SEMANTIC WEB APPROACHES IN E-RECRUITMENT: A SYSTEMATIC LITERATURE REVIEW

Mihaela-Irina ENACHESCU
Bucharest University of Economic Studies, Bucharest, Romania
irina.enachescu@csie.ase.ro

Abstract. *This paper aims to provide an overview about the efforts done in the last decade to automate the recruitment process carried out using the Internet, taking advantage of the benefits brought by the Semantic Web. A systematic literature review was conducted in order to highlight the common approaches used when trying to develop a semantic job matchmaking system, focusing mainly on the design of the human resource ontology. As the simple use of an ontology followed by the annotation of jobs and resumes has its shortcomings we concentrate on describing some of the proven strategies that were employed to improve the semantic matchmaking process, like the use of semantic clusters, query relaxation, under-qualified match and exploiting multiple semantic resources. In order to prove that the use of Semantic Web technologies in eRecruitment does achieve better results in the matching process, this paper gave an overview about the common methods applied to validate the proposed online recruitment systems.*

Keywords: eRecruitment, Human Resource Ontology, Semantic Matching, Semantic Web
JEL classification: J21, M51, O35

1. Introduction

Nowadays, the Internet has become the main stage for performing the recruitment process. Although there are companies that still use offline medium to carry out the management of human resources, big companies and national/international authorities have focused their attention on automating the process and seeking for candidates in online media. They search for platforms that enable them to fill vacant positions faster and cost-effectively and also deliver to the users a richer and more effective user experience.

Using the Internet for recruitment improves the quality of jobs and applicants' profiles matching, reduces search costs, increases contact opportunities for both of the involved parties and removes the subjectivism from the screening process of job applicants [1].

Search engines, like Google, Bing etc., are becoming more popular but they are miles away from resolving the requirements of the eRecruiting. One problem of the general search engines is that recruiters and job seekers find them difficult to use, due to their overwhelming syntax. Also, when querying to obtain the desired information, because of the ambiguous meaning of the words chosen, traditional search engines fail to deliver the expected results. That is the dilemma that semantic technology aims to address by labelling data with information related to its meaning and context.

Semantic Web is providing "a basis for coding, exchanging, and reusing structured metadata among applications exchanging machine understandable information on the Web" [2]. [3] claims that extending job portals with semantics "overcome the problems of distribution, heterogeneity and machine non-processability of existing job information and substantially increase market transparency, lower transaction costs and speed up the procurement process for businesses".

This research aims to provide an overview towards the efforts done in the last ten years in order to create and use semantic job matchmaking systems in the recruitment processes. First, we tried to gather information about the most valuable researches done in this particular area. After building a list with them, we did an analysis in terms of the main steps considered when such systems were proposed. We spotted the critical phase of the process and took an in depth look at it, trying to introduce several approaches and highlighting the advantages and disadvantages of each of them. Further, we paid special attention to the most used and proven strategies that were applied to gain better results when using Semantic Web in eRecruitment. Finally, we focused on identifying the methods used to review the job matchmaking prototypes, so we can validate if they accomplish what was initially proposed and how accurate their results are, compared with existing solutions.
The rest of the paper is organized as follows. The next section presents the main steps considered when building a semantic job matchmaking system, with more focus on the ontology development. In section 3 we describe the strategies used to improve the semantic matchmaking process. Section 4 presents the common used approaches to validate the prototypes of the eRecruitment systems built with Semantic Web technologies. Finally, section 5 presents the conclusions to our literature review and suggest some future work remarks, which can be used to improve existing solutions.

