item. As indicated by the theory, the user can be considered as an optimal item, so the more an item is like this client the more it is pertinent, in other words, it compares more to the necessities of the client.

$$sim(user(u); item(i)) = cosinus(\vec{u}; \vec{i}) = \frac{\vec{u} * \vec{i}}{||\vec{u}|| * ||\vec{i}||} \quad (1)$$

3. Related works

In the literature, there are only systems dedicated to the company to assign the suitable e-portfolios to the job offer proposed, there are no recommendation systems that are dedicated to the candidate and which seek the job offers published on the internet. In this section, we conducted a study on the recommendation system based on job postings. Four research works were selected for this study.

The processor project of the word [17] for the classification of job offers uses the explicit-rules and machine learning. This project consists foremost of the text preprocessing such as tokenization, lower case reduction, HTML substitution of special characters, removing of stop words, elimination of numbers, and the recovery of the word stemming. After preprocessing, the project uses the rule-based approach, which can be defined by the user based on taxonomies modeling terminology in certain domains. A matrix was built with the titles of the online job offers and the columns represent the characteristics (i.e., the word counts of the different words stemming) two machine learning classifiers were used to perform the text classifications. The technique developed is based only on job titles. It offers a set of documents and is based on two different stages: extraction and classification of characteristics. The objective of the feature extraction task is to derive for each job category a particular data structure, called Weighted Word Pairs (WWP) [18] and containing the most relevant pairs of elements by exploiting an automatic processing of classical natural language (NLP) as well as the probabilistic dependencies characterizing the titles of the job offers. On the other hand, the classification procedure is based on the matching between the terms taken from the job title job and the set of pairs linked to the different ISTAT categories [19]. For each category, the number of matches is determined, and each move is weighted by the probability dependence of the related pairs in the WWP structure. The category with the highest score is finally selected as the winner and selected for classification.

The SAJ project [2] presents the method of extracting data from the job offer description and from the candidate profile in order to be able to match them; they used the Linked Open Data system, ontology job posting description domain, and domain-specific dictionaries for data mining. SAJ enriches and builds the context between the extracted data to minimize the loss of information in the extraction process. Unstructured job description text extracted from any document format, such as MS Word or PDF. Plain text is extracted using Apache Tika[1]. Then the text is segmented into predefined categories using a self-generated dictionary. Automatic Natural Language Processing (NLP) and the dictionary help in data identification. Data entities are passed to two parallel processes, Context Building and Entity Enrichment. The result of these two processes is integrated and stored in the knowledge base. On the other hand, SAJ not only extracts job description entities but also enriches them, unlike existing e-recruiting systems. The entities extracted by SAJ and their connections can facilitate the search and recovery, scoring, and ranking of candidates against the job description. SAJ combines various processes to extract and enrich contextual information from the job description in online recruiting.

The Teambuilder platform [1] offers an e-portfolio recommendation engine that is based on the micro-services architecture and uses ontologies combined with text mining, machine learning, and pseudonymization techniques that preserve privacy. Teambuilder allows you to make a multi-

[1] https://tika.apache.org

criteria reconciliation between an assignment sheet inserted by a user, with the e-portfolios previously stored. All this data is stored in a JSON file format in the MongoDB database. Teambuilder is also a semantic e-portfolio search engine, it is based on an ontology of skills, in particular the European standard EN 16234-1: 2016 [5] which describes 40 skills for IT professionals, and ontologies relating to company names, publisher certifications, diplomas and functions and their aliases for semantic disambiguation (skills with aliases, sectorized clients, etc.). The display of e-portfolios matching the search is based on two scores: Static score corresponds to the frequency and position of the skill in the portfolio (for example the skill mentioned in the title has the highest score), and the score dynamic: these are the scores calculated and used at the time of the execution of a search query weighting the static scores in order to obtain a new “dynamic score” allowing to sort the results and return the N best profiles.

