Proceedings of IOE Graduate Conference, 2017
Volume: 5 ISSN: 2350-8914 (Online), 2350-8906 (Print)

Ontology based Job-Candidate Matching using Skill Sets

Subhash Paudel [a], Aman Shakya [b]

[a, b] *Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University*
Corresponding Email: [a] subhash.paudel@gmail.com, [b] aman.shakya@ioe.edu.np

Abstract

Adopting an efficient skills matching approach is necessary for discovering right skills needed in the labour market. Hence, skills management has been recently acknowledged as one of the key factors to adequately face the increasing competitiveness among different companies. In fact, suitable knowledge representation and matching of skills and other competences in the job and individual profiles—if properly chosen—could support human resources management automation through suitable matching and ranking services. This paper presents an approach for matchmaking between skills demand and supply through the implementation of methodology for skill profiles enrichment and matching supply and demand profiles over multiple criteria. In this respect, this work brings together research from different fields – profile modelling, information enrichment and multi-criteria matching. Methodology for harmonization and enrichment of heterogeneous profile models and skill set description by making use of standard ontology is the first contribution of this work. Secondly the formulated solution utilizes algorithm for similarity matching across multi-criteria for discovering set of profiles that best fits the job description criteria. The prototype system developed in the scope of this paper can provide a foundation for realization of a sustainable virtual marketplace for employees and employers to discover the best fitting job or resource respectively.

Keywords

HR Management – Ontology – Job Matching – Similarity Matching

1. Introduction

Due to the information overload on a given topic, the user has to filter out the proper results satisfying his information needs. In complex scenarios with multiple criteria and fuzziness of selection criteria, the information retrieval strategy based only on relational database is not sufficient. This problem is further deepened by the unstructured and heterogeneous nature of web data. This problem is strongly prevailed in the domain of human resource management because of high dependency on user profiles being created by different internet based recruitment solutions.

Nowadays human resource managers rated the internet as an important recruitment channel and over half of all personnel recruitment is the result of on-line job postings. Although job portals are an increasingly important source for job applicants and recruitment managers, they still exhibit shortcomings in retrieval and precision as the stored job offers are in syntactic formats, i.e. searches are subject to the ambiguities of natural language in job descriptions and characteristics lack relations to similar or interdependent concepts. Particularly, queries which are over-specified or inconsistent return no matches while relevant job offers could actually still be found if the consistency or specificity problem were to be resolved. Moreover, if exact matches are lacking, worse alternatives must be often accepted or the original requirements have to be negotiated for compromises [1].

No matter what the need might be, the job provider has certain requirements when it comes to the skill set, experience, education and expected salary of the people they are hiring. The goal of proper decision-making is to optimize one or more criteria in order to achieve the desired result. In human resource scenario there is a need of optimizing several criteria simultaneously. These specific problems are handled using the matching technique with the help of ontology, and similarity based strategy is used in order to compute the degree of similarity between each of the job seekers and a job description and to rank the job seekers according to their similarity score. [2]

This paper is focused on the modelling a solution for information enrichment by utilizing human resource ontology ESCO (the multilingual classification of European Skills, Competences, Qualifications and Occupations) and development of a matching approach for comparisons of applicant profiles and job openings with focus on skills, occupations, and experience as well as industry sector descriptions.

1.1 Motivation

Job recruitment often involves processing a big number of applications for an open position. It is not efficient and effective to short-list candidates manually. Therefore Web has become more and more important platform for job recruitment. There are many job portals and big organizations who set up on-line application systems, where basic data of job profiles and application profiles are entered so that the data can be used for automated selection of candidates who have satisfied the competencies and other requirement in the job profiles.

In determining relative suitability of applicants with different skill sets with regards to a specific job offer, number of questions arises as:

- How does one select people from the database with the sought after skills?
- If no exact match is found, can a selection be done where people with similar skills to the sought ones can be recommended?
- How can CV be ranked with regards to their skills, and as such, compared to each other with regards to suitability to given job offer?

In order to understand the problem let's take sample data of candidate profile and the job offer criteria containing following data:

Table 1: Example of job offer criteria

Job Title	Skill Required	Experience	Qualification
Senior web programmer in .Net	Asp.Net, C#	>5	Bachelor in computer engineering or equivalent
Ruby on Rails Developer	Ruby on Rails, MongoDB	>2	Bachelor in computer engineering or equivalent
System analyst	Architect, .Net , AWS	>2	Bachelor in computer engineering or equivalent

Table 2: Example of candidate profile.