2. Semantic Web Approaches in eRecruitment

Developing a semantic job matchmaking system can follow many available approaches. However, no matter which is the chosen approach three main steps are always considered [4]:
a. Creating a formal data shared model (ontology) that provides the standard to define all the relevant information for the job seeker or the job recruiter. This can be done: from scratch or by merging already existing standards/sub-ontologies.
b. Annotating the resumes/job descriptions with meta-information extracted from the ontology model build at the first step. This process can be done: automatically, semi-automatically or manually.
c. Computing the match between the two entities involved (candidate's profile and job advertisement). In this phase several algorithms are used to measure the semantic distance among concepts and calculate the overall similarity.
The main phase when trying to develop a job recruitment portal that uses Semantic Web technologies consists in designing a *human resource ontology* that will allow a uniform representation of job ads and job seekers' profiles. The robustness and completeness of the modelled ontology will have significant impact on the success of the implemented system.
For implementing the ontology researchers rely on empirical evidence ([3], [5], [6], [7], [8]) or apply already known and validated methods ([9]). In [7], M. Mochol et al. define a list of guidelines for the ontology engineering process, based on practical lessons learned in the Knowledge Nets project. The authors suggest three steps for developing an ontology based on existing standards: ontology discovery (a considerable effort need to be invested to identify existing ontologies, classifications, taxonomies, controlled vocabularies, glossaries etc.), ontology evaluation (in terms of usability and provenience – if the source is trusted and with good reputation) and ontology integration and merging (linguistic and taxonomic matchers algorithms can be used, but the final results need human experts evaluation). On the other hand, [9] applies for the design of the ontology the method developed by the University of Stanford, including the following seven steps: determining the domain and scope of the ontology, considering reusing existing ontologies, identifying the important terms of the ontology, defining the classes and the hierarchy of classes, defining the properties of the classes (the attributes), defining the facets of the attributes and creating instances of the classes.

Analyzing the approaches used when creating an ontology for eRecruitment, two alternatives stand out: using already defined standards and classifications (this is most used, [3], [5], [6], [7], [10], [11]) or designing it from scratch ([8], [9], [12], [13], [14], [15]).
Some of the commonly known standards and classifications used in eRecruitment are: HR-BA-XML (XML specifications to enable ebusiness and the automation of human resources related data exchanges), SOC (Standard Occupational Classification), BKZ (German version of SOC), WZ2003 (German standard classification of economic activities), NAICS (North American Industry Classification System), KOWIEN (skill ontology), ISCED 1997 (International Standard Classification of Education), ISCO 88 (International Standard classification of occupations), NACE (the European standard for industry classifications), CEF (Common European Framework of Reference for Languages), ISO 3166 (a standard published by the International Organization for Standardization, used for representing nationality of an applicant) etc.
A brief summary of the ups and downs of the two approaches is presented in table 1.

Table 1. Creating an ontology from scratch vs. using existing standards/classifications

Approach	Advantages	Disadvantages
Using available standards / classifications	- benefit after the work and time invested by domain experts in aggregating all the relevant knowledge in taxonomies, standards and classifications - use comprehensive classifications that cover unambiguous and well-documented descriptions of occupational titles, skills, qualifications etc.	- considerable amount of time is needed to integrate all the components (time effort split when creating an ontology reusing existing sources: 15% for source gathering, 30% customizing selected ontologies, **45%** for entire integration, 10% for refinement and evaluation) [3] - the merging results still need to be evaluated by human experts to validate the correctness
From scratch	- creation of an ontology that fits perfectly the needs of the application - avoiding to include in the ontology terms that are not going to be used (for example when using an existing ontology it may contain concepts that are not needed in the application)	- lack of knowledge to assure a comprehensive domain representation - significant amount of time is needed to design the ontology, but might be slightly less than the time required when reusing existing sub-ontologies

Almost all of the proposed ontologies rely on four main entities ([13], [16]): the *applicant* (who is looking for a professional opportunity), the *employer* (who is posting a job advertisement), *candidate's profile* (including his experience and qualification) and *job description*.

