Matching Applicants with Positions for Better Allocation of Employees in the Job Market

Ziad Elgammal Abdullah Barmu Hamza Hassan Khaled Elgammal Tansel Özyer Reda Alhajj

Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey
Department of Computer Engineering, TOBB University of Science and Economics, Ankara, Turkey
Department of Computer Science, University of Calgary, Alberta, Canada
Department of Heath Informatics, University of Southern Denmark, Odense, Denmark

2021 22nd International Arab Conference on Information Technology (ACIT) | 978-1-6654-1995-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACIT53391.2021.9677374

***Abstract*—Nowadays most people who are looking for a job use the Internet, visiting websites like Linkedin or Indeed so they must face hundreds of recruitment companies and job ads. The process of applying for a job is time consuming especially in screening, preparing and attending tests and interviews. In addition, job applicants do not know which companies are most proper for them, this job-hunting strategy can easily lead to employment dissatisfaction or failure. therefore, it is more efficient to recommend a few most suitable jobs. Also, manual screening for a position is time consuming and expensive. Experienced recruiters may be able to speed up the process by noting patterns in the resumes. The aim is therefore also to identify these patterns so they can be implemented in the system. In this work we try to introduce a system that would help both recruiters and job applicants in the job-hunting process.**

***Index Terms*—natural language processing, job recommendations, matching systems, screening systems**

I. Introduction

Technology has enabled the easy manipulation and reproduction of electronically encoded information. One effect of this is that it has become very easy to modify, reproduce and send out applications for jobs with tailored resumes. This process was exacerbated by the increased numbers of graduates and the scarcity in available positions. It is not uncommon that companies who post-opening are receiving hundreds or even thousands of resumes for a given position. Processing these applications puts a significant strain on human resources departments and they look for expert help. A job posting might attract thousands of resumes, making it difficult and time-consuming for recruiters to select appropriate candidates[17]. Hiring a wrong person can cause great damage: “Steve Jobs had a saying that A players hire B players; B players hire C players and C players hire D players. It doesn’t take long to get to Z players. This trickle-down effect causes bozo explosions in companies.”

Thus, it is necessary to provide an automated process for recruiting assistance in the form of assessments of the set of candidates for a given job. This could be realized by integrating techniques from natural language processing [18], ma- chine learning and network analysis into a system which may enhance multiple aspects of the resume assessment process, leading to improved recruitment decisions and more satisfied employers and employees.

Matching and recommendations systems have been widely used in many domains and applications, with the goal of providing the best 2 compatible items from 2 or more different sets. With various implementations and parameters[19][20].

In this work, we introduce a job matching system that takes jobs description and CVs as input, analyzes them and extracts important traits using NLP. Match all these traits against each other by. And finally, a scoring system gives a score for each job-cv pair and rank them.

II. Related works

A job matching system is a system that matches candidates to jobs based on a number of factors, making the screening process much easier for recruiters and other hiring authorities.

Reference [1] and [2] explained the main problems in the process: recommendation problem, screening problem. Also, they covered the most common techniques and approaches that have been implemented by different researchers.

A. Recommendation problem:

A huge amount of job postings is scattered online across many websites and job portals. A candidate looking for a job will have to go through all these job ads until he finds the best job suitable for his preferences. this is a very long and inefficient process. In order to tackle this problem and recommend the right job for the right person; many systems were implemented.

Recommendations systems may be classified to four main categories:

1) Collaborative Filtering Recommenders (CFRs) [11]: The CFRs find users with similar interests as the target user and suggest recommendations to him/her based on theirliked items. The key function in CFRs is the computation ofsimilarities among users. This type of recommenders face thechallenge of cold-start.

2) Content-Based Recommenders (CBRs) [13]: Items are recommended based on similarities between the user profile and the job profile. A CBR analyzes a set of characteristics of items which are rated by the target user and build the profileof the interests of this user based on the features of the items which are rated by him. This type of recommenders mightface the problem of overspecialization.

3) Knowledge-Based Recommenders (KBRs) [12]: Recom- mendations are generated by utilizing domain knowledge. These systems have the advantage of enhanced reliability sincethey usually contain less noise.

4) Hybrid Recommenders (HR): HRSs combine two or more techniques to overcome limitations that face any technique from the previous ones.

B. Screening problem:

certain job ads may attract thousands of applications, whichmake the screening of all these applications a very long and tedious task; especially with the limited time HR departments have. To solve this issue, some information systems were

introduced to assist on the task with various techniques and implementations:

1) Boolean Search techniques: using only keywords to search and filter applications. These methods lack the capabilities to deal with problems related to natural language processing or capture underlying attributes.