The development of a set of predictive systems [20] called Work4Oracle is able to estimate the audience (number of clicks) that a job offer would get posted on Facebook, LinkedIn, or Twitter. These systems combine both recommender system techniques and machine learning methods. The results made it possible to quantify the factors that influence the audience (and therefore the attractiveness) of job offers posted on social networks. The authors used user data from Facebook (job), LinkedIn (position), and job posting (title). The reconciliation between job postings ’title’ to Facebook’s ’job’ and LinkedIn’s ’position’ is done using O * NET-SOC taxonomy [21]. The latter extracts the O * NET-SOC vectors and gives importance to the most significant data, then the vectors pass to a similarity function to help the system to predict the audience of job offers published on the networks. LinkedIn, Facebook, Twitter thanks to Work4. Work4 collects data concerning the fields to which the O * NET-SOC taxonomy applies while respecting the privacy of users. The authors used SVM (Support Vector Machines) to train the system to better select the audience for job offers.

4. Extraction, filtering and analysis of job offers data

In this section, we will present our research work for extracting data job offers from websites according to our data model defined based on a set of criteria. Subsequently, the extracted data will be subjected to specific processing for the preparation of our data set. Finally, we will classify job offers and e-portfolios according to the skill specialty.

A. Extraction of job offers data

Our case study is based on extracting job offers in Morocco. We have looked at all the websites offering job offers. Among these websites, we mention Rekrute[2], Emploi.ma[3], Linkedin[4], etc. Job offers from websites such as Emploi-public[5], dreamjob[6], and alwadifa-maroc[7] are collected in PDF format. Each job offer is presented as a PDF file. The collection of job offers from PDF documents does not represent the object of our research in this article. We will rely on the collection directly from websites. We have defined our data model according to the criteria we need in the context of our research, such as the job title; the area of the activity; the mission; level of education; years of experience; the city; type of contract; etc. The job offers collected are in a structured format and are extracted in Excel file format. We have collected over 4000 job offers from websites that offer job postings like Rekrute, Emploi.ma, and LinkedIn, using Data Miner[8].

B. Data pre-processing of job offers

The extracted job offers are subjected to processing using natural language processing (NLP) techniques to obtain high quality-data. Among these treatments, we performed the removing stop words, conversion of non-numeric values, text tokenization, PostTagging, and named entity recognition (NER) using the StanfordCoreNLP library [22]. We have converted the non-numeric values for educational level and years of experience, which are extracted in text format. For example, ”Bac + 4”, is converted to an integer format which is ”4”, the same has been applied for years of experience; For removing stop words, we’ve removed records that contain null values. For example, if there is a job title containing an empty value, the record is deleted, so as not to block subsequent processing; We applied the text tokenization to the job offer’s title, and mission to retrieve all the information from the job in token format. Tokenization is based on morphological analysis, which consists of segmenting the text into elementary units and also determining the different characteristics of these units. For the job title, two types of information can be retrieved, the profile title and the city. For example, we have retrieved the following two tokens ”back-end developer (Java)” and ”Casablanca” from the job title ”back-end developer (Java) — Casablanca”. Tokenization then applies to the profile title to extract more information if exists; Part-of-Speech Tagging assigns each token its category such as verb, adjective, noun, etc. using the StanfordCoreNLP POSTagger tool. For example, organization recruits analyst engineers : ”organization” receives its category which is ”Name”, ”recruits” receives the category ”verb”, and so on; The named entity recognition (NER) technique allows mapping between tokens (people, organizations, places, dates, times, etc.) and their categories. For example, “Web developer (Drupal) — Rabat ”, contains the word” Rabat ”which corresponds to the name of the city, which is identified by the method of named entity recognition.

[2]https://www.rekrute.com/
[3]https://www.emploi.ma/
[4]https://fr.linkedin.com/
[5]https://www.emploi-public.ma/
[6]https://www.dreamjob.ma/
[7]http://www.alwadifa-maroc.com/
[8]https://dataminer.io/