Name	Skills	Designation	Experience
Paolo	Web Programming, C, SQL, MVC,,jQuery, Python, Java	Analyst Programmer	11
John	Architect, Technical Lead, Development Lead, .Net,Asp.Net, C#, VB, Jquery, MVC	Senior Programmer	6
Dominic	JavaScript, Ajax, Asp.Net, HTML, Oracle, MySQL, C, C#, C++, Ruby on Rails, SmallTalk, web Programming	Programmer	3

If we need a simple piece of information for satisfying the job offer having job title "Senior web programmer in .Net" as in Table 1 from candidate data in Table 2. Let us suppose that there are three candidates John, Paolo, Dominic skilled as presented in Table 2 and that the three of them have a bachelor degree fulfilling the strict constraint of the user request. Looking at the three profile descriptions and at the original request, we will rank the three candidates as (1) A (2) B (3) C w.r.t. the preference expressed by the user. In fact, reasonably, the candidate Paolo has higher ranking than Dominic because of eleven years of experience in web programming and Java even if he does not fully satisfy the criteria Asp.Net, c#. On the other side, Paolo has experience in MVC, JQuery so skills seems to be more useful than Dominic ones. To get this ranking the normal relational database query is not sufficient. It's clear from this scenario that we have to give attention to multiple criteria the candidates have in their CV with some semantic matchmaking for getting proper desired result.

2. Related Works

Some state of the art in the field of recruitment by matching the candidate skills and competencies with job description using ontological technique, similarity matching technique and logic base technique can be summarize as below.

In the paper "Improving the recruitment process through ontology-based querying" [1], the authors present the query relaxation technique which is able to return results even in the cases of inconsistent or overly specific queries which would return no results is presented . Sub-symbolic methods estimate the (quantitative) similarity between job or applicant descriptions. Symbolic approaches allow a more intuitive way to formulate and handle preferences and domain knowledge. But due to their partial preference order they cannot rank all results in practice like sub-symbolic approaches. In this paper author propose a query relaxation method which combines both methods. This method demonstrates that by having data based on formal ontologies, one can improve the retrieval.

Bird Mating Optimization method for one-to-n skill matching is proposed in “Bird Mating Optimization method for one-to-n skill matching” [2]. The method finds the optimal combination of skills from two or more CVs that best satisfies a job description. In this approach the CV sets as well as the job description are described semantically by using a skill taxonomy. To evaluate the quality of a solution (i.e. a set of CVs that satisfies the job description considered) a fitness function is defined that evaluates the degree of semantic matching of the combination of skills part of the considered solution to the set of skills of the job description.

The authors propose an ontology based method that matches a job seeker to job offers in ”Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach,” [3] . In this approach, the job seeker and the job offers are described semantically by using a skill ontology, and the type of match between a job seeker and a job offer is determined by using a description logic based classification. Additionally, a similarity based strategy is used in order to compute the degree of similarity between each of the job seekers and a job description and to rank the seekers according to their similarity score.

The author explains Logic-based techniques and technologies permit to make more efficient and flexible the recruitment process in “A system for retrieving top-k candidates to job positions” [4]. And the system they propose automatically performs a matchmaking process between available candidate profiles and vacant job positions according to mandatory requirements and preferences provided by a recruiter. In order to perform it, we need a language suitable for data intensive applications with a good compromise between expressiveness and computational complexity. The system performs non-exact match through top-k retrieval techniques: it uses a match engine which performs top-k queries over a DLR-lite Knowledge Base providing a ranked list of candidates.

The hybrid Ant Colony Optimization based method for solving the multi-skill resource-constrained project scheduling problem is proposed in “Hybrid ant colony optimization in solving multi-skill resource-constrained project scheduling problem,” [5]. In this approach, an ant is mapped to an artificial ant, every edge in a path (i.e. solution) that the ant tries to build is represented as a given task together with the resources that are capable of performing the task, the pheromone value is a value that specifies the probability of assigning a given resource to given task, the path (i.e. a solution) is a set of tasks and their associated resources, while the surface is represented by the set of all feasible solutions. Moreover, there is a special ant that leaves much more pheromone than any other ant in the colony. This ant is selected by using a Tabu-Search strategy.

