An Ontology-Based Approach for improving Job Search in Online Job Portals

Article · June 2018

CITATIONS
0

READS
434

3 authors, including:

Oluwatobi S. Dada
Federal University, Kashere
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

A.F. Donfack Kana
Ahmadu Bello University
16 PUBLICATIONS **21** CITATIONS
SEE PROFILE

THE JOURNAL OF COMPUTER SCIENCE AND ITS APPLICATIONS
Vol. 25, No 1, June, 2018

AN ONTOLOGY BASED APPROACH FOR IMPROVING JOB SEARCH IN ONLINE JOB PORTALS

O. S. Dada[1], A. F. D. Kana[2], S. E. Abdullahi[3]

[1,2,3]Ahmadu Bello University, Zaria

[1]saintdada2000@gmail.com; [2] donfackkana@gmail.com

ABSTRACT
Internet has become the primary medium for Human Resource Management, specifically job recruitment and employment process. Classical job recruitment portals on the Internet rely on the keyword based search technique in plain text to locate jobs. However, this technique results in high recall, low precision and without considering the semantic similarity between these keywords. Many researchers have proposed semantic matching approaches by developing ontologies as a reference to determine matching accuracy qualitatively, however these approaches do not quantify how closely matched applicants and employers are, based on core skills. This paper proposes a technique that uses an ontology based approach to enhance keyword searching by leveraging on the similarity between concepts in the ontology, which represent core skills needed and required for a job in order to determine how closely matched an applicant is to a job advertisement and vice-versa. This was achieved by developing a CV Ontology based on core skills, annotating applicant profiles and job profiles using a common vocabulary and modifying the semantic concept similarity algorithm to accurately compute and rank matching score between profiles when a query is performed. The results showed improvements of 54% and 36% for Recall and F-measure respectively, over [21].

Keywords: Ontology, Semantic, Algorithm, Core Skills, OWL.

1.0 INTRODUCTION

The Internet has become the primary medium for recruitment and employment processes. Jobberman's Online Recruitment Service Report (2015) claims that applications on its job recruitment portal increased by over 50% between May and September 2015. This clearly indicates an upward trajectory in online job portals being a major player in contemporary job recruitment process. The relevance of the Internet in job recruitment process cannot be overemphasized, more than three-quarter of the age class qualified for recruitment are active internet users and there is an increasing number of companies that publish their job vacancies on the web [17]. Most classical search engines and search mechanism adopted by online job recruitment portals rely heavily on containment

of keywords in free text before search results are returned. This may produce a lot of result from a submitted keyword or phrase but many of these results may be irrelevant to user's need. Therefore, a user may have to navigate through a large number of results to find a domain specific results. Many researchers have proposed several semantic matching approaches and have developed prototype job portals to effectively match job seekers with corresponding job postings [15]. They achieved this by developing human resource or Curriculum vitae (CV) ontologies using controlled vocabularies to determine how applicants are closely related to job positions advertised. however, how to quantitatively and precisely match job seekers with available job postings based on semantic similarity between their core skills and competences relative to the core skills and competences required for the advertised jobs and also ranking the search according to the semantic closeness between applicant profile and job profile relative to their respective core skills set was not provided. Lack of such matching may lead to imprecision and lack of overall effectiveness in matching between available jobs and qualified candidates. This paper therefore seeks to address the above mentioned problems by developing a CV Ontology based on core skills, annotating applicant profiles and job profiles using a common vocabulary and modifying and implementing the semantic concept similarity matching algorithm to accurately compute and rank matching score between profiles when a query is performed.

