Tawzef: Improving Recruitment Portals Performance via AI Technology, A Comparative Analysis

Ass. Prof. Dr. Magdy Elhennawy
Higher Institute of Computers and Information Technology, Computer Depart. , El. Shorouk Academy, Cairo, Egypt
Email: dr.magdy.ezzat@sha.edu.eg
Layla Reda
Higher Institute of Computers and Information Technology, Computer Depart. , El. Shorouk Academy, Cairo, Egypt
Email: luila.reda@sha.edu.eg

Abstract:

Organizations can achieve their objectives by selecting appropriate employees with appropriate qualifications and skills, which can be well defined through the employment job description. However, most of the problems that employees face are finding a job that suits their qualifications, as for companies, the problem is searching for an employee that fits the job requirements. On the other hand, the appropriate connection between the job seekers and employers represents a real problem, since employers should search in tons of CVs and job description and skills to select the appropriate and related ones

Today's most online job portals, concentrated on how to automate the process of recruitment. Job seekers profiles include professional experience, education, skills, certifications, and upload resumes. Employer's profiles include company culture, products, and job openings. Employers can post job openings, which users can search and apply for directly through the platform. Job seekers can apply to jobs with their profile and can share updates, post articles, and engage with content shared by others through likes, comments, and shares. Job seekers can see who has viewed their profile. Online job portals, a web-based application, have emerged as a popular platform for job seekers and employers to connect with each other. They have become the appropriate platform for recruitment process. However, they faces a vital challenges. Receiving a huge amount of career and job search information, matching between both is a difficult and exhaustive process, leading to time and effort consuming besides a probability of mismatch.

This research focuses on how to employ AI and other latest technologies, resume recognition, matching between resumes and employer job description, to provide a comprehensive solution to the above challenges. The results shows that matching time. The measured results shows that using the AI technologies has minimized the matching time make it negligible regarding to the recognition time.

I. Introduction

The traditional methods of job search and recruitment have been replaced by the technology-driven solutions, like Online job portals that offer more efficient and effective ways of connecting job seekers and employers. Online job portal is a website where people who need jobs can find suitable job opportunities and companies looking for job seekers can find the right talent employees for their organizations **[1].**

A well-known example of job portal is LinkedIn. LinkedIn counts over 900 million users worldwide, 59.1%, of them are in the age range of 25-34. Every second, there are three new registrations on LinkedIn, or ninety each minute. Expected to have a user population of 1 billion by 2025. In 2022, the annual revenue of LinkedIn was $14.5 billion **[2].** To make them good enough, online job portals should have some basic properties such as: multilingualism, live Chat Support, user friendly interface, besides the fast response to satisfy a good performance for millions of posted requests on the portals **[3].** they can be classified according to different views: according to type (General, Industry-specific, Government, University …) or according to nature (Internship, Freelance Writing …), and others**.** They have become the optimal platform for job seekers and employers to find the right match for their requirements. However, even though online job portals are the today's most efficient, you can be overwhelmed by responses that do not match your requirements wasting valuable time effort and money. So, it may face some basic difficulties such as: 1) Vast, sometimes unwieldy amounts of career and job search information, leading to time and effort consuming besides a probability of mismatch. 2) Time-consuming, tedious process, besides filling out online applications can be cumbersome. 3) Privacy issues related to posting résumé online.

Always, online job portals will receive hundreds, if not thousands, of resumes. Most of these resumes will be from candidates who are not remotely qualified for the requested position. However, eliminating the qualified from the non-qualified applicant takes time and effort, and someone in your company must review these resumes **[4]**. That is why it is important to find an intelligent way to overcome the above drawbacks upon huge number of received resumes or CVs, in a way that minimum effort, time, besides a confident result can be achieved.

The advancements in artificial intelligence (AI) and other latest technologies have made online employment portals become more sophisticated and effective in overcoming above challenges. We focus on our developed system to leverage AI and other latest technologies to provide a comprehensive solution to the above challenges.

II. Online Job Portal Technical Characteristics

The goal of the **Online Job Portal** is to allow job seekers and employers to build their profiles to establish and document their social networks they know and trust professionally to support employers with a wealth of data to enable them align workforce supply with actual demands [**5**].

They provide a set of functionalities. Employers and job seekers can create their profiles. Job seekers profiles include professional experience, education, skills, certifications, and upload resume. Employer's profiles include company culture, products, and job openings. Employers can post job openings, which users can search and apply for directly through the platform. Job seekers can apply to jobs. Job seekers can share updates, post articles, and engage with content shared by others through likes, comments, and shares. Job seekers can see who has viewed their profile. On the other hand [**6**].