2) Information extraction techniques: in which the system mines the resumes and the job profile in order to get the similarity between both resume and job requirements based on different features. Although this method works fine with feature extraction, it did not solve NLP related problems.

3) Machine learning techniques: using support vector machine algorithm (SVR) to learn how to rank candidates using previous screening decisions.

Following are some of the recent publications on this issue:

Leah and Enrico(2019) introduced a job matching system to help recruitment agencies in their work; First by identifying and extracting some main features from resumes and jobs to build a profile for each of them. Then using a clustering technique, profiles of applicants are matched against the jobs requirements [4].

Susilowati et al.(2018) developed a decision support system to help local government in the process of transfer of duties and careers. Based on the results of profile matching of 2 main criteria;Intellectual aspects and work aspects [5].

Anika and Deepak(2014) applied data mining techniques to provide job recommendation based on the applicant's profile matching as well as his preferences and behaviour [6].

Dheeraj et al.(2016) made a hybrid recommender system that exploits both job and user profile. Focusing on Job seeker profile and preferences, the system provides personalized job recommendation considering various aspects and techniques. The system ranks 3A items(actors, assets, activities) according to their relevance to the user [7].

Srika and R.Raja(2020) designed a hybrid system that provides personalized news recommendations using on a new profiling approach. based on(T-UP-Tree) algorithm which shows better retrival performance than personalized and content-based recommender systems [8].

Kanyanut and Kridsada(2020) tried to match older people with jobs profile considering a set of parameters like skills, gender, age, job type alongside with the job seeker health information.using a vector model to process these parameters and calculate the Euclidean distance measure between them [9].

Yi and Han(2020) utilized machine learning and text mining to provide personalized resume analysis and generate recommendations In large job fairs or job-matching events, providing a cv analysis that results in a score for 3 main indicators [16].

Since most of the previous mentioned implementations have its own weakness, or works only in a specific environment under some conditions to serve a certain task. we try to introduce a system that has the minimum pre-conditions and can work efficiently under any circumstances; providing recommendations to both job applicants and recruiters.

III. METHODOLOGY

Our system architecture depicted in Fig. 1 shows the main processes in our system.

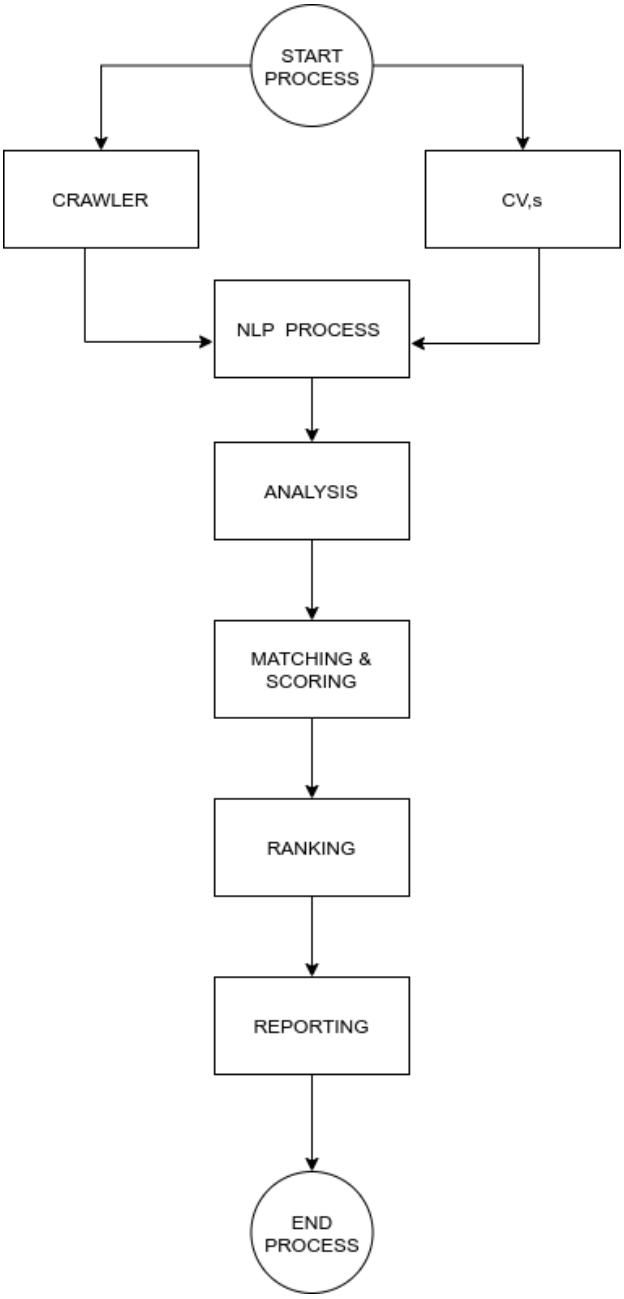

Fig. 1. system architecture.