C. *Filtering and data analysis*

For a good understanding of the fields of the data obtained, we must analyze and filter the data since they are very clean and can be explored. This step allows us to better understand the different behaviors. The result obtained from the NLP processing of the job title and mission is ready to send to WordNet ontology and YAGO2 ontology [4] in order to extract the skills. Then, the TF (Term Frequency) weighting is applied to the set of skills in order to retrieve the number of occurrences of each skill on the job offer, which reflects the importance of the skill requested. After obtaining the weighting of the skills of a job offer, we pass to identifying the areas of belonging (skill specialty) from the skills (or skill alias). For this, we need an integrated knowledge base that combines the DICE competence center and a standardized hierarchy of occupational specialties O*NET [23]. In this context, we use DICE to classify skills that belong to the field of Information, Communication Technologies (ICT), economics, and O*NET to classify skills related to the Medical and Artistic domains. A comparative analysis [23] was carried out between the DICE and O*NET classifications. Some skills acronyms are not recognized and not classified by O*NET, for example, JCA, JPA which refer to ”Java Connector Architecture” and ”Java Persistence API” respectively, are classified in the specialty ”Nursing Assistants” and ”Accountants” By O*NET, on the other hand, they are classified in the specialty ”Software development”. The skills ”Radiography” and ”Medical analysis” are classified respectively ”Technicians in radiology” and ”Medical and clinical laboratory categories” in O*NET, but are not classified in DICE. For the classification of job offers according to the skill specialty, the skills (Aliases) collected through previous processing are sent to the skill's knowledge base. We get then the specialties to which they belong, to each skill. A skill can belong to none, one or more specialties. We are obtaining as output a list of skills specialties that may contain one specialty or more. The classification of the skill specialties in the list will be made according to the skill weighting. As a result, our data model will be enriched by the list of skills with their weightings and the list of skill specialties also ordered according to the weighting of its skill. For the classification of e-portfolios according to skill specialty, we extract first the skills through the e-portfolio platform. Then, the skills with their candidate expertise levels (20%, 40%, 60%, 80%, 100%) are sent to the skill's knowledge base to retrieve in return the specialty belonging to each skill. We get a list of skills specialties, which is ordered according to the level of expertise for each skill.

5. Job offers recommendation system

Our job offers recommendation system is based on the architecture described in Figure 1. We have already described the steps of data extraction, filtering, and analysis. We will then explain the steps of calculating similarity, ranking, and the role of the e-portfolio platform.

A. *Similarity calcul and ranking*

After the stage of filtering by skill specialty for job offers and candidate profiles, we come to the most important stage of building the recommendation system (is the content-based system) based on a vector space model. Candidate profiles and job offers are represented as vectors in a vector space of n dimensions (n criteria) according to the data model that we have constructed. Among these criteria, we cite profile title; skills; level of education; years of experience; type of contract; area of activity; city; etc. Each dimension (criterion) of the space is a potential instance of the criteria used to describe job offers and profiles.

Similarity applies to all job offers and e-portfolios that have the same skill specialty with the greatest weighting. For example, the job offer which has the greatest weighting in the specialty ”Software development”, the similarity is applied on all e-portfolios, which have the specialty ”Software development” with the greatest weighting compared to the other specialties mentioned on the e- portfolio. Subsequently, the similarity is applied to the second specialty which has the weighting of its competence in the second degree of importance and so on.

1) *Cosinus Similarity calculator*

For the similarity calculation, a measure of similarity of the characteristic vectors must be defined. This similarity measure is required to determine the proximity between the profile and the job offer. We choose the use of cosine similarity after a study carried out similarity measures, according to the following formula:

Each profile “p” and job offer “o” are represented by a vector in a n-dimensional vector space, such as:
p = $\{v_1 ; v_2; \ldots ;v_n\}$, where v_i is the weight of the criterion i in the profile p.
o = $\{w_1 ; w_2; \ldots ;w_n\}$, where w_i is the weight of the criterion i in the job offer o.

$$sim(p, o) = cos(\vec{p}, \vec{o}) = \frac{{}_{i=1}^{n} v_i * w_i}{\sqrt{{}_{i=1}^{n} v_i^2} * \sqrt{{}_{i=1}^{n} w_i^2}} \tag{2}$$

For implementing the formula of cosine similarity, we used the Python language to program the similarity calculation and make the desired classification, which is a very rich language in terms of artificial intelligence algorithms and allows managing very large databases. We based on the Sklearn Library [24] which is an open-source machine learning library that contains a wide range of algorithms and supports supervised and unsupervised learning. It also provides various tools for model fitting, data preprocessing, model selection and evaluation, and many other utilities. Our case requires the treatments found in the submodule ”sklearn.metrics.pairwise” to implement utilities to evaluate pairwise distances or the affinity of sets of samples.