The illustration of a method for matching jobs to workers, which is able to deal with incomplete and inaccurate information is done in “Expansion Methods for Job-Candidate Matching Amidst Unreliable and Sparse Data,” [6]. This approach is based on a probabilistic weighted ontology model that assigns weights to different attributes (i.e. location, skills, and qualification) and is able to perform a probabilistic conversion of audio content to text. In the case of location as attribute, the Euclidean distance is used to compute the distance between two points, while in the case of skills as attributes, a WordNet based strategy is applied to establish the distance between two skills. In the case of qualification as attribute, a lattice based approach is used. The quality of the method proposed has been evaluated by using a set of metrics from information retrieval.

3. Methodology

The methodology for development of technical solution, the first step is data collection which includes data set for both the job seekers and recruiters. The base data type for the job seekers is collection of CVs while for the recruiters is job description and necessary requirements. In order to harmonize such types of data that can be collected from different sources, the first step is cleaning, integration, selection and transformation which basically address the need to model heterogeneous data into one uniform data model. The data after pre-processing contains the personal skillsets, experience, education and expected salary data which are the factors that will be considered for further analysis. These pre-processed data of CV sets as well as the job description can be semantically linked by using domain ontology of skillset. The main purpose of this semantic linking is to find semantic distance between different terms that are used for defining skillset. So, until this step uniform representation of both the CV and job description with necessary semantic linking with skillset domain ontology will have acquired.

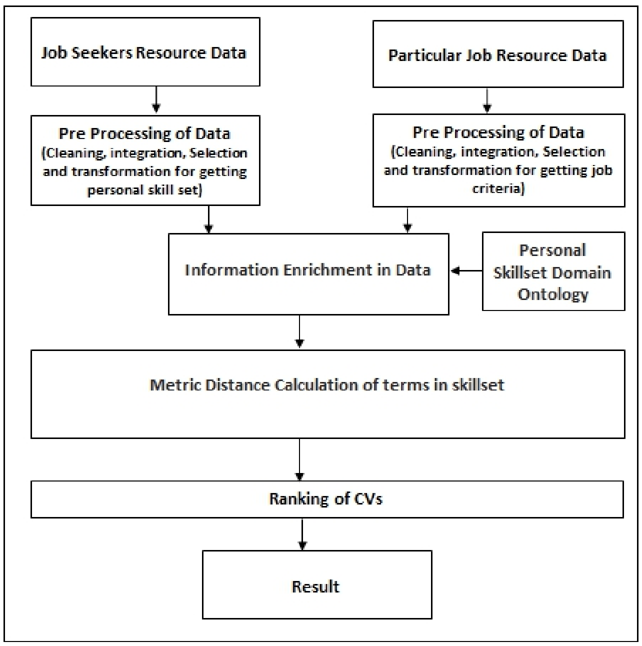


Figure 1: Model for the System

The next step is generation of suitable matching CVs in correspondence to the job description. In doing so, the first step is to filter out widely dissimilar CVs, which is done by computing semantic distance between skill sets in the CV and job description. Next, as stated in the problem definition section, the matching problem is finding best fit between the CV and job description by considering multiple criteria that define the job description. The Figure1 above shows the model of the system including these steps. Finally, the end users will have two main advantages for making decision for selecting the best candidate and or job, which are:

- The end user (Recruiter) now has a good view of the CVs that are closely related to the job description sorted by relevance.
- The end user (Job Seeker) can have the perspective into his missing skills over which (s)he can improve to fit more closely to some hot jobs in the market.

3.1 Model Development

3.1.1 Object Model

In the scenario of matching between employee and employers, it is necessary to identify the different types of actors and the respective models to describe them.The basic model for representing different entities that will be part of the match making process are:

3.1.2 Job seekers

Job seekers represent the group of people who are looking for jobs in the domain of their skills. The number of job seekers in the market is I $\in$ 1, 2, .., N. Each job seeker owns a set of non-transferable endowments: they cannot be exchanged among job seekers. The set of endowments for individual is termed as CV and is characterized by skill vector and experience. Skill vector for individual i is characterized by si = s(1,i) , .., s(K,i) , where K is the number of skills. Experience for individual i is characterized by ei=x $\in$ $\mathbb{R}$:x¡0 Skills are heterogeneously distributed among the job seekers, so the vectors is a multivariate random variable s $\sim$ Fs(s1, .., sK) and represents the set of skills available in the specified domain.