TABLE I
SAMPLE OF NER RESULTS FROM A JOB CV

entity.text	entity.type
Aisha Yadav	PERSON
Bachelor of Engineering	EDUCATION
April 2016 to present	DATE
computer assistant	EXPERIENCE
HTML	skills
spring	skills
SQL	Skills
Ernst & Young	ORG
24 months - Python - Experience	EXPERIENCE

TABLE II
SAMPLE OF NER RESULTS FROM A JOB POSTING

entity.text	entity.type
software developer	Job Title
30.000$	MONEY
Russia	GPE
Russia	Language
12/12/2019	DATE
SQL	Skills
Full-Time	Employment Type
HTML	Skills

A. Data collection

Data were collected from different resources, having 2 main datasets, containing 14000 CV and 3000 job postings. In addation, a scrapper was implemented to collect job postings data from indeed.com.

B. Natural language processing

NLP is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages, it is used in major fields such as machine translation speech recognition and text processing. NLP is mainly concerned with making human and computer interaction easy, by process and analyzing large amounts of natural language data [14] [15]. in this work, NLP is used to extract all important aspects from both CVs and job postings. After testing various NLP libraries and tools, results were not as desired; we decided to use SpaCy NLP library since it allows to train NER process in an easy and flexible way in order to auto-detect new entities such as experience and skills from both CVs and job postings [3]. By training NER process; we developed 2 NLP models which allowed them to auto-detect new entities such as experience and skills from both CVs and job postings. the training dataset contained 147 record for CVs model, and 470 record for the job model. TABLE 1 and TABLE 2 show a sample of NER results on a CV and job posting respectively.

C. Analysis

the main purpose of this process is to prepare the extracted data and store it for later usage. Before being used in matching, NLP results are pre-processed; they are cleaned and organized and saved. This includes transforming all letters to lowercase, removing duplicated words and extra spaces, mapping some special characters and removing others.

D. Matching and Scoring

For each job posting, we try to find how much each of its components matches the corresponding component in each CV. First we calculate the similarity between each category in the CV and its corresponding one in the job posting using SpaCy Similarity function. Equation(1) shows how the score is calculated for each component:

$$score = onesSum + \frac{positiveSum}{length(Tokens1 + Tokens2)} \quad (1)$$

The final score for each (job - CV) pair is the sum of scores for all components.

E. Ranking and Reporting

In this stage, the scores obtained are used to rank candidates or jobs according to the required task. When recommending candidates for a certain job to the recruiter; after ranking scores, a table is created showing the better candidates at the top, and the less suitable one at the end.

Finally in reporting phase, the system prepare a summary of all credentials of the highest-ranking CVs for the job.

IV. RESULTS AND DISCUSSION

In this research dataset, there were more than 14 thousand CVs and 30 thousand jobs posting. Python was used as implementation language for all processes. BeautifulSoup[10], pandas, requests were used in the implementation of the scrapper. Also, SpaCy and other python libraries were used in the NLP process and the following stages.

The following example shows how the system works and the workflow of data:
suppose we have a list of some candidate's CVs and job postings.
- First, CV's and job postings will be each sent to its NLP models in order to extract the needed features.
- extracted features will be pre-processed and saved for later use.
- saved results undergo the matching process, in which a score is given for each 2 corresponding features in the job-CV pair.
- A final score is given for each pair.
- Finally, the pairs are ranked according to their score.

Based on the final score, the system can either recommend an ordered list of most suitable jobs to a certain candidate, or recommend the most eligible candidates for a specific job to the recruiter.
A sample of the dataset(jobs and CVs) were matched manually based on found skills and experience, then compared with system results. Overall system accuracy was equal to 83 percent.