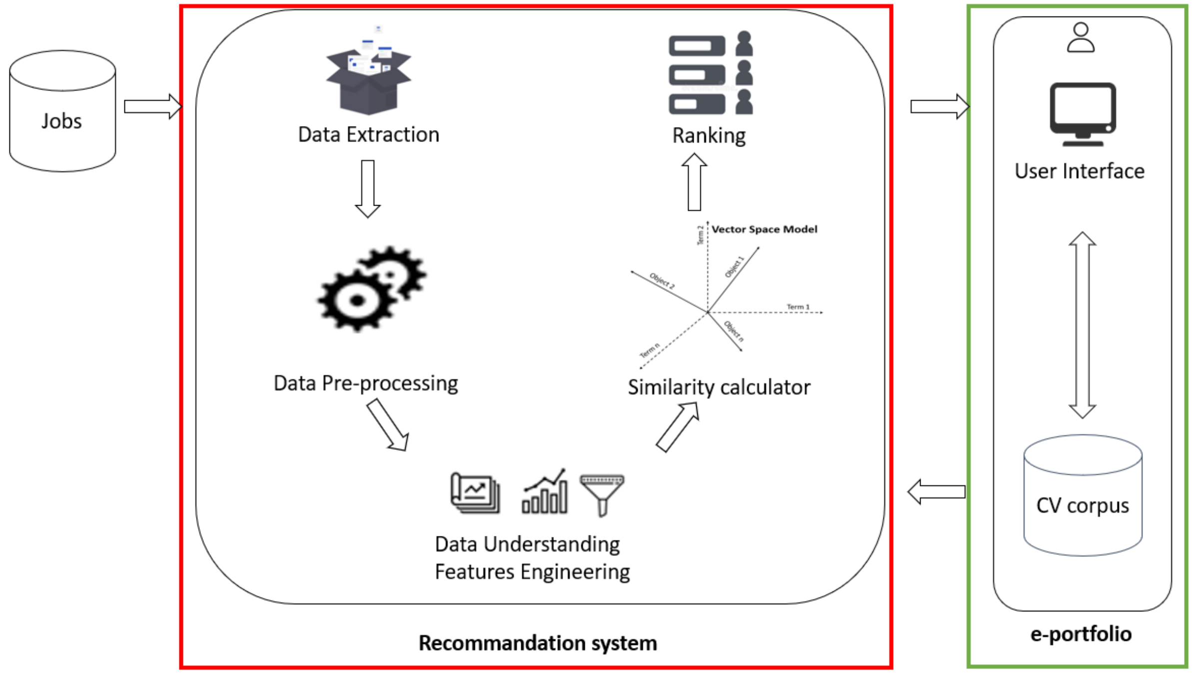

Figure 1. General architecture of the recommendation system

2) Euclidian Similarity calculator

We like to calculate also the similarity using the Euclidean distance between the profile vector and the job offer vector. In the end, we will compare the result of the two algorithms and choose the most efficient. To apply the similarity by using the Euclidean distance, we used the "Cross Distances" operator with the RapidMiner platform, as shown in Figure 2. The input is characterized by two elements, "request set" represented by all job offers and "reference set" represented by all e-portfolios. The output is made up of 3 objects: the first is the result of the similarity between each job offer of all the 'request set' job offers and each candidate profile of the 'reference set' CVThese object is named "result set" and it is in the form of a data table, the second is the set of input "request set" and the third is the set of input "reference set". To work better, we must have the attributes of "request set" and "reference set" in the same order. The "request set" and "result set" must have an ID, in case there is no "Cross Distances" operator able to generate it. [9]

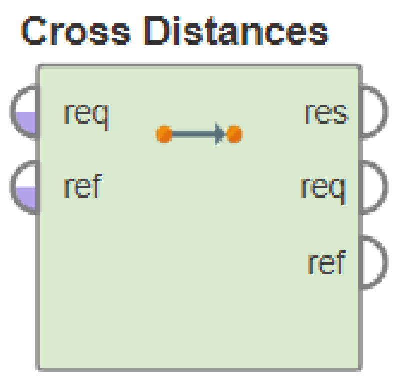

Figure 2. The cross Distances operator

3) Ranking

After calculating the similarities between each job offer and candidate profile of the same skill specialty, the result is ordered in descending order of similarity to display the most relevant at least relevant job offers for each profile.