3.1.3 Employers (Companies or Firms)

This represents the companies that have various businesses in different domains. The number of firms in the given business domain is represented as F$\in$ 1, 2, .., N and have different hiring requirements which is represented as job description.

3.1.4 Job Description

This represents the way a firm opens vacancies for different positions that are available at the firm. The job description is represented as skill vector exactly as used for the job seeker.

3.1.5 Mapping to ESCO model

ESCO is the multilingual classification of European Skills, Competences, Qualifications and Occupations. For a given skillset si in Job/CV the skillset is expanded by mapping with skills defined in the ESCO ontologies by considering the similarity between the skill s(j,i) and s(m,esco) where s(j,i) is a skill in CV and s(m,esco) is the skill from the ESCO ontology. If w(j,m) is the similarity between the skills then the skill is added in the skillset si by adding the weight to the skill in the skillset vector. In the next iteration similar mapping is done with the similar skills that are related to the occupations that correspond to the skills in the skillset si.. After this mapping the skillset vector is enriched with necessary similarity weight, which is represented as: si = (s(1,i), w(1,i)), ..,(s(M,i), w(M,i)) s.t. M ¿ K i.e. the original count of skills in the skillset from the CV. This expansion of skillset will give us more detailed association with various skillsets that have been defined for different domains of works from the ESCO ontology, thus helping to build model that can lead towards better match between skillsets in demand (job descriptions) and offer (CVs)

3.1.6 Mapping of jobs and cvs

Matching between skills of individuals and job description is by a number of factors. The choice of how to manage these factors and/or different factors that are considered or discarded can lead to different solutions. Other factors are typical of the matching of skills and can be neglected in other contexts. We outline in the following some of the factors we believe characterize skill matching scenarios and then focus on our particular setting.

Negative Information treatment:

This factor affects the choice of the language in which descriptions have to be expressed and is fundamental in the matching process of any kind of description. We may itemize possibilities as follows:

<u>Absent:</u> all information allowed in profile descriptions are positive and all others are considered unknown.
<u>Implicit:</u> lacking information in a description are implicitly managed as negative.
<u>Explicit:</u> negative information can be elicited in descriptions together with positive ones, but all not elicited information are considered unknown.

Notice that considering negative information as absent or implicit in a CV can lead to having limited match results. Instead the absence of a characteristic in the description of a profile should not be interpreted as a constraint of absence but as an item that can be either refined later, or left unknown if irrelevant for a user.

Multiplicity of Relationship between Individuals and Jobs:

This issue is typical of the skill matching process, because in the matching of other kinds of good the multiplicity is always one to one. We have for example a demand describing one particular good and we search for one supply fulfilling the demand. When turning to skill matching, instead, one offered profile may be assigned more than one task and vice versa. Match relationship between Individuals and Jobs may be characterized by a multiplicity:

<u>One to one:</u> we have one job profile to match with one individual; offered and requested profile descriptions may be relative to more than one skill. The scenario is typical of temporary work agencies or counselling companies, in which one person is employed if s/he is able to attend one task.
<u>Many to one:</u> we have one job to assign to several people. This happens for example in the selection of a working team for a project, representing in this case the task to assign. For this case, each person is assigned no more than one task.
<u>One to many:</u> we search for one individual attending to many simple tasks. The scenario is similar to time-sharing in Operating Systems, in which we have one resource to share between several users. In this context many tasks share the same human resource and several constraints may ensue.
<u>Many to many:</u> we have many tasks to assign and many individuals available and we have to search for the best scheduling of human resources on the different tasks. In this work we concentrate on one-to-one skill matching and highlight some intuitive properties that a semantic approach should take into account. First of all notice that we make the open-world assumption.

Within the scope of this paper we have considered the scenario is only limited to Negative Information treatment – absent and Multiplicity of Relationship between Individuals and Jobs – one to one. The Negative Information treatment – explicit is partially handled during the skills expansion by utilizing the ESCO taxonomy as explained in the sub-section mapping to ESCO ontology.