The goal of these websites is to match qualified candidates with available positions. To organize all these openings and candidates, many platforms employ **AI-powered recommendation algorithms**. These algorithms, referred to as matching engines, process information from both the job seeker and the employer to curate a list of recommendations for each **[7].**

III. Literature Review

A. Introduction to Artificial Intelligence, AI

In today's digital age, life styles are constantly seeking innovative ways to upgrade, enhance, improves, and drive growth. One technology that has emerged is Artificial Intelligence, AI. With its ability to analyse vast amounts of data, make predictions, and automate processes, AI has the potential to transform life styles **[8].** It encompasses a wide range of technologies, including machine learning, natural language processing, computer vision, and robotics. AI enables machines to learn from experience, adapt to new inputs, and perform tasks that typically require human intelligence **[8].** AI has the power to revolutionize numerous aspects of our life, specifically, business operations. From customer service and marketing to supply chain management and product development, AI can bring significant improvements and efficiency gains covering: customer Insights, process automation, predictive analytics, intelligent virtual assistants, fraud detection, and others **[8].** By embracing AI technologies strategically, organizations can unlock new opportunities, optimize operations, and deliver exceptional customer experiences. As the AI landscape continues to evolve, staying informed and adapting to emerging trends will be key to realizing the full potential of AI in business. Embrace the power of AI and embark on a journey of transformative growth today **[8]**

B. Introduction to AI Technologies

Named Entity Recognition, NER

Named Entity Recognition (NER) serves as a bridge between unstructured text and structured data, enabling machines to sift through vast amounts of textual information and extract nuggets of valuable data in categorized forms. The NER's primary objective is to comb through unstructured text and identify specific chunks as named entities, subsequently classifying them into predefined categories. This conversion of raw text into structured information makes data more actionable, facilitating tasks like data analysis, information retrieval, and knowledge graph construction **[9].**

NER can be broken down into several steps, namely: tokenization, entity identification, entity classification, contextual analysis, and post-processing. It has the ability to understand and interpret unstructured text. By identifying and classifying named entities, NER adds a layer of structure and meaning to this vast textual landscape **[9].** NER has a set of methods, each tailored to address the unique challenges of extracting and categorizing named entities from vast textual landscapes. Such methods can be: rule-based Methods, statistical Methods, machine Learning Methods, deep Learning Methods, and hybrid Methods **[9].** However, NER has a set of challenges, such as: ambiguity, context dependency**,** language variations**, d**ata sparsity**, and** model generalization **[9].**

TF-IDF (term frequency-inverse document frequency)

Frequency–inverse document frequency, tf–idf (also TFIDF, or TF–IDF), is a measure of importance of a word to a document in a collection or corpus, adjusted for the fact that some words appear more frequently in general.[1] It was often used as a weighting factor in searches of information retrieval, text mining, and user modelling. A survey conducted in 2015 showed that 83% of text-based recommender systems in digital libraries used tf–idf. Variations of the tf–idf weighting scheme were often used by search engines as a central tool in scoring and ranking a document's relevance given a user query. The specificity of a term can be quantified as an inverse function of the number of documents in which it occurs **[10].**

Cosine Similarity

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis. A document can be represented by thousands of attributes, each recording the frequency of a particular word (such as a keyword) or phrase in the document **[11].** Continuing with the example of the document vectors, where attributes represent either the presence or absence of a word. It is possible to construct a more informational vector with the number of occurrences in the document, instead of just 1 and 0. Document datasets are usually long vectors with thousands of variables or attributes.
Example: Given the vectors X (1,2,0,0,3,4,0) and Y (5,0,0,6,7,0,0), the cosine similarity measure for two vectors is given by:

Cosine similarity (|X, Y|)= $\frac{x.y}{||x||\ ||y||}$, where: $x.y = \sqrt{\sum(x_i.y_i)}$, i= 1, 2, .., n, and $||x|| = \sqrt{x.x}$, $||y|| = \sqrt{y.y}$, so: Cosine similarity (|X, Y|)= $\frac{x.y}{||x||\ ||y||}$ = 0.08

The cosine similarity measure is one of the most used similarity measures, but the determination of the optimal measure comes down to the data structures. The choice of distance or similarity measure can also be parameterized, where multiple models are created with each different measure. The model with a distance measure that best fits the data with the smallest generalization error can be the appropriate proximity measure for the data **[11].**

C. Introduction to Web Development

Web development is the process of creating websites or web-based applications that run on the internet. It involves a variety of technologies, languages, and frameworks that work together to produce the final product. Modern web development has seen a significant evolution in recent years, with new technologies emerging and existing ones becoming more advanced **[12].**

One of the most notable trends in web development is the rise of front-end frameworks dynamic, interactive web applications, like React, Angular, and Vue.js. These frameworks are built on top of JavaScript **[12].** Meanwhile, the growing importance of back-end technologies like Node.js, PHP, MongoDB, and Ruby on Rails allow developers to build powerful, scalable web applications that can handle large amounts of traffic and data. On the other hand, web developers must be familiar with a wide range of supporting tools and technologies, including version control systems like Git, task runners like Grunt and Gulp, and package managers like NPM and Yarn. They must also be familiar with cloud computing platforms like AWS and Azure **[12].** **Figure 1**shows the architecture of web application.