B. Interaction with e-portfolio platform

The e-portfolio [25] can be defined as an evolving set of documents and electronic resources, capitalized in a digital environment, describing and illustrating the learning, experience, skills, or background of its author. Accessible remotely via interoperable technology, it is based on a personal database (information, documents, or links accessible via the Internet) and one (or more) collective selective publication space (s) . E-portfolio platform is a digital environment based on the e-portfolio approach, allowing students to build and enhance academic and extra-academic achievements while being part of a lifelong learning approach, to develop their digital visibility by capitalizing on training achievements and skills obtained, and to present themselves digitally. The students can achieve their personal projects: personal, professional, and training courses (Initial, continuing, and lifelong learning) and the development of the culture of digital identity. To complete

[9] https://rapidminer.com/

the e-portfolio, the student can add all the information about his e-portfolio through several sections proposed by the platform. After completion, the student will be able to share his e-portfolio with companies in both PDF and web formats. The web e-portfolio contains the video and links to the student's supporting documents, while the PDF e-portfolio contains only textual information. E-portfolios are stored in JSON format in the MongoDB database, from which we import the data to the recommendation system. After all the processing operations have been carried out, our job offers recommendation system gives users of the e-portfolio platform the job offers classified in decreasing order of similarity.

6. Conclusion

In this article, we made a job offer recommendation system based on the candidate's resume (e-portfolio). This work was done by passing on several steps. Firstly, we extract data job offers from websites according to our data model defined based on a set of criteria. Subsequently, the extracted data were subjected to specific processing for the preparation of our data set as text tokenization, post tagging, and named entity recognition (NER). The next step was to classify job offers and e-portfolios according to the skill specialty. After calculating the similarities between each job offer and candidate profile of the same skill specialty, the result is ordered in descending order of similarity to display the most relevant at least relevant job offers for each profile. After all the processing operations have been carried out, our job offers recommendation system gives users of the e-portfolio platform the job offers classified in decreasing order of similarity. The next step of this work is to extend our job offers recommendation system to other types of data as PDF, to have a complete recommendation system. Also, we plan to integrate the Arabic language into our recommendation system. And evaluate the result of two similarity methods and chose the method which returns the best results for our case.

References

[1] P. Darmon, R. Mazouzi, O. Manad, M. Bentounsi, and S. Berkani, "Teambuilder, vers un moteur de recommandation de cv scorés et rankés," 06 2018, big Data Mining and Visualization, Journées des groupes de travail EGC.

[2] N. Ahmed, S. Khan, K. Latif, and A. Khattak, "A new approach to information extraction in user-centric e-recruitment systems," *Applied Sciences*, vol. 9, p. 2852, 07 2019.

[3] J. J. Webster and C. Kit, "Tokenization as the initial phase in nlp," in *The 14th International Conference on Computational Linguistics*, 1992.

[4] A. B. Kmail, M. Maree, M. Belkhatir, and S. M. Alhashmi, "An automatic online recruitment system based on exploiting multiple semantic resources and concept-relatedness measures," in *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, 11 2015, pp. 620–627.

[5] B. Mohit, *Named Entity Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 221–245. [Online]. Available: https://doi.org/10.1007/978-3-642-45358-8_7

[6] R. Mishra and S. Rathi, "Efficient and scalable job recommender system using collaborative filtering," in *ICDSMLA 2019*, A. Kumar, M. Paprzycki, and V. K. Gunjan, Eds. Singapore: Springer Singapore, 2020, pp. 842–856.

[7] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, *Collaborative Filtering Recommender Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 291–324. [Online]. Available: https://doi.org/10.1007/978-3-540-72079-9_9

[8] F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, 08 2015.

[9] J. Séguéla, "Fouille de données textuelles et systèmes de recommandation appliqués aux offres d'emploi diffusées sur le web," Ph.D. dissertation, 05 2012.

[10] B. Djordjevic, A. Mane, and E. Krmac, "Analysis of dependency and importance of key indicators for railway sustainability monitoring: A new integrated approach with dea and pearson correlation," *Research in Transportation Business Management*, p. 100650, 03 2021.

[11] B. Li and L. Han, "Distance weighted cosine similarity measure for text classification," in *Intelligent Data Engineering and Automated Learning – IDEAL 2013*, H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, and X. Yao, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 611–618.