3. Strategies to improve the semantic matchmaking process

The first created prototypes were relying on an ontology used to annotate the job ads and the candidates' profiles, so that these two can be automatically matched. However, using it in real life scenarios obtained inconsistent results, mostly because of the fact that a perfect match between a resume and a job description was hard to find. In order to surpass these kind of difficulties, some strategies to improve the semantic matchmaking process were developed. Forwards, the most known strategies are presented.

3.1. Calculation of the degree of semantic similarity between thematic clusters

Several researches, like [5] [6] and [10], suggest grouping the pieces of information from both resumes and job postings into "thematic clusters" (skills, salary information, industry sector details etc.) and then calculate the similarity between corresponding ones. Also, every job

requirement can have a specified associated importance, justified by the indicated weight. The higher the indicated weight, the greater influence has on the calculated similarity. The final similarity between the job description and the candidate's profile is computed as the sum or average of the similarities between these thematic clusters.

3.2. *Query relaxation*

It usually happens that the description of the job requirements is too specific and consequently no resumes are matched. [3] [10] and [17] propose an approach based on rewriting the rules that compose the query in order to achieve what is known as "relaxation", meaning a set of approximated queries. The relaxation process is semi-automated, in order to stop it whenever the results become irrelevant. At first it tries to deliver the best results, applying exactly the requirements in a strong query, but in case the result set is empty then the query is changed replacing or deleting parts of it. For example, a C# skill can be replaced with OOP (Object Oriented Programing) which is more general.

3.3. *Under-Qualified match*

The case when the job seeker profile fulfils exactly the requirements specified in the job offer seldom happens. Therefore, instead of receiving an empty list with the best ranked candidates, [12] introduces the notion of Under-Qualified match, consisting of two types. Type 1 describes the case when the candidate has the required competency but its level is lower than the expected one or has experience but the number of years is smaller. Type 2 takes into account the fact that it is possible to have some missing competencies and in this case tries to replace those skills with more general ones. Applying one of these two Under-Qualified matches increases the chance to get relevant results although they do not completely meet the requirements. Under-Qualified match is related with the query relaxation technique.

3.4. *Using multiple semantic resources together with statistical concept-relatedness measures*

To overcome the limited domain coverage of semantic based techniques, researches like [18] and [19] propose an approach based on exploiting multiple semantic resources. In order to construct the semantic networks for the resume and the job post WordNet ontology is used, in first instance. Then, to solve the semantic knowledge incompleteness issue, concepts that were missing from the previous ontology are further submitted to a second ontology (YAGO2). Even though more than one ontology is used, most likely this will not guarantee for a full recognition of the concepts. To compensate for the missing background knowledge Hiring Solved (HS) dataset, composed from a huge number of skills and the weights of semantic closeness between them, is used. Based on the results retrieved by the HS dataset, the measures of semantic relatedness are replaced with the "related-to" relation.

4. Evaluation of Sematic Web prototypes for online recruitment

All the reviewed papers were focused on designing a prototype of an application that aimed to solve the problems of existing traditional solutions used in eRecruitment, taking advantage of Semantic Web facilities. Although a part of them, like [3], [5], [6], [12], [13] and [17] remained only at an initiative stage, without implementing or testing it, the others went a step further, developing and providing the results of the tests performed using the proposed prototype ([4], [8], [9], [14], [18], [19], [20]).

This section continues with the presentation of the common methods used to validate the online recruitment systems that make use of the advantages semantic technologies bring.

4.1. Comparing the automated results with the results calculated manually using experts judgements

In order to confirm the platform-based selection process the hiring staff has to manually check each resume to see if it satisfies the requirements. On the other hand, all the proposed jobs for a particular profile should be analysed to see if they are relevant or not, considering the

background knowledge of the candidate. This is the most straight-forward solution to validate if the results obtained using the Semantic Web recruitment platform are accurate and trustful.

4.2. Statistical metrics: Precision, Recall, and F-measure

These are widely used parameters. A high precision (exactness) shows that all the obtained results were relevant, while a high recall (completeness) reveals that nothing was missed.