3.1.7 Algorithms

Algorithm 1 Identifying matching skillset

STEP 1: *START*
STEP 2: *Take Skillset (array of skill) of Jobs or CVs and ontology as input.*
STEP 3: *Find the metric similarity between the skill from skillset and each skills from the ontology*
STEP 4: ***CaseI*** : *If the skill is from the ontology, make a skill object using the similarity weight, URI of the skill and name of skill from the ontology.*
case *II* : *If the skill is from the skillset, make a skill object using the similarity weight, URI of the skill from skillset and name of skill from the skillset.*
STEP 5: *Repeat from Step 3 for each skill in skillset to get the array of skill object*
STEP 6: *Choose the skill object array having similarity weight greater than 'α'*
STEP 7: *For each skill object find the corresponding occupation using ontology*
STEP 8: *For each occupation from step 7 find corresponding skill object array as in step 5 using Ontology. Multiply the similarity weight by factor 'β' for this skill object Array*
STEP 9: *Choose the skill object array having similarity weight greater than 'α'*
STEP 10: *Merge the skill object array from step 6 and step 9 then final skill object array as output*
STEP 11: *STOP*

Algorithm 2 Match CVs according to job description

STEP 1: *START*
STEP 2: *Take CVs object array (arrCVObj), Job object array (arrJobObj), Job Input, Experience Criteria (EC) as input*
STEP 3: *For Job Input find the corresponding job object (JobObj) from Job object array (arrJobObj)*
STEP 4: *Take CV object (CVObj) from CVs Object array (arrCVObj)*
if *CVObj.experience satisfy EC* **then**
    $SumofSkillWeight = \frac{FindMetricsimilarityofCVOnj.skillandJobObj.skill}{1000*(1+CVObj.experience)}$.
    SkillCompareWeight = Find Metric similarity of job.jobInput.skill and CV.CVInput.skill
    $skillWeight = SumofSkillWeight + SkillCompareWeight * \gamma + CV.expereince * \lambda$
End If;
STEP 5: *Repeat step 4 for each CVObj in arrCVObj and create the array of CVObjRank*
STEP 6: *Rank the CVObjRank array according to the skillweight and select no of ranked CVs.*
STEP 7: *STOP*

3.2 Experiments and Results

Jobs and CVs raw data for this study are taken from the online job advertisement site www.hirefire.co.uk for the period of 5 years. These data are in the form of SQL tables so they can be easily processed. The CVs taken initially on this study for the experiment are 1000 and the jobs are 100. The new cv and jobs can be added from the system manually as well.

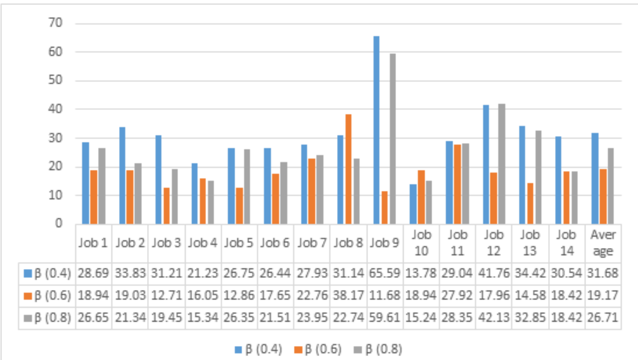
The ontology used in this paper is ESCO ontology. The ESCO ontology identifies and categorises skills, competences, qualifications and occupations relevant for the EU labour market and education and training.In this paper the focus is only on the skill, competence, qualifications and occupations of computing and information technology of ESCO ontology.

The validation data given to the individual evaluator which included the criteria-job description and list of CVs. Each CV has an identifier, skillset and experience.

For the analysis purpose difference between the score by system and evaluators average score is calculated and it is marked as error. In the output we can observe the CVs that are matching to the job description. The `Score_Actual` provides the score for each CV calculated by the system, while the `Score_Avg` provides the average score given by all the evaluators. Error gives the absolute value of difference between the score by system and evaluators average score.

The algorithms that have been developed in the scope of this work utilized a number of coefficients for skill set matches at different steps. The coefficient 'α' is threshold match score between skills in CVs or Jobs to ontology skills. The factor 'β' is used to evaluate skill weight during enrichment of skills from the second layer ontology. The factor 'γ' is used to evaluate skill which matches the skills of CV and Jobs without any enrichment. The factor 'λ' that is used to evaluate the impact of experience.

The value of 'α' is dependent on the number of skills considered for the further processing. The higher the value of 'α' lower the number of skills to be selected and the purpose is to filter out highly unrelated skills. From the experimentation value of 'α' chosen is 0.3. The value 'γ' gives the importance to the skills that are usually written in unstructured Job and CV skills. In order to use data coming from unstructured sources lower value for this coefficient give the good reuslt and for the data obtained from standard CVs and Jobs description higher value of 'γ' give good results. For this work the value chosen is to be 2. The value of 'λ' can be freely chosen based on the importance of experience. This value is 0.1 on the basis of different experiment on the data.