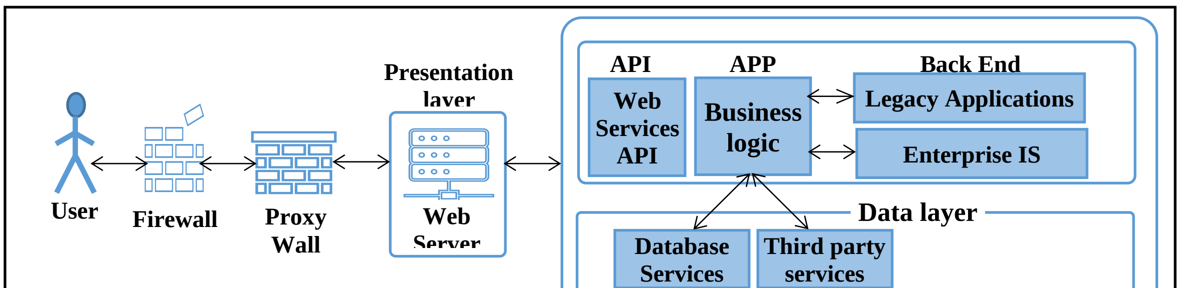


Figure 1: Architecture of Web Application

IV. Problem definition

A. Nature of the Problem

The today's online job portals are facing a huge number of **employer requests** (career) and **job searches** (CVs) submitted on the portals. Each employer request may contain more than one specialty and each job search may contain more than one specialty. Similarly, each of them will contain a set of skills for the employer to fulfill and for the CV owner to satisfy. To do the matching process, it requires a set of operations to be executed, such as: retrieve the **employer request**, retrieve the **job search**, matching each specialty in each of them, matching each skill in each of them, collect and sort the matched ones according the importance of each of them as stated in **employer request**, then select the most five or three tops of for more investigation if required. All of the above processes are done manually which is a time and effort consuming, in addition to the probability of error.

B. Problem Statement

The online employment portal system is facing several challenges that hinder its ability to effectively match job seekers with suitable job opportunities. In front of such challenges are the matching performance and efficiency, time and effort consumed in matching process, and matching result accuracy. Additionally, the system lacks a robust verification mechanism to ensure that the information provided by both employers and job seekers is accurate and up-to-date, leading to potential mismatches. These issues ultimately result in a low success rate for job placement and a dissatisfied user base. Therefore, there is a need to address these challenges and improve the functionality of the online employment portal system to increase user engagement, accuracy, and overall satisfaction.

C. Problem Formulation (Refer to **Figure 2** below)

Assuming that there is an n specialties and m skills for each candidate and u specialties and v skills for each employer to be matched together.

For the employer request:
Let SP_i denotes the i_{th} requested specialty, where i =1, 2, … n.
Let SK_j denotes the j_{th} requested skill, where j =1, 2, … m.

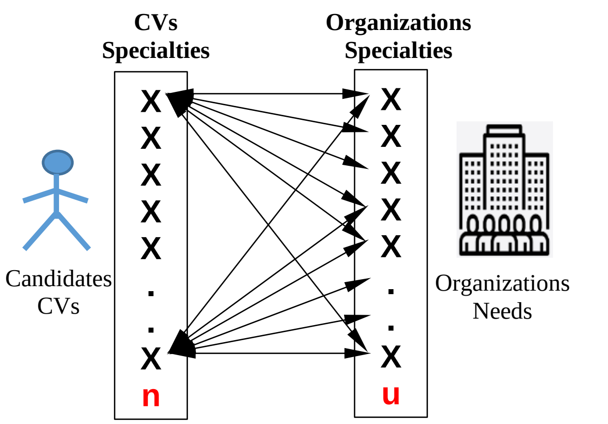
For the candidate request:
Let ESP_k denotes the k_{th} candidate posting his available specialty, where k =1, 2, … u.
Let ESK_l denotes the l_{th} candidate posting his available skill, where l =1, 2, … v.

It is required to match between SP_i and ESP_k for every i =1, 2, … n, k = 1, 2, … u, and to match between SK_j and ESK_l for every j =1, 2, … m, l = 1, 2, … v.

The number of matching processes required to do the specialties matching = n.u times
The number of matching processes required to do the skills matching = m.v times

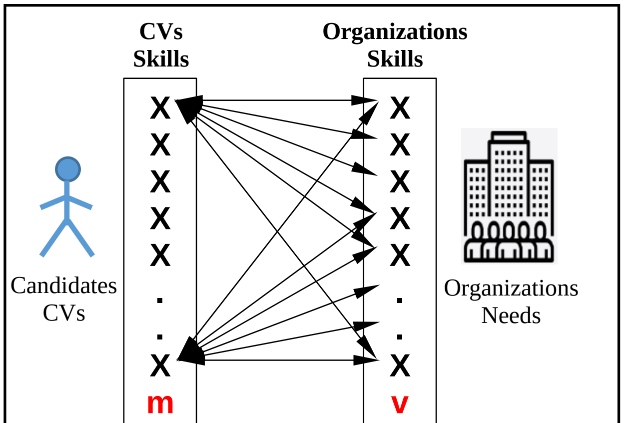
Let T_{sp} is the total time consumed in matching specialties in each of them together.
Let T_{sk} is the total time consumed in matching skills in each of them together.
Then, total time consumed in matching process is $T_{spk} = T_{sp} + T_{sk}$.