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|} \quad (1)$$

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

In our case, the documents are either the resumes or the job descriptions.

F-measure is the harmonic mean of precision and recall, where β has non-negative real values.

$$F - measure_{\beta} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (3)$$

4.3. Comparing the results with the ones provided by other state of-the-art semantics-based automatic recruitment systems

Another solution to provide a ground for evaluating the quality of the new proposed solution is to make a comparison between the results of its prototype and the ones acquired using a similar platform, known for its high valuable results. In this manner, it can be easily highlighted what are the improvements achieved with the new approach. In the comparison process should be evaluated both: the *accuracy* of the given results and also the needed *time* to retrieve them. In order to assure a relevant and objective evaluation a consistent dataset against which the results are verified should be considered. Also, the two systems can be compared in terms of precision, recall and F-measure metrics.

Ever increasing use of *mobile devices* is driving *behavioral changes* in today's ever-connected workforce [21]. None of the analyzed platforms thought of a mobile prototype in order to extend its audience. However, in a world where 51.3% of internet usage worldwide is done using a mobile and tablet device (as per statistics provided in [22], in October 2016) enabling eRecruitment through mobile devices is becoming a natural step in the evolution of the job portals.

5. Conclusions and future work

Making use in the online recruitment platforms of the power semantic data brings in, has proven to increase their efficiency. All the prototypes developed using Semantic Web technologies acquired better and more accurate results in the matching process. Using an eRecruitment platform based on Semantic Web technologies usually rely on two end-users that benefit from its advantages: job seeker and recruiter. However, some researches, like [4] and [23] make an interesting point and add a new actor in the schema: the *training provider*. Whenever a job seeker fails to find job advertisements that suits his profile, the recruitment job portal should provide him with advice about trainings that will help him to increase his chances of getting a new job. Also, when a company reaches a candidate that does not fully meet all the job requirements, the platform should suggest the list of trainings that will transform the job seeker in the desired candidate. In this way the recruiter can decide if it makes sense to hire him without having all the needed expertise, but to provide him with additional trainings that will fill the discovered gaps.

All the reviewed papers were considering technical skills, beside the education and known languages. However, *soft skills* represent a valuable feature that should not be neglected. Beyond all the knowledge a person possess, his abilities to work in a team, to know how to

communicate with his co-workers and to effectively share information, are key aspects that recruiters look for. Evaluating a candidate soft abilities was performed until now only in the final phases of an interview, when human involvement was imperative. Nowadays, continuous development of technology offer the means to track the soft skills of a job seeker analyzing his activity in social media (Facebook, LinkedIn, Tweeter, Meetup and so on) using data mining techniques. The results of such profiling could be even more relevant than the abilities the candidate praise to have during the interview, when it is expected that he will highlight his outstanding skills rather than speak about his weaknesses.

