Boosting the Recruitment Process through Semi-Automatic Semantic Skills Identification

Alexandra Cernian, Valentin Sgarciu

***Abstract*—This paper presents the design and implementation of a semi-automatic semantic tool for boosting the recruitment process. It is able to analyze a set of CVs in order to identify a suitable candidate for a particular job. The process is based on semantic processing of the resumes and matching the candidates' skills with the requirements for each particular job description. The purpose is to reduce the recruiter's processing time by eliminating certain repetitive activities in the resume analysis procedure and to obtain a qualitative improvement by highlighting the competencies and qualities of candidates based on complex and customized semantic criteria.**

***Index Terms*—data mining, semi-automatic tool, recruitment process, semantic technologies**

I. Introduction

People are the most valuable asset of an organization. One of the key roles of HR specialists is to find and hire the right people for the right job. But how do we find them? Their profile must match the organizational vision and mission, their set of values must be convergent with the organizational culture and their skills must match the job description and requirements. The evaluation process during recruitment is the key to finding the appropriate candidates in a timely manner.

At present, the recruitment process is mainly assisted by human experts, which can be a time consuming task. Moreover, we assist at a shift in in recruitment methods, from the traditional one to a social recruiting, where decisions are based on social media analysis and on networking decision factors [1].

Companies currently create and use networks of leads, candidates, employees and alumni in order to leverage and increase the quality and performance of the recruitment process.

An automation we can bring into the process is a smarter way of processing candidates' resumes. If we build an improved and more intelligent parsing and skills matching algorithm, the entire recruiting process would benefit and we would be able to efficiently identify the suitable candidates for a particular job. The final objective is to minimize the HR specialist input in the process and limit the time they spend in processing resumes by developing digital tools to automate the candidates identification phase. Such a tool would bring two main benefits:

1. Eliminate certain repetitive activities in the resume analysis procedure, currently performed by HR specialists mostly empirically and
2. Improve the process from a qualitative point of view by identifying and highlighting the competencies and skills of candidates based on customized criteria [2].

This paper presents a semi-automatic tool for boosting the recruitment process through a semantic approach for identifying candidates skills and matching them against organizational requirements and job descriptions.

The rest of the paper is organized as follows: Section 2 presents and overview of the system, emphasizing the main objectives and benefits of this approach, Section 3 presents the design and implementation of the tool and Section 4 draws the conclusions of the hereby presented work.

II. Overview of the Semi-Automatic Tool for Identifying Candidates Skills

This objective of this paper is to present the design and implementation of a semi-automatic tool able to assist HR specialists in the recruitment process by providing an automatic analysis of the candidates resumes and identifying their key skills, by also matching their competencies against the requirements in the job description. The system will be able to recommend the best suited candidates for specific positions.

The process is based on a semantic processing of the candidates resumes, based on a dictionary approach. The flow of the application is presented in figure 1.

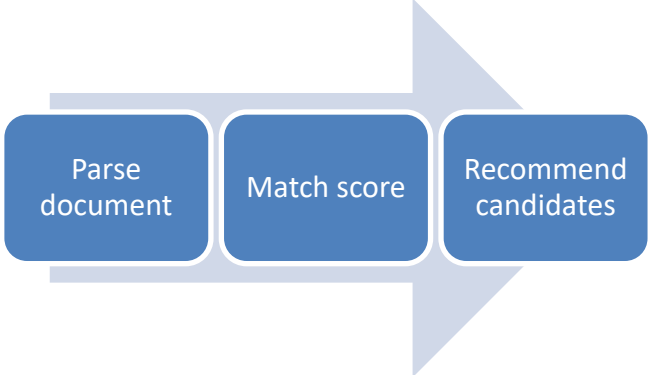

Figure 1. The main components of the tool

Alexandra Cernian is with the University Politehnica of Bucharest, Faculty of Automatic Control and Computers, Bucharest, 060042, Romania (phone: +40 744 632 031; e-mail: Alexandra.cernian@aii.pub.ro).

Valentin Sgarciu is with the University Politehnica of Bucharest, Faculty of Automatic Control and Computers, Bucharest, 060042, Romania; (e-mail: vsgarciu@aii.pub.ro).

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The main functional specifications of the tool are more exhaustive and include various actions conducted by the HR department, namely:

1. Provide the possibility to define and update job descriptions
2. Keep a database of candidates resumes
3. Perform searches to identify best suited candidates

The main objectives of the application are the following:

1. Implement a semi-automatic tool to assist HR specialist in the recruitment process
2. Implement an automatic module for parsing the resumes of candidates and identify their key skills and competencies
3. Develop an algorithm that automatically finds the best matches for a specific job profile based on candidates skills

III. The Design and Implementation of the Tool

The candidates' selection algorithm starts each search process based on the configurations made by the user and has the following flow:

- The user configures the job description and candidate profile
- A special section is dedicated to the main competencies and skills required
- The candidates resumes are saved into the database
- The parsing algorithm is used to process the candidates resumes
- The candidates profile is matched against the job description and a similarity score is computed
- The list of best matches is generated by the application for each specific job description