It is required to minimize T_{spk}, meaning the matching time is: **Min of {Sum ($T_{sp} + T_{sk}$)}** such that **{Sum ($T_{sp} + T_{sk}$)}$_{\text{AI-Automated}}$ < {Sum ($T_{sp} + T_{sk}$)}$_{\text{Traditional}}$**



Of matching process = n.u

Figure 2: Number of Matching n Candidates CVs Specialties with m Organizations Needs Specialties



Of matching process = m.v

Figure 2: Number of Matching n Candidates CVs Skills with m Organizations Needs Skills

V. Proposed System Specifications

The proposed system architecture outlines the framework for the enhanced online job portal referred to as **Tawzef Online Portal** (referred to as **TOP**). It incorporates modern technologies to facilitate the process of searching for a job that suits the employee's skills, help the company to find the right employee for the job, improve search accuracy, enhance user experience, and foster effective communication between job seekers and employers.

A. System Description

The proposed system consists of two main modules, one of them is responsible for CV and **resumes recognition**, the second is responsible about performing the job **matching** between job Seekers CVs' recognized pattern and uploaded job description of the company. The above modules are responsible for the tasks referred to as **specific tasks** of the portal, besides the other traditional modules that are responsible for other **basic tasks** such as: login, registration, searching, CV and resumes applications, and other traditional tasks.

As stated before, the most tedious tasks that consumes much time, much effort, with less accuracy are the **resumes recognition**, and the job **matching.** That is why the AI technologies have been invented and used in our system to empower the system and overcome the above difficulties. **Figure 3** shows the Proposed Tawzef Online Portal, ***TOP***, and its system functional description.

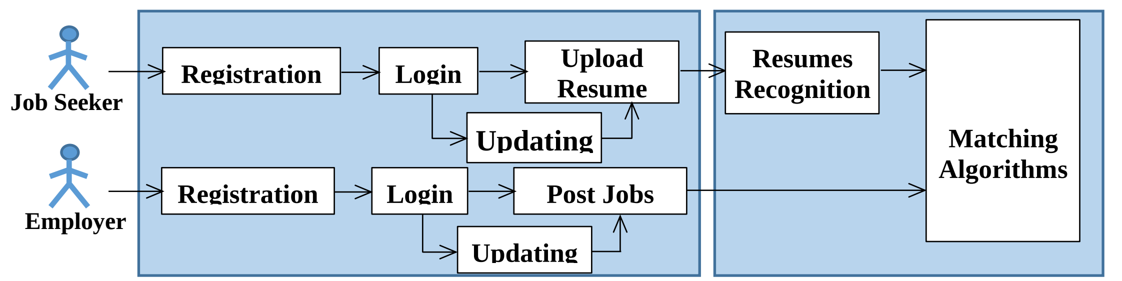


Figure 3: Proposed Tawzef Online Portal, *TOP*, System Functional Description

System Stockholders

There are mainly three stockholders in the application:

- **Job seekers:** called in this research also **Candidate.** The role of a job seeker on an online employment portal management system is to utilize the platform to search for and apply to job opportunities that match their skills, qualifications, and career objectives. Job seekers typically create an account on the portal, upload cv, build a profile, and actively engage with the available features and functionalities to enhance their job search process
- **Company:** called in this research also **Organization** or **Employer**. The role of a company on an online employment portal management system is to utilize the platform to advertise job openings, connect with potential candidates, and efficiently manage the recruitment process. Companies leverage the features and functionalities provided by the portal to attract qualified applicants and streamline their hiring efforts. The portal help company to Creating a Company Profile, Posting Job Openings, Managing Application, Communication with Applicants and Applicant Screening and Selection
- **Administration:** The system administration, ***Admin***: The role of system administration in an online employment portal management system is crucial for ensuring the smooth operation, security, and maintenance of the system. System administrators are responsible for managing

the technical aspects of the portal and ensuring that it remains functional and accessible to both job seekers and companies

B. System Requirements

The system business requirements can be classified as follows:

Basic requirements:

Uploading the career and jobs description on the online job portals. Applying the job seeker resumes on the online job portals. Allows employer and job seeker to define their profiles through registration functionality, logging-in, Job Search, view vacancies, see candidates, updating resumes and jobs description, and other traditional functionality.

Specific requirements:

Allow resume recognition via AI technology using the **pattern recognition algorithms**, instead of traditional posting way and allow executing matching process between career and job description via AI technology to overcome above stated difficulties.

Non-functional requirements

Enforcing the privacy issues related to posting résumé online.

C. Basic Functionalities (System Use Cases)

The following use cases describe the main functionalities in the system for both the ***jobseeker*** or ***employee*** and ***employer*** or ***organization,*** **Figure 4** shows the above actors use case for the employee, and **figure 5** shows the above actors use case for the organization, employer.