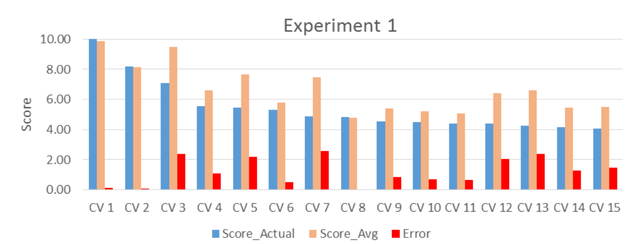


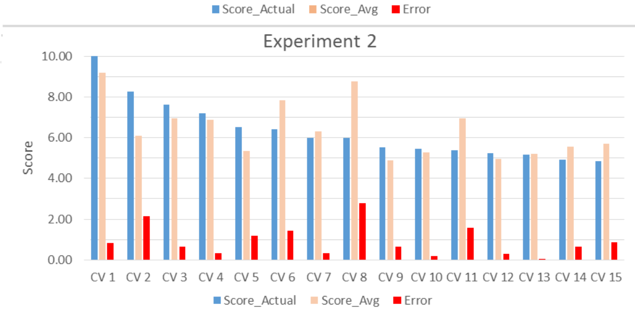
	Job 1	Job 2	Job 3	Job 4	Job 5	Job 6	Job 7	Job 8	Job 9	Job 10	Job 11	Job 12	Job 13	Job 14	Aver age
β (0.4)	28.69	33.83	31.21	21.23	26.75	26.44	27.93	31.14	65.59	13.78	29.04	41.76	34.42	30.54	31.68
β (0.6)	18.94	19.03	12.71	16.05	12.86	17.65	22.76	38.17	11.68	18.94	27.92	17.96	14.58	18.42	19.17
β (0.8)	26.65	21.34	19.45	15.34	26.35	21.51	23.95	22.74	59.61	15.24	28.35	42.13	32.85	18.42	26.71

Figure 2: Error comparison for different values of 'β'

From the experiment on changing the value of 'β' it is observe that the data are more close to the expert score for 'β' =0.6. So for this work the value chosen is 0.6.

The following figure gives the comparison of system and evaluators score.





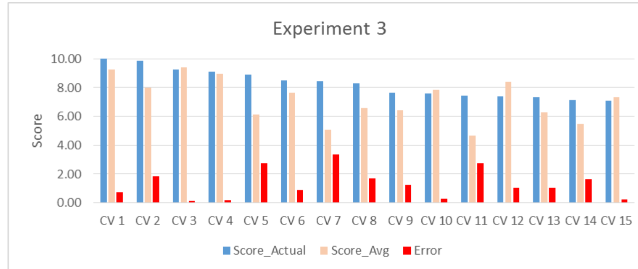


Figure 3: The comparison of system and evaluator score

We observed some errors in the scoring by the system and evaluator which is approximately 18.94%, 12.71%, 16.05% for three experiments respectively shown in figure above. But the CVs which have been ranked highest have smaller error to the CVs which have been ranked lowest. Also the profiles having different skillset described in the job also profited and ranked higher because of their skillset association with the job description.

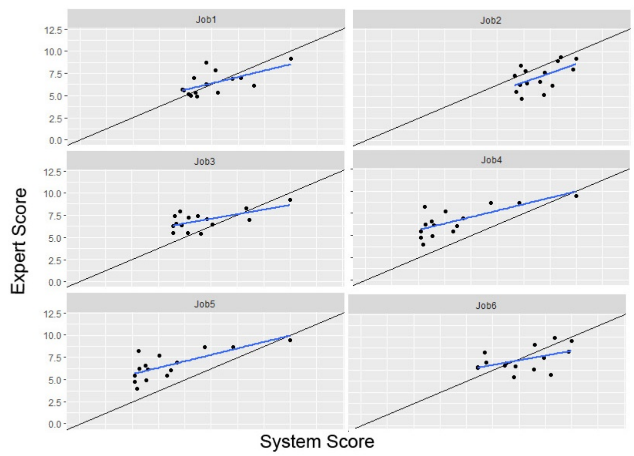


Figure 4: The expert score verses system score scatter graph for each job

The black line represent the median line for both the axes (x-axis- System Score, y-axis-Evaluator Score). The blue line represent the regression line. From the graph it is obvious that the plot for all job description is evenly distributed.