Questo documento cerca quindi di affrontare i problemi sopra menzionati sviluppando un'ontologia CV basata sulle competenze di base, annotando i profili dei candidati e dei profili professionali utilizzando un vocabolario comune e modificando e implementando l'algoritmo di corrispondenza della similarità dei concetti semantici per calcolare e classificare accuratamente il punteggio di corrispondenza tra i profili quando viene eseguita una query

2.0 Related work

Several works have been carried out in recent times in order to improve the quality of online job recruitment using ontologies in particular and semantic web technologies in general. A system that focuses on the semantic modeling of online recruitment documents was proposed in [8], the author developed an ontology for the database field. The proposed ontology is inspired from the common parts, which were considered significant to CVs and job offers in the field of databases. Furthermore they bring clarifications regarding the essential concepts for the semantic modeling of online recruitment systems, field ontology, semantic annotation, semantic indexing, and semantic association of documents.

[10] Proposed a framework for building intelligent interoperable applications for employment system. This was achieved by collaborating between distributed heterogeneous data models using semantic web technologies. The objective of their work is to provide a better inference system for the query against dynamic collection of information in collaborating data model. Their employment exchange system provides interface for the users to register their details thereby managing the knowledge base dynamically. Semantic server transforms the queries from the employer and jobseeker semantically for possible integration of the two heterogeneous data models to drive intelligent inference. The semantic agent reconciles the syntax and semantic conflicts which exists among the contributing ontologies in different granularity levels and performs automatic integration of two source ontologies and gives response to the user.

[11] Proposed an approach to job matching with user provided information, referred to as parameters. Common parameters for job matching includes domain of job, job title, position, knowledge, experience, location, salary etc. Predefined rules assign weighting factor to each parameter and defining how matching results could be filtered and ranked to produce job matching results. The used an auto-filling technique in places where a candidate has missed out certain important information in their resume. The auto-filing utilized the self-learning engine to collect information, analyses the data and auto generates standard template for different categories group. The standard template is categorized based on some parameters such as qualification, education background and job experience. Their proposed self-learning engine

uses the advantage of ontology to make inference from data in order to discover missing parameters as well as new relationship among the parameters. The inference techniques also improve the possible inconsistencies of various parameters. The system then performs a final job matching based on direct parameters extracted from user input and dynamically populated parameters from matched standard template.

[22] Proposed a qualitative assessment of resumes on the basis of different quality parameters using a simple text analytic based approach for a resume collection. The resume collection is assessed for two qualitative aspects, coverage and comprehensibility; and these ratings are transformed into a comprehensive quality rating. All the quality parameters are collectively measured into a combined 1 to 5 rating scale for determining the quality metric for the resumes. While for coverage, it is simpler; but in case of comprehensibility, it is a bit complex and tricky to transform computed values to 1 to 5 scale rating. Nevertheless, the algorithmic formulation was used in an annotation and recommendation system.

[21] Proposed a technique that uses ontologies to implement a query augmentation that improves defining the context through users adding suggestions of relevant keywords which represent core skills needed for a particular job or task. The intuition is that by searching based on core skills, the context narrows, making it easier to search for any consultant matching a specific assignment or a job seeker searching for a particular job. In his work he claims that the Human Resource Management Ontology was better suited for storing information rather than applying relations between the information. The author created an ontology based on core skills which gives more freedom when querying with SPARQL.

3.0 Proposed Approach

In This section we discuss the system architecture and the modified semantic concept similarity Matching Algorithm