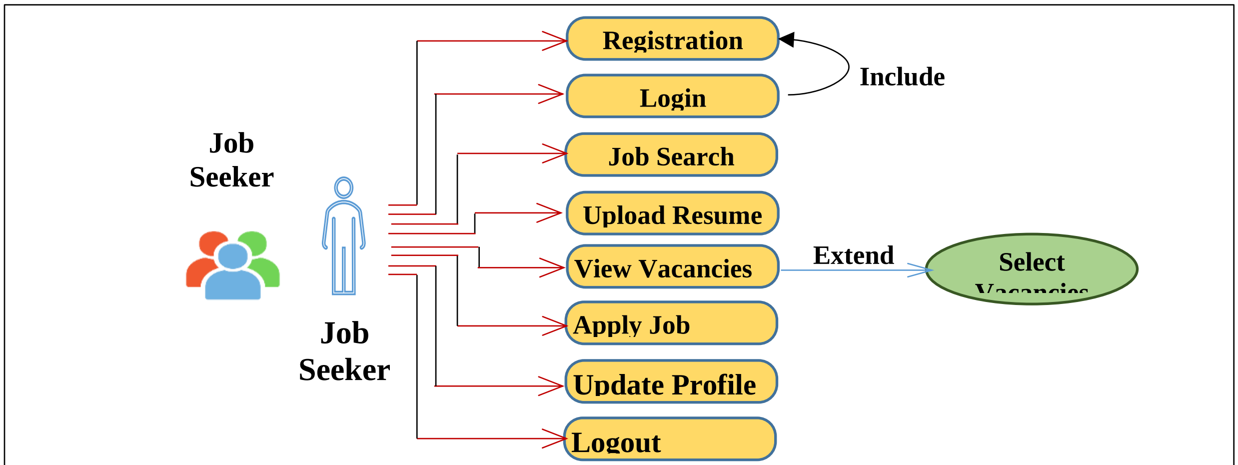


Figure (4): Current System Use Case for the Job Seeker

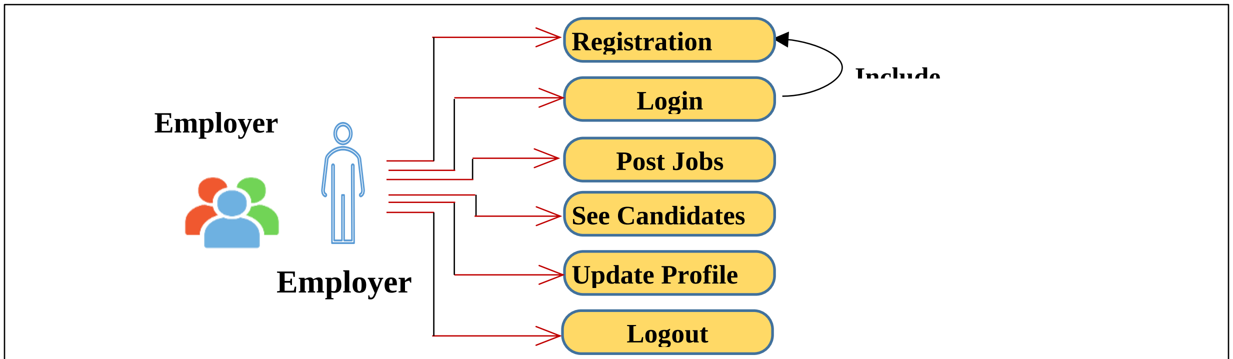


Figure (5): Current System Use Case for the Employer

D. Implementation Aspects

- **Extraction of information from resumes using Natural Language Toolkit (NLTK) and entity recognition.** The following steps have been followed:

  To extract information from a resume, various NLP and pattern-matching techniques are used. Names are identified through NLP methods like tokenization, POS tagging, NER, and chunking. Phone numbers are extracted using a regular expression that handles optional symbols and ensures a valid number format. Emails are captured using a regular expression that matches typical email structures. Skills are identified by first cleaning the text, removing stop words and punctuation, and then analyzing bigrams and trigrams to detect related terms. The processed text is then compared against a predefined skills dataset to identify relevant skills mentioned in the resume and **Annex (A)** show the detailed description of implementation

- **Extraction of information from resumes using the Heuristic Approach.** The following steps have been followed:

  This recognition algorithm extracts information from a CV using a combination of regular expressions and heuristic approaches. Emails are identified with a regular expression matching common email patterns, while phone numbers are extracted using another regular expression that handles various formats. Names are detected using a heuristic approach that assumes a name is a line where each word starts with an uppercase letter. For skills extraction, the algorithm removes stop words and punctuation, then uses bigrams and trigrams to identify related terms. Finally, the processed text is matched against a predefined skills dataset to identify relevant skills listed in the CV and **Annex (A)** show the detailed description of implementation

- **Matching using skills count** in which the job description contains the requested skills. The job seeker has applied to such skills. Looping over resumes of each job seeker that applied to this job and count the skills of job seeker that match the skills defined in the job description.

- **Matching using cosine similarity** in which the skills of each job seeker and the skills required in job description into a numerical data using tf-idf factorization, then calculating cosine similarity between the required skills and each candidate's (job seeker) skills.