The following graph shows the standard error plot for both system and evaluator score. The results in the system shows the even distribution of deviation of score from system and evaluator score across all job description which shows the system is consistent. If we consider the other criteria which are not included in this thesis work the system can perform efficiently. From the observation of the result of the system, the ranked list deals efficiently with the cases where the exact match of given requirement does not met by the candidate profile.

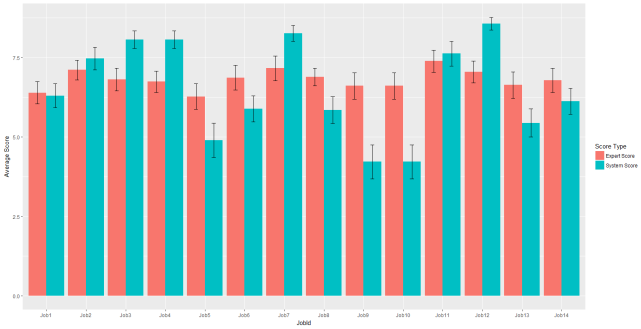


Figure 5: The graph for the standard error

From the observation the most important point to note is the even distribution of error across all job description. This implies that the performance for matching by the system is consistent. Even though the average error percent is 19.12% we can consistently improve the system performance by tuning the various factors in matching algorithm.

4. Conclusions and Future Works

In this paper an approach for matchmaking between skills demand and supply through the implementation of methodology for skill profiles enrichment and matching supply and demand profiles over multiple criteria is carried out. This work brings together research from different fields – profile modelling, information enrichment and multi-criteria matching. The Methodology for enrichment of heterogeneous profile models and skill set description of the candidate and jobs by making use of standard ontologies is the first contribution of this thesis work. Secondly the formulated solution utilizes algorithm for similarity matching across multi-criteria for discovering set of profiles that best fits the job description criteria. This thesis work mainly focus on modelling the existing data in the field of job recruitment with standard ontology and similarity matching algorithm.

The result obtained from the system was evaluated by comparing the results collected from experts for same inputs. The observation during the validation have an evenly distributed deviation. This deviation can be corrected by tuning different coefficient defined in the algorithm. On the whole this research work can provide a foundation for realization of a sustainable virtual marketplace for employees and employers to discover the best fitting job or resource respectively.

Some recommendation for future work are:

- Automatically handle the integration of changed in ESCO ontology which is improving on continuous basis. One of the recommendation is to improve the system by making dynamic enough to automatically link with ESCO ontology.
- The matching algorithm can be further improved by including other criteria such as degree, previous projects, expectations, designation, soft skills, location etc.
- And finally, some further work on experimental analysis with more criteria can be important future work for fine tuning the algorithm. Additionally we recommend to extend the system with learning algorithm to improve matches over time.

References

[1] Malgorzata Mochól, Lyndon J. B. Nixon, and Holger Wache. Improving the recruitment process through ontology-based querying. In *SEBIZ*, volume 226 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.

[2] Simona Corde, Viorica Rozina Chifu, Ioan Salomie, Emil St Chifu, and Andreea Iepure. Bird Mating Optimization method for one-to-N skill matching. *Proceedings - 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing, ICCP 2016*, pages 155–162, 2016.

[3] Maryam Fazel-Zarandi and Mark S. Fox. Semantic matchmaking for job recruitment: An ontology-based hybrid approach. 2009.

[4] Umberto Straccia, Eufemia Tinelli, Simona Colucci, Tommaso Di Noia, and Eugenio Di Sciascio. A system for retrieving top-k candidates to job positions. *CEUR Workshop Proceedings*, 477, 2009.

[5] Paweł B. Myszkowski, Marek E. Skowroński, Łukasz P. Olech, and Krzysztof Oślizło. Hybrid ant colony optimization in solving multi-skill resource-constrained project scheduling problem. *Soft Computing*, 19(12):3599–3619, 2015.

[6] Jerome White, Krishna Kummamuru, and Nitendra Rajput. Expansion Methods for Job-Candidate Matching Amidst Unreliable and Sparse Data. *COLING'12 - Proceedings of the 24th International Conference on Computational Linguistics*, (December):1321–1330, 2012.