V. CONCLUSION

The overall system consists of 5 main stages: Scrapping, NLP, Matching & Scoring, Ranking, Reporting First, the scrapper collects job postings from Indeed.com website. Some of the obstacles with scrapping phase are that you should know the website structure to get the same exact data we want to have. These are sent alongside resumes to the NLP process, in which the system extracts all important information and identifies needed aspects such as (Skills, Experience, Location, etc...) Later at the matching and scoring process, the extracted information from cv's and job postings are matched, and a score is produced for each (job - cv) pair; the higher the score, the better match it is. Based on these scores, a ranking can be made of cv's for a specific job and vice versa, The outcome from our project will relieve hiring companies from diving into huge piles of CVs to identify potential candidates for a given job posting. This is a tedious task and mostly impractical if at all possible, especially in this era where the economic recession is severe and the number of qualified employees who have lost their jobs is rapidly increasing. Accordingly, the number of applicants for every job posted is expected to exponentially increase. Hence, the need for effective automated matching systems like the one developed in this project is in higher demand than ever before. As a closing remark, we can state that the work completed in this project is timely and has a social and scientific impact.

VI. FUTURE WORK

This work can be extended by guiding applicants who want to apply for a specific job, showing them all the required skills and knowledge they need, and providing them with a summarized document of all needed resources and references. Also the system can be enhanced by adding more training to NLP models allowing it to detect new features.

REFERENCES

[1] Landing on the right job: A Machine Learning approach to match candidates with jobs applying semantic embeddings (Unpublished master's thesis). NOVA Information Management School (NIMS) NIMS - Dissertações de Mestrado em Métodos Analíticos Avançados. (2019).

[2] Guo, S. (2015). ResuMatcher: A personalized RESUME-JOB matching system (Unpublished master's thesis). Texas A&M University.

[3] Honnibal, M. & Montani, I., 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

[4] L. G. Rodriguez and E. P. Chavez, "Feature Selection for Job Matching Application using Profile Matching Model," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 263-266, doi: 10.1109/CCOMS.2019.8821682.

[5] Susilowati, Tri & Anggraeni, Elisabet & Fauzi, Widi & Andewi, Yeti & Handayani, Andino & Maseleno, Andino. (2018). Using Profile Matching Method to Employee Position Movement. International Journal of Pure and Applied Mathematics. 118. 415-423.

[6] A. Gupta and D. Garg, "Applying data mining techniques in job recommender system for considering candidate job preferences," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, 2014, pp. 1458-1465, doi: 10.1109/ICACCI.2014.6968361.

[7] Bahl, D., Desai, V., Vibhandik, S. and Fatma, I., 2016. A Profile Based Job Recommender System.

[8] Amara, S. and Subramanian, R.R., 2020, March. Collaborating personalized recommender system and content-based recommender system using TextCorpus. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 105-109). IEEE.

[9] K. Homsapaya and K. Budsara, "Machine Learning for Older Jobseeker and Employment Matching," 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 2020, pp. 427-430, doi: 10.1109/ECTI-CON49241.2020.9158291.

[10] Richardson, L., 2007. Beautiful soup documentation. Dosegljivo: https://www. crummy. com/software/BeautifulSoup/bs4/doc/.[Dostopano: 7. 7. 2018].

[11] Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S., 2007. Collaborative filtering recommender systems. In The adaptive web (pp. 291-324). Springer, Berlin, Heidelberg.

[12] Pazzani, M.J. and Billsus, D., 2007. Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer, Berlin, Heidelberg.

[13] Burke, R., 2000. Knowledge-based recommender systems. Encyclopedia of library and information systems, 69(Supplement 32), pp.175-186.

[14] Chowdhury, G.G., 2003. Natural language processing. Annual review of information science and technology, 37(1), pp.51-89.

[15] Indurkhya, N. and Damerau, F.J. eds., 2010. Handbook of natural language processing (Vol. 2). CRC Press.

[16] Y. -C. Chou and H. -Y. Yu, "Based on the application of AI technology in resume analysis and job recommendation," 2020 IEEE International Conference on Computational Electromagnetics (ICCEM), Singapore, 2020, pp. 291-296, doi: 10.1109/ICCEM47450.2020.9219491.

[17] Ramasubbareddy, S., Walia, A. S., Luhach, A. K., Kannayaram, G., & Evakattu, S. (2021). Big Data Analysis on Job Trends Using R. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, *14*(1), 100-108.

[18] Liddy, E. D. (2001). Natural language processing.

[19] Nagarnaik, P., & Thomas, A. (2015, February). Survey on recommendation system methods. In *2015 2nd International Conference on Electronics and Communication Systems (ICECS)* (pp. 1603-1608). IEEE.

[20] Sharma, L., & Gera, A. (2013). A survey of recommendation system: Research challenges. *International Journal of Engineering Trends and Technology (IJETT)*, *4*(5), 1989-1992.