IV. System Measurements, Results and Analysis

The issue for our system is the response of the system against the resume recognition and job matching. The above specific functionalities have been implemented and response time is measured using different recognition and matching algorithms.

Measurements Formulation

- The measured **Processing Time, PT,** will be calculated for one published job belonging to an employer, with n job seeker resumes applied for such job.

- In average, **Total Processing Time** can be calculated according to the following formula:

Total Processing Time = ∑ **{ (**Resume Recognition Time for one resume
+ Matching Time**) i, i = 1, 2, 00, n }**

That is:
TPT = ∑ { (RRT + MT)$_i$, i = 1, 2, 00, n }

Where:
TPT …... Total processing time.
RRT ….. Resume Recognition Time for one resume.
MT …… Matching time

The performance of matching time, ***MT,*** using *skills count* and using *cosine similarity* have been measured. **Annex (A)** show the actual results. **Figure 6** shows the relation of the processing time of both cases. It is clear that the performance of Skills Count is better than Cosine Similarity algorithms.

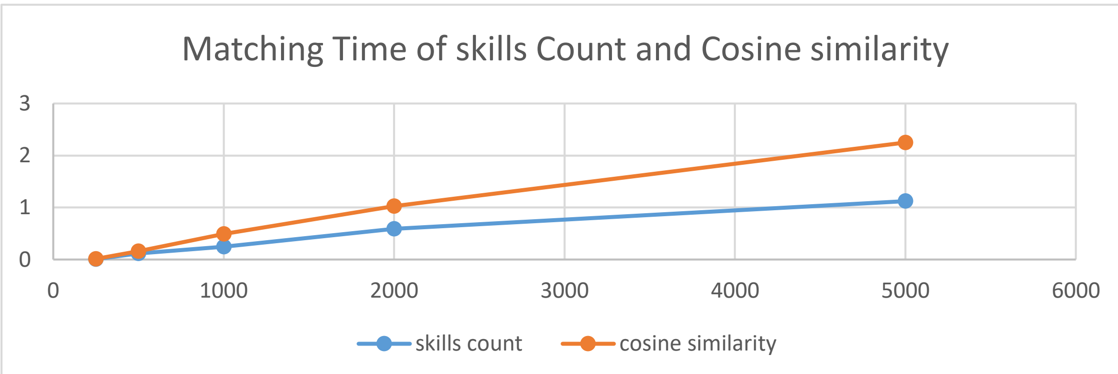


Figure 6: The Processing Time of the Matching Using Skills Count and Using Cosine Similarity

Relation between performance of total processing time, ***TPT***, have been measured using *skills count* and *cosine similarity*. **Annex (A)** show the actual results. **Figure 7** shows the relation between performance of total processing time, ***TPT***, using *skills count* and *cosine similarity*. It seems that they are identical, meaning that the matching time is negligible against recognition time.

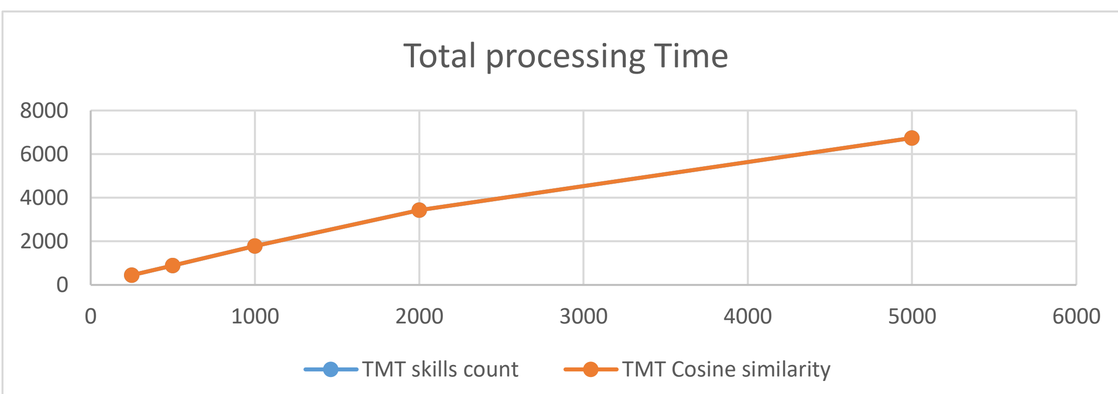


Figure 7: Relation Between Performance of Total Matching Time, ***TPT,*** using Skills Count and Cosine Similarity

Relation between performance of total matching, *TPT*, in case of using *skills count* with NLP and heuristic approaches. **Annex (A)** show the actual results. **Figure 8** shows the performance of total processing time, *TPT*, for both cases. It seems that the performance of heuristic approach is much better than NLP. Accordingly, the appropriate combination of the above approaches is to use the *skills count* for matching process and heuristic approaches for recognition process.