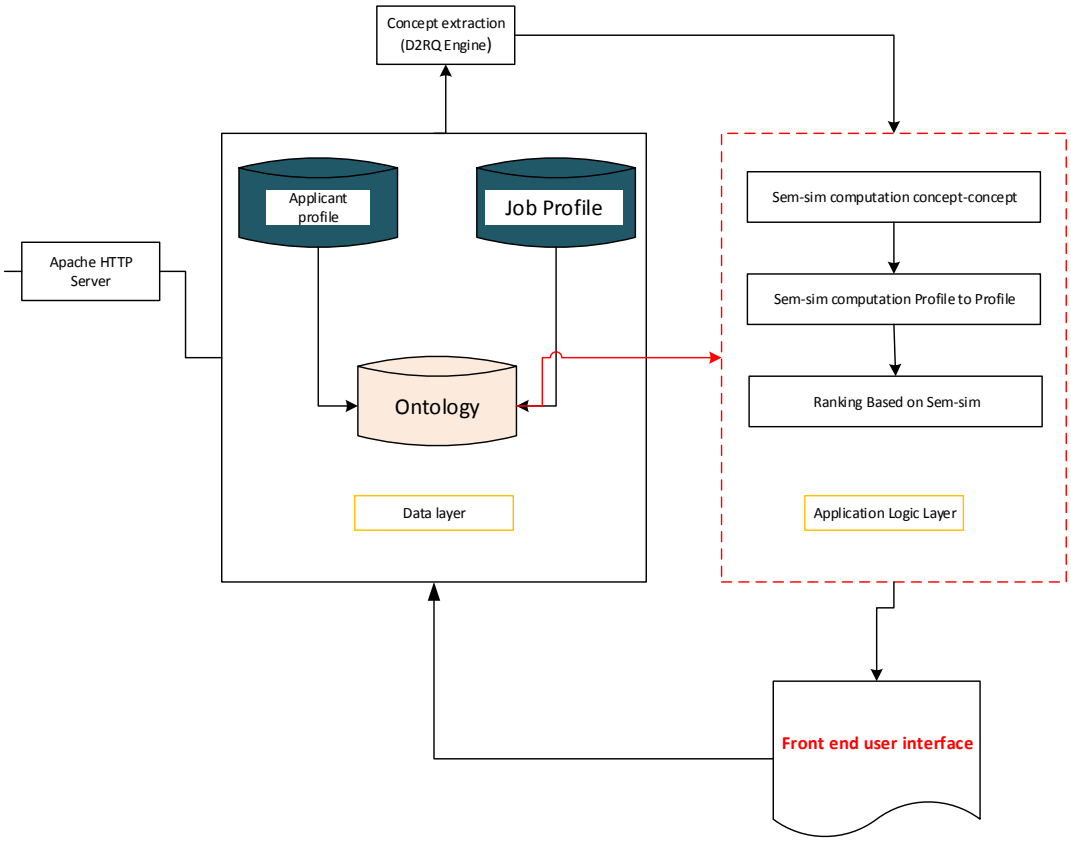

Figure 1: System Architecture

3.1 System Architecture

The system architecture illustrates the different levels of functionality of the proposed ontology based job recruitment portal. The system comprises of three major components or layers.

3.1.2 Data Layer

At this layer, the job applicant's full profile and the Employer profiles with job adverts originally stored in a MySQL database are converted to RDF format using the D2RQ platform. The D2RQ engine, is a system for accessing relational databases as RDF graphs and also allows data stored in RDF to be queried using SPARQL. When a query is performed, core skills are extracted from the applicant and job profiles. Core skills represents concepts on the ontology. These core skills are accessed by the application logic engine from D2RQ for semantic similarity computation

3.1.3 Application Logic Layer

At this layer, classes which represent core skills required in the programming domain initially implemented in an OWL file, are accessed in Java using the OWL2Java framework. The Modified Semantic similarity matching Algorithm is deployed to the ontology where the weight values are assigned to the edges between classes. Furthermore the semantic matching of job profiles and applicant profiles is performed by computing the semantic similarity of the core skills and further ranking based on semantic similarity in relation to the ontology. .

3.1.4 User interface/Front end layer

On top of the architecture is the front end layer which offers a browser-based user interface, which accepts input from users by forms: These forms are used for login and entering details on user profiles. Ranked query results are also displayed to the users.

3.2 The Modified Algorithm

The modified concept similarity matching algorithm based on semantic distance adopts the approach used in [9] where four macro steps are considered in computing semantic similarity between two concepts. However, in this work the semantic similarity is not between only two concepts, but rather a job profile and an applicant profile which both contain multiple skills. The algorithm therefore computes the semantic similarity between all the skills in a job profile with all the skills in an applicant profile in pairs and obtains a total. For a particular applicant profile, it repeats this against all the available job profiles and vice versa on the portal. It further sorts and returns in ascending order all the jobs for which a particular candidate is qualified for or all the available candidates qualified for a particular job which represents a ranking mechanism based on semantic similarity.