References

[1] R. Kessler, N. Béchet, M. Roche, J. Torres-Moreno and M. El-Bèze, "A hybrid approach to managing job offers and candidates", *Information Processing & Management*, vol. 48, no. 6, pp. 1124-1135, 2012.
[2] T. Ermilov, A. Khalili and S. Auer, "Ubiquitous Semantic Applications: A Systematic Literature Review", *International Journal on Semantic Web and Information Systems*, vol. 10, no. 1, pp. 66-99, 2014.
[3] M. Mochol, H. Wache and L. Nixon, "Improving the recruitment process through ontology-based querying", 2006. [Online]. Available: http://ceur-ws.org/Vol-226/paper05.pdf.
[4] P. Montuschi, V. Gatteschi, F. Lamberti, A. Sanna and C. Demartini, "Job Recruitment and Job Seeking Processes: How Technology Can Help", *IT Professional*, vol. 16, no. 5, pp. 41-49, 2014.
[5] M. Mochol, R. Oldakowski and R. Heese, "Ontology based recruitment process", 2005. [Online].Available: https://pdfs.semanticscholar.org/0ea2/6e30e4037c493f06ec45eaef20d6b0ac0dd6.pdf.
[6] C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf and R. Eckstein, "The Impact of Semantic Web Technologies on Job Recruitment Processes", *Wirtschaftsinformatik 2005*, pp. 1367-1381, 2005.
[7] M. Mochol, E. Paslaru, and B. Simperl, "Practical Guidelines for Building Semantic eRecruitment Applications," in *International Conference on Knowledge Management (IKNOW 06), Special Track: Advanced Semantic Technologies (AST'06),* Graz, Austria, 2006.
[8] F. Halvorsen, "Semantic Web technology in HR Management", Master thesis, University of Oslo, 2008.
[9] W. Karaa and N. Mhimdi, "Using ontology for resume annotation", *International Journal of Metadata, Semantics and Ontologies*, vol. 6, no. 34, p. 166, 2011.
[10] M. Mochol, H. Wache, L. Nixon, "Improving the Accuracy of Job Search with Semantic Techniques", *Proc. 10th Int'l Conf. Business Information Systems*, pp. 301-313, 2007.
[11] P. Nilaphruek, "A Job Recruitment System Using Semantic Web Technology", in The 15th International Conference of International Academy of Physical Sciences, Pathumthani, Thailand, 2012.
[12] M. Fazel-Zarandi and M. Fox, "Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach", Proceedings of the 3rd International Workshop on Service Matchmaking and Resource Retrieval, 2009.
[13] B. Ionescu, I. Ionescu, V. Florescu and A. Tinca, "Semantic Annotation and Association of Web Documents: A Proposal for Semantic Modelling in the Context of E-Recruitment in the IT Field", *Accounting and Management Information Systems*, vol. 11, no. 1, pp. 76-96, 2012.
[14] D. Çelik, A. Karakaş, G. Bal, C. Gültunca, A. Elçi, B. Buluz and M. C. Alevli, "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs", in *COMPSAC Workshops*, 2013, pp. 333–338.

[15] M. Enachescu, "A Prototype for an e-Recruitment Platform using Semantic Web Technologies", Informatica Economica, vol. 20, no. 4, pp. 62-75, 2016.
[16] F. M. Hassan, I. Ghani, M. Faheem and A. Ali Hajji, "Ontology Matching Approaches for eRecruitment", *International Journal of Computer Applications*, vol. 51, no. 2, pp. 39-45, 2012.
[17] M. Mochol, A. Jentzsch and H. Wache, "Suitable employees wanted? Find them with semantic techniques", in *European Semantic Technology Conference*, Vienna, Austria, 2007.
[18] A. Kmail, M. Maree, M. Belkhatir and S. Alhashmi, "An Automatic Online Recruitment System Based on Exploiting Multiple Semantic Resources and Concept-Relatedness Measures", *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2015.
[19] A. B. Kmail, M. Maree, and M. Belkhatir, "MatchingSem: Online recruitment system based on multiple semantic resources.," in *FSKD*, pp. 2654–2659, 2015.
[20] P. Nilaphruek and R. Khanankhoaw, "The Enhancement of Efficiency in e-Recruitment System using Semantic Matching Technique", *Science and Technology RMUTT Journal*, vol. 5, no. 1, pp. 83–99, 2015.
[21] J. Rossheim, "How Technology is Changing the Recruiting Landscape", *Monster*. [Online]. Available: http://hiring.monster.com/hr/hr-best-practices/recruiting-hiring-advice/attracting-job-candidates/new-recruiting-strategies.aspx.
[22] "Mobile and tablet internet usage exceeds desktop for first time worldwide", *StatCounter Global Stats*, 2016. [Online]. Available: http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide.
[23] V. Radevski and F. Trichet, "Ontology-Based Systems Dedicated to Human Resources Management: An Application in e-Recruitment", *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pp. 1068-1077, 2006.