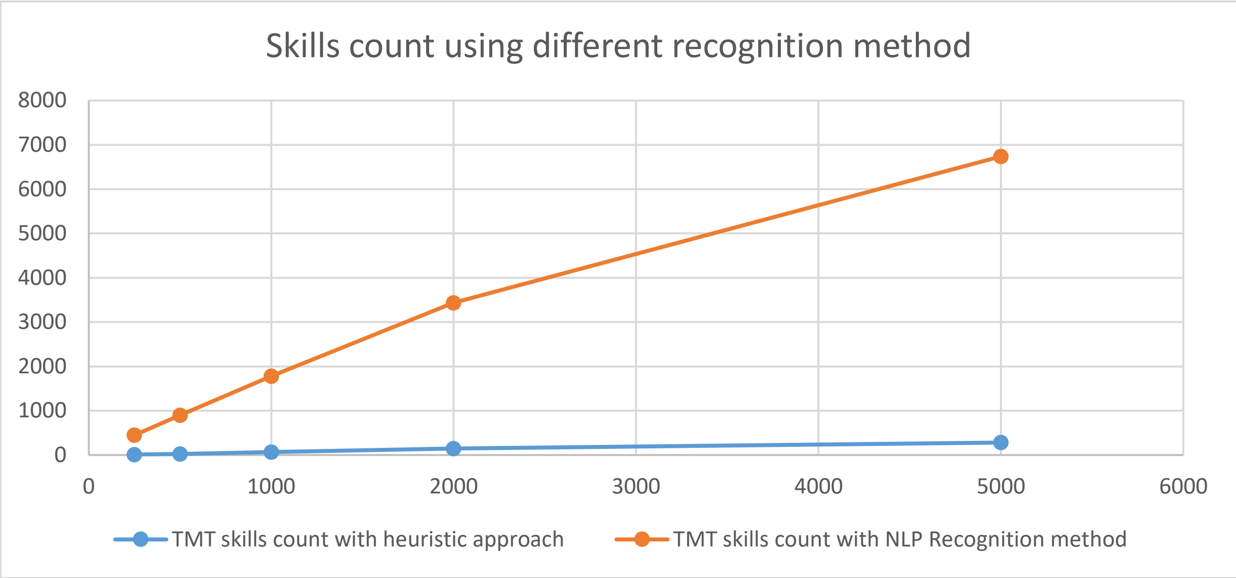


Figure 8: Relation Between Performance of Total Processing Time Using Skills Count with Different recognition Method (Heuristic and NLP Approaches)

VI. Conclusion

A big challenge facing human resources departments in various organization seeking for recruitment as well as the candidates seeking for the opportunity for jobs. Online job portals have solved such problem by automating the process of recruitment including automating the application of resumes, uploading the employer job description, matching between them, and other related functionalities. However, as the number of resumes and employer job description has increased dramatically, the processing time is a real challenge.

This research has introduced a proposal to solve such challenge by using AI technologies. According to the measurements results stated above, the performance of matching process has enhanced according to the used matching algorithm and recognition approaches. Attached **Annex (A)** shows the implementation and measurements details. The basic conclusions are as follows:

- **The first conclusion:** relating to the matching time, the performance of the skills count algorithm is better than the cosine similarity algorithm.
- **The second conclusion:** the time of recognition of the resume or CV is much slower than the matching time, that is why the difference in matching time of the skills count and the cosine similarity algorithms are not recognized.
- **The third conclusion:** When applying the recognition algorithms, heuristic and NLP approaches, the performance of the heuristic is much better.

- **The fourth conclusion:** The appropriate combination of approaches is to use the *skills count* for matching process and *heuristic* approaches for recognition process.
- **The fifth conclusion:** the automated system using AI algorithms is much better to fulfill the objective of improving the performance of the online job portals.

VII. Reference

[1] https://www.linkedin.com/?trk=article-ssr-frontend-pulse_nav-header-logo, March, 2023.

[2] https://www.linkedin.com/?trk=article-ssr-frontend-pulse_nav-header-logo,

[3] https://www.skillfuel.com/category/recruit-management/, The Essential Job Portal Features List of 3033, September 2022.

[4] https://smallbusiness.chron.com/, Disadvantages of Online Recruitment, By Chris Joseph, March 04, 2019

[5] https://www.techtarget.com/whatis/definition/LinkedIn

[6] https://www.linkedin.com/help/linkedin/answer/a548441

[7] https://www.technologyreview.com/2021/06/23/1026825/linkedin-ai-bias-ziprecruiter-monster-artificial-intelligence/, Sheridan Wallarchive, Hilke Schellmannarchive, June 23, 2021.

[8] https://revstarconsulting.com/blog/the-rise-of-ai-how-machine-learning-is-changing-the-face-of-business, Transforming Business with AI: A Comprehensive Guide

[9] https://www.datacamp.com/blog/what-is-named-entity-recognition-ner, "What is Named Entity Recognition (NER)? Methods, Use Cases, and Challenges", Sep 2023.

[10] https://en.wikipedia.org/wiki/Tf%E2%80%93idf, tf–idf, 17 March 2024.

[11] https://www.sciencedirect.com/science/article/pii/S0950584919301648, Information and Software Technology, 2019.