The algorithm allocates the weight value to the edge between concept nodes using the weight allocation function defined in [9]. Given two concepts c_1 *and* c_2, the weight allocation function is given as.

$$\mathbf{W[sub}(c_1, c_2)\mathbf{]} = 1 + \frac{1}{k^{depth(c_2)}} \qquad (1)$$

Equation 1 is the weight allocation function for Parent-child node and sibling nodes. Where, depth(C) is the depth of concept C from the root concept to node., K is a predefined factor larger than 1 indicating the rate at which the weight values decrease along the ontology hierarchy. The equation has two fundamental properties: (a) The semantic differences between upper level concepts are higher than those between lower level concepts, in other words, two general concepts are less similar than two specialized ones. (b) The distance between sibling concepts is greater than the distance between parent and child concepts. Specially, the depth of root is zero and the depth of other concepts is equal to their path length to root concept node.

ALGORITHM: Modified Concept Similarity Matching Algorithm for Profile Matching and ranking

Algorithm Searching (P, $Z\{\ z_0 . . . z_n\ \}$)
// P is the profile to search
// Z is the set of profiles being searched
// f [0 . . . n] is an array of skills of P
// k [i] is an array of skills of z
// n_0 is the root node
// v is the total number of concepts on the ontology
// H [][] is an array which hold weights between node edges
1 H [][] ← Weights
2 For each $z\ in\ Z$
3 k [ ] ← core skills of z
4 Sum ← 0
5 For i = 0 to k.length - 1
6 For j = 0 to f.length - 1
7 If k[i] = f[j] then
8 Sem_Dis (k[i], f[j]) = 0
9 Else if there exists the direct path relation between k[i] and f[j]
10 Sem_Dis (k[i], f[j]) = W_d [sub (H, k[i], f[j])]
11 Else if there exists the indirect path relation between k[i] and f[j]
12 Sem_Dis (k[i], f[j]) = $\sum W_i$ [sub (H, k[i], f[j])
13 Else
14 Sem_Dis (k[i], f[j]) = min {Sem_Dis($k[i], n_0$)} + min {Sem_Dis(f[j], n_0)}
15 End for loop
16 Sum+= Sem_Dis
17 End for loop
18 Q[]. Profile = Z
19 Q[]. Sem_Dis = Sum
// sort Q based on the value
20 End for each
21 Q[].Sort (Sem_Dis)

3.3 Designing the Ontology

The first step requires that the domain and scope of the ontology to be designed must be determined. For this ontology, the domain is the computer programming languages and the scope is to model the basic information as regards core skills/concepts that are associated with Job positions in the aforementioned

domain. Two super classes were created for the ontology in this work. Firstly is the "Person class" which comprises of two subclasses which are the employer and the applicant class. The second is the "Programming language class" which comprises of two major subclasses which are, Hardware development based programming languages subclass and Software development based programming language subclass. Hardware development based programming languages subclass: This class consists of programming languages often required by employers who need programmers for the purpose of hardware development. These programs are commonly used in coding electronic circuits and digital logic circuits. Some of the sub classes used under this category include Assembly, Lisp, Matlab, Pascal and also Hardware Development Language (HDL). HDL class comprises of two sub classes which are Analog and Digital. Software development based programming languages subclass: This class was populated with major programming languages required in the software development circles relative to the needs of the employer. They include C++, C#, Delphi, Java, Ruby, Perl, Objective C, Python, MacRuby, Swift and Visual Basic. The Software Development class had one major sub-class which is Web development based programming which is further broken down into two other major sub-classes these are Classical web programming languages and Semantic web programming languages. Classical web based programming languages are programming languages required by employers for the development of classical web systems and applications, This was further subdivided into two main classes often times as required by employers, which are Front-end Path and Full-Stack Path. These two subclasses were populated with sub classes as shown in the figures 2 and 3. Semantic Web based programming languages includes Languages often required by employers for the development of semantic web based systems and applications basically OWL RDF and XML.