[12] N. Kamnitui, C. Genest, P. Jaworski, and W. Trutschnig, "On the size of the class of bivariate extreme-value copulas with a fixed value of spearman's rho or kendall's tau," *Journal of Mathematical Analysis and Applications*, vol. 472, 11 2018.

[13] D. Kim, "Group-theoretical vector space model," *International Journal of Computer Mathematics*, vol. 92, pp. 1536–1550, 05 2015.

[14] P. Lops, M. de Gemmis, and G. Semeraro, *Content-based Recommender Systems: State of the Art and Trends*. Springer US, 01 2011, pp. 73–105.

[15] D. Werner, "Indexation et recommandation d'informations : vers une qualification précise des items par une approche ontologique, fondée sur une modélisation métier du domaine : application à la recommandation d'articles économiques. (information indexing and recommendation : toward a precise description if items by an ontological approach based on business domain modeling : application to recommander system of economic news)," Theses, University of Burgundy, Dijon, France, 07 2015. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01558628

[16] S. Zhao, S. Zhang, J. Liu, H. Wang, J. Zhu, D. Li, and R. Zhao, "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture*, vol. 540, p. 736724, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0044848621003860

[17] F. Amato, R. Boselli, M. Cesarini, F. Mercorio, M. Mezzanzanica, V. Moscato, F. Persia, and A. Picariello, "Challenge: Processing web texts for classifying job offers," in *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, 02 2015, pp. 460–463.

[18] F. Colace, P. Napoletano, L. Greco, and M. De Santo, "A query expansion method based on a weighted word pairs approach," vol. 964, 02 2013.

[19] G. Barcaroli, G. Bianchi, R. Bruni, A. Nurra, S. Salamone, and M. Scarnò, "Machine learning and statistical inference: the case of istat survey on ict," 06 2016.

[20] M. Diaby, E. Viennet, and T. Launay, "Toward the next generation of recruitment tools: An online social network-based job recommender system," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 821–828.

[21] "About occupational information network (o*net)," Tech. Rep., 2010. [Online]. Available: https://www.onetcenter.org/taxonomy/2010/list.html

[22] X. Cheng, X. Kong, L. Liao, and B. Li, "A combined method for usage of nlp libraries towards analyzing software documents," in *Advanced Information Systems Engineering*, S. Dustdar, E. Yu, C. Salinesi, D. Rieu, and V. Pant, Eds. Cham: Springer International Publishing, 2020, pp. 515–529.

[23] A. Zaroor, M. Maree, and M. Sabha, "A hybrid approach to conceptual classification and ranking of resumes and their corresponding job posts," in *Intelligent Decision Technologies 2017*, I. Czarnowski, R. J. Howlett, and L. C. Jain, Eds. Cham: Springer International Publishing, 2018, pp. 107–119.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, and G. Louppe, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 01 2012.

[25] A. L. Daunert and L. Price, *E-Portfolio: A Practical Tool for Self-Directed, Reflective, and Collaborative Professional Learning*. Dordrecht: Springer Netherlands, 2014, pp. 231–251. [Online]. Available: https://doi.org/10.1007/978-94-007-7012-6_13

Hanae Mgarbi Hanae Mgarbi is PHD student at the research laboratory of the Information System and Software Engineering (SIGL) at National School of Applied Sciences, and the computer engineering at the Abdelmalek Essaadi University, Tetouan, Morocco. (email: hanae.mgarbi@etu.uae.ac.ma).

Mohamed Yassin Chkouri Mohamed Yassin Chkouri is the leader and founder of the accredited research laboratory of the Information System and Software Engineering (SIGL) and the Head of Department of Computer Engineering since 10/2016 at National School of Applied Sciences, Abdelmalek Essaadi University, Tetouan, Morocco. (email: mychkouri@uae.ac.ma).

Abderrahim Tahiri Abderrahim Tahiri received his engineering degree and PHD in Computer Science from the University of Abdelmalek Essaadi-Morocco. Since 2010, he has been an Associate Professor in the Department of Informatics at the National School of Applied Sciences of Tetouan, School of Engineering at Abdelmalek Essaadi University. He has been a member of the school Council and was appointed Coordinator of the Informatics branch then head of department. (email: t.abderrahim@uae.ac.ma).