[12] https://www.techopedia.com/definition/23889/web-development, Web Development, 14 May 2024

Annex (A): Implementation and Measurements Details

- <u>**Recognition Using Heuristic approach**</u>
  - **<u>Extract email</u>** using regular expression -> "\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b"
  - **<u>Extract phone number</u>** using regular expression -> '\b\d{10}\b|\+?\d[\d -]{8,}\d'
  - **<u>uses a heuristic approach to extract names from text</u>**
    - Heuristics are rules of thumb or educated guesses that are used to solve problems more quickly when traditional methods are too slow or to find an approximate solution when exact solutions are not possible.
    - This approach uses a heuristic method, where it assumes that a name is likely to be a line where each word starts with an uppercase letter.
  - **<u>Extract skills</u>**

    We have a skills data set that contain all skills that may be included in cv the next paragraph shows the detailed Steps:
    - Remove stop word: in (NLP) is the process of eliminating words that occur frequently in a language but carry little or no meaning like "and", "the", "is", etc.,
    - remove the punctuation.
    - Bigrams are pairs of consecutive words, and trigrams are triples of consecutive words, so it is helpful to find related words like (web development, mobile application development and other related words in resume are together).

- We match the text after all these steps with the skills dataset to finally found the existing skills in this cv.

- **Recognition Using Natural Language Processing, NLP (using the Natural Language Toolkit, NLTK, and entity recognition)**
  - **Extract name**
    - Converting the image into text using nltk.sent_tokenize().
    - Iterating over each sentence in the resume.
    - Iterating over each token in sentence and the tokens are tagged with their parts of speech (POS) using **nltk.pos_tag**, it is called "chunk".
    - The chunked entities are checked to see if they have a label attribute and if that label is 'PERSON', indicating a person name.
    - Chunks is combined together into a person name.

  - **Extract phone number using regular expression**

    Regular expressions are used for pattern matching in strings. Phone number regular expression: [\+\ (]? [1-9] [0-9. \-\ (\)] {8,} [0-9]
    - [\+\ (]?: Matches an optional '+' or '(' at the beginning of the phone number.
    - [1-9]: Ensures that the first digit (after an optional '+' or '(') is between 1 and 9 (i.e., not 0).
    - [0-9. \-\(\)]{8,}: Matches any combination of digits, spaces, dots, hyphens, parentheses, or more of these characters, at least 8 times.
    - [0-9]: Ensures that the phone number ends with a digit.

    After applying the regular expression and find the phone numbers, a lot of validation must be done, such as:
    - Ensures that the found number is present in the original resume text.
    - Ensures that the length of the phone number is less than 16 characters.
    - Then finally we found the phone number.

  - **Extract email using regular expression**

    Also, by using regular expression: Email regular expression: [a-z0-9\.\-+_]+@[a-z0-9\.\-+_]+\.[a-z]+
    - One or more lowercase letters, digits, dots, hyphens, pluses, or underscores.
    - Followed by an @ symbol.
    - Followed by one or more lowercase letters, digits, dots, hyphens, pluses, or underscores.
    - Followed by a literal dot.
    - Followed by one or more lowercase letters (representing the TLD).
    - Then finally we found the email address

  - **Extract skills**

    The text contains the skills data set that is included in the resume. We have a skills data set that contain all skills that may be included in resume. The steps followed to extract the skills are:
    - Remove stop word: in (NLP) is the process of eliminating words that occur frequently in a language but carry little or no meaning like "and", "the", "is", etc.
    - Remove the punctuations.
    - Bigrams are pairs of consecutive words, and trigrams are triples of consecutive words, so it is helpful to find related words like (web development, mobile application development and other related words in resume).
    - We match the text after all these steps with the skills dataset to finally found the existing skills in the resume.

Measurements Actual Results:

Matching time using two different algorithms first algorithm is the skills count and the second algorithm is the cosine similarity.

Algorithm/Approach	**250 Cases**	**500 Cases**	**1000 cases**	**2000 cases**	**5000 cases**
Matching using skills count	0.00835	0.1173	0.24428	0.5886	1.1219
Matching using cosine similarity	0.0125	0.1589	0.4945	1.024	2.2497

Total processing time of applying the recognition algorithm and the matching algorithm using 2 different ways first is skills count and second is cosine similarity

	250 cases	**500 cases**	**1000 cases**	**2000 cases**	**5000 cases**
NLP RRT	447.85	895.05	1777.23	3433.81	6736.64
Matching using skill count	0.00835	0.1173	0.24428	0.4886	1.019
matching using cosine similarity	0.0125	0.1589	0.4945	1.024	2.2497
TPT skills count	447.85835	895.1673	1777.4742	3434.2986	6737.659
TPT cosine similarity	447.8625	895.2089	1777.7245	3434.834	6738.8897

Total processing time of applying another recognition algorithm and the matching algorithm using 2 different ways first is skills count and second is cosine similarity

	250 cases	**500 cases**	**1000 cases**	**2000 cases**	**5000 cases**
Heuristic RRT	11.7977	24.46196	67.7524	147.74511	280.652
Matching using skills count	0.002122	0.01257	0.0589	0.1389	0.4256
TPT skills count	11.799822	24.47453	67.8113	147.88401	281.0776