It is important to note that the classes and sub-classes created in this ontology which represents core skills and concepts required for employment in the programming languages domain are not exhaustive but good enough for our proposed modified concept similarity matching algorithm to operate upon. The ontology was developed in Protégé.

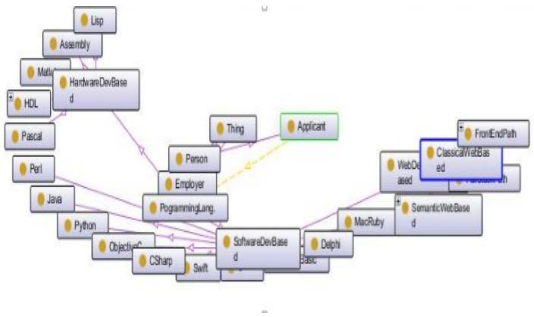

Figure 2: View of CV ontology on Ontograph

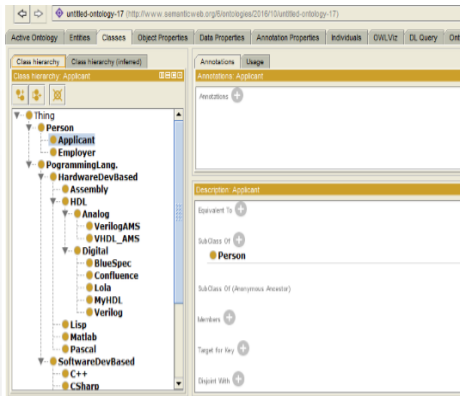

Figure 3: View of CV ontology on Protégé

4.0 Results and Discussion

The following performance metrics were used to determine the performance of our system:

i. Precision
ii. Recall.
iii. F-Measure

Comparison with work by [21] using measure (i), (ii) & (iii) based on the number of web pages to be used for evaluation.

Precision, also known as Positive Predictive Value (PPV), primarily it is used to show how useful search results are. Precision and recall are units of measurement that show an indication of how well a system performs when querying for relevant documents. Given a set of documents divided into relevant and non -relevant documents, Precision measures how many of the documents retrieved by a search are relevant [21]. It's given by Eq. (2)

$$\frac{relevant\ documents\ \cap\ retrieved\ documents}{retrieved\ documents} \quad (2)$$

Recall, sometimes also known as Sensitivity or True Positive Rate (TPR), is similar to precision but looks at how many of the relevant documents are actually retrieved[21].It's given by Eq. (3)

$$\frac{relevant\ documents\ \cap\ retrieved\ documents}{relevant\ documents} \quad (3)$$

F-measure is a combination of both precision and recall, and is a harmonic mean of both the measurements. It tries to give a better measurement of "effectiveness" than recall and precision alone are able to accomplish
F-measure, sometimes also called F-score, has multiple definitions but its main one is defined as follows [21]. It's given by Eq. (4)

$$\textbf{F-measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

4.1 Data Set

A Total of two hundred and forty (160) applicant and employer profiles were created. There were eighty (80) profiles for two core skills, three, and four respectively, all relative to the ontology. A total of sixty (60) profiles were used to test the Precision, Recall and F-measure of the search. The same set of data set were also used to test the performance of [21].

Table 1: Results Obtained for Proposed Portal

No of Skills	Precision	Recall	F-Measure
Two skills	0.91	1	0.95
Three Skills	0.90	1	0.94
Four skills	0.90	1	0.94
Average	**0.90**	**1**	**0.94**

Table 2: Results Obtained (Tran, 2016)

No of Skills	Precision	Recall	F-Measure
Two skills	1	0.22	0.36
Three Skills	1	0.56	0.71
Four skills	1	0.60	0.74
Average	**1**	**0.46**	**0.60**

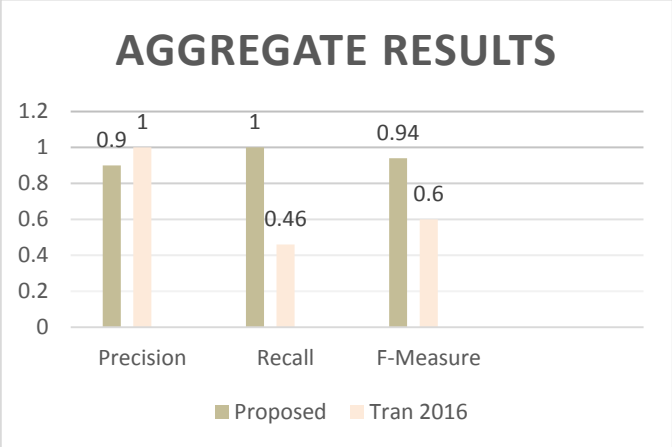

Figure 4**:** The aggregates of the results obtained in this work as compared to Tran (2016) shows that an improvement of 54% and 36% were obtained over Tran (2016) for Recall and F-Measure Respectively

Some of the observations made from the results obtained are summarized as follows:
i) Our experiments indicate that the search and matching is performed to specification since it was able to get a ranked list of all jobs available on the portal relative to the ontology according to the core skills of the applicant or employer.

ii) From the results obtained from this work there were relative improvements of 54% and 36% over Tran (2016) for Recall and F-Measure respectively these improvements were obtained as a result of the implementation of the concept similarity matching algorithm deployed on the ontology as against just leveraging SPARQL query used in [21]. However in the work of [21] a 100% Precision is recorded as against an average of 90% in this work. This is because SPARQL query retrieves only profiles that exactly match the query, thereby making all retrieved profiles relevant to the search.

iii) Query relaxation was achieved as evident in the recall rate, which is 100%. This indicates that all candidate profile on the portal are compared against all the jobs and are indexed in the search result. This is contrary to [21] where only best fit candidates are returned.

5.0 Conclusion

In this research work, a modified concept similarity matching algorithm was deployed to a CV ontology in order to match core skills of applicants to employers more accurately. From the results of experiments carried out on the implemented system, the searching and matching technique recorded an improvement in accuracy of matching. Results affirm the relevance of the concept similarity matching algorithm in determining ontolological closeness or similarities between concepts on an ontology and how it can be applied to improve the employment process.

REFERENCES

[1] Abrouk, L. (2006). Annotation de documents par lecontexte de citation basée sur une ontologie. M.Sc. Thesis, Universite Montpelier II, France. Available from http://tel.archives-ouvertes.fr/docs/00/14/25/68/PDF/these.pdf Retrieved May 17, 2016

[2] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Available from http://www.scientificamerican.com Retrieved March 14, 2016

[3] Bizer, C., Heese, R., Mochol, M., Oldawaski, R., Tolksdorf, R., & Ecstein, R. (2005). The Impact of semantic web technologies on job recruitment process. In: proceedings of International Conference Wirtschaftsinformatik (WI), Bamberg, Germany.

[4] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F. M., Mongiello, M., Mottola, M. A. (2003). Formal Approach to Ontology-Based Semantic Match of Skills Descriptions. Journal of Universal Computer Science., 9(12), 1437-1454.

[5] Dafoulas, G. A., Nikolaou, A. N., Turega, M., & (2003). E-Services in the Internet Job Market. In: proceedings of 36th Hawaii International Conference on System Sciences, Hawaii, USA.

[6] Fazel, Z., & Fox, M. (2009). Semantic Matchmaking for Job Recruitment: An Ontology Based Hybrid Approach. In: proceedings of the 3rd International Workshop on Service Matchmaking and Resource Retrieval in the Semantic Web at the 8th International Semantic Web Conference (ISWC), Washington D.C., USA.

[7] Guissé, A., Lévy, F., Nazarenko, A., & Szulman, S. R. (2009). Annotation sémantique pour l'indexation de règles métiers. Available from

http://www.irit.fr/TIA09/thekey/articles/guisse-levy-nazarenko-szulman.pdf Retrieved March 16, 2016

[8] Ionecusu, B., Ionecusu, I., Florescu, V., & Tinca, A. (2012). Semantic Annotation and Association of Web Documents: A Proposal for Semantic Modelling in the Context of E-recruitment in the IT Field. Journal of accounting and management information system, 11(1), 76-96.

[9] Jike, G., & Yuhui, Q. (2008). Concept Similarity Matching Based on Semantic Distance. In: proceedings of Fourth International Conference on Semantics, Knowledge and Grid. IEEE computer society, Beijing, China.

[10] Kavidha, A., & Saradha, A. (2013). Intelligent interoperable application for employment Exchange using ontology. Webology, 10(2).

[11] May, F., & Yew, C. (2015). Intelligent job matching with self-learning recommendation engine. In: proceedings of 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences,

[12] Mochol, M., Jentzsch, A., & Wache, H. (2007). Suitable employees wanted Find them with semantic techniques. In: proceedings of Workshop on Making Semantics Web For Business at European Semantic Technology Conference (ESTC2007), Vienna, Austria.

[13] Mochol, M., & Paslaru, B. (2006). Practical Guidelines for Building Semantic eRecruitment Applications. . In: proceedings of International Conference on Knowledge Management, Special Track: Advanced Semantic Technologies, Berlin, Germany.

[14] Mulder, W. (2010). Personal informations system e-Entwicklungsstand, Funktionalitat und Trends. Wirtschafts informatik. Journal of Universal Computer Science., 16(2), 98–10.

[15] Niaphruek, P. (2012). Job Recruitment System Using Semantic Web Technology. In: proceedings of 15th International conference of international academy of physical science Pathumthani, Thailand.

[16] Noy, N. F., & McGuinness, D. L. (2016). Ontology Development 101: A Guide to Creating Your First Ontology. Available from URL: http : / / protege . stanford .edu/publications/ontology%7B%5C_%7Ddevelopment/ontology101-noy-mcguinness.htm Retrieved May 5, 2016

[17] Report., J. s. O. R. S. (2015). Available from http://www.jobberman.com/INFOGRAPHIC-Online-Recruitment-Services-Report-2015.htm Retrieved April 21, 2016

[18] Schmidt, A. (2005). Bridging the Gap Between E-Learning and Knowledge Management with Context-Aware Corporate Learning Solutions. Professional Knowledge Management. . In: proceedings of Third Biennial Conference WM, Revised Selected Papers, Lecture Notes in Artificial Intelligence (LNAI).

[19] Schmidt, A. (2006). Context-Aware Workplace Learning Support: Concept, Experiences, and Remaining Challenges In: proceedings of European Conference on Technology-Enhanced Learning (EC-TEL 06).

[20] Sheth, A., & Thirunarayan, K. (2013). Multimedia Semantics empowered Web 3.0 Available from https://books.google.com.ng/books?id=IK4v4fluDJgC&pg=RA2PA7&lpg=RAA7 Retrieved April 20, 2016

[21] Tran, H. (2016). Human Resource Matching Through Query Augmentation for Improving Search Context. (M.Sc. thesis), Linkopings University Sweeden. Available from https://liu.diva-portal.org/smash/get/diva2:941053/FULLTEXT01.pdf

[22] Vinaya, R., Munjula, R., & Sidna, N. (2015). An unstructured text analytics approach for qualitative evaluation of resumes. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 8(2), 2349-2163.

[23] Yahiaoui, L., Boufaida, Z., & Prié, Y. (2006). Semantic Annotation of Documents Applied to e-recruitment. Available from https://www.semanticscholar.org/paper/Semantic-Annotation-of-Documents-Applied-to- E-Yahiaoui-Boufa/pdf Retrieved April 14, 2016