The tool has three main modules:

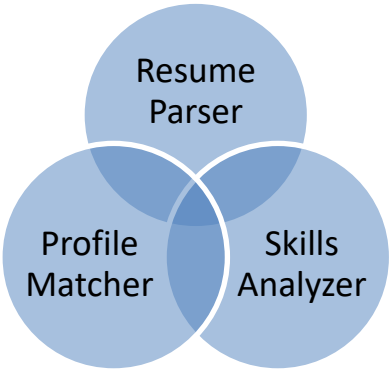

Figure 2. The three main modules

A. The Resumes Parser

This component is responsible with processing the candidates' resumes in order to extract information about their education, work experience and different skills. The parser processes information in four main phases:

1. Get input data from the user (one or more resumes) and transform them into text files
2. Parse the CV in order to split the information into individual sections: personal details, education, work experience etc.
3. Classify specific words in each section in order to leverage the candidates' selection model
4. Store the information in a database.

The tool used to parse the resumes is an open source library called Apache PDFBox [3][5], which facilitates document management, providing functions to create documents and extract content from existing documents in different formats.

B. The Skills Analyzer Module

This module is the most complex of the tool and it is based on a three steps process (Figure 3):

Figure 3. The skills analyzer module architecture

1. Skills dictionary (Figure 4): We developed a specialized taxonomy for abilities and skills dedicated to resumes processing in order to assist HR recruitment process. The taxonomy contains 75 keywords and we plan to extend it to provide an exhaustive list of skills looked for upon recruiting and selecting candidates.

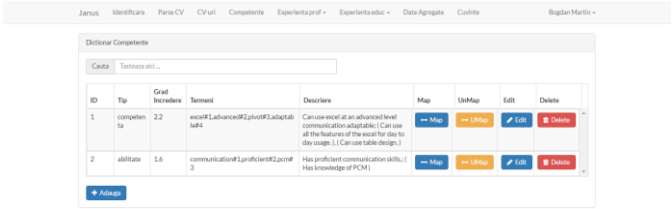

Figure 4. Skills taxonomy

2. Identify skills in resumes: This component processes the resumes and checks for skills by looking them up in the skills taxonomy. They will be further used by the Profile Matcher module to determine the candidates that best suit each specific job description.
3. Extract educational background: The resumes are parsed and the education section is processed. This component is based on two dedicated dictionaries (taxonomies) and has the purpose to provide a classification of the abilities and competencies that the candidates have acquired during their studies and work experience. The dictionary components are based on the SentiWordNet semantic taxonomy [4].

Figure 5 shows the database tables associated with the dictionary (taxonomy) components [5]:

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

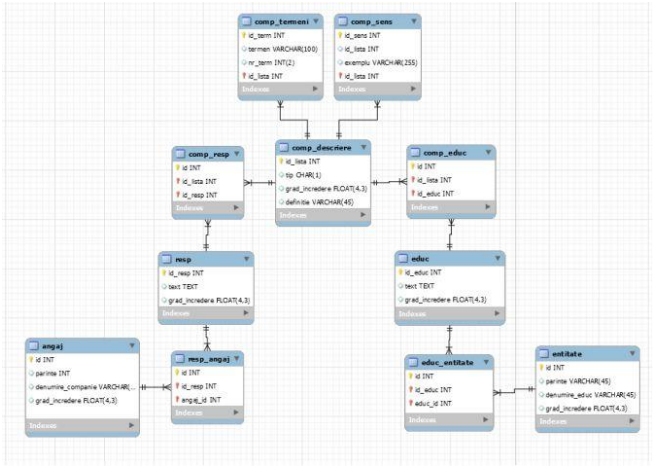

Figure 5. Taxonomy database tables

At this point, the CVs are parsed and specific key terms have been identified and classified. The next step is to compute a matching score for each CV, in relation to the requirements for a specific position in a company.

C. The Profile Matcher Module

This module uses a score-based algorithm to determine the candidates that best match specific job descriptions, by taking into account the work experience, skills and educational background of applicants.

The score associated with a CV and its position between all the analyzed CVs is based on the number of identified terms that correspond to the search criteria, such as skills, job title, keywords or job responsibilities, as well as educational items related to employers and educational institutions in relation to the required skills.

The profile match phase has several steps, as follows (figure 6):

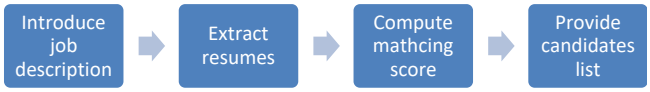

Figure 6. The profile matching process

1. *Introduce job description*

The user introduces the job description with the following input: competencies, abilities, free text, job title, professional responsibilities, education and interests. If some elements (skills, education, experience) have a higher priority for a specific job description, the HR specialist can provide a weight for them, which will be taken into account when computing the matching scores.

2. *Extract resumes rom database.*

The resumes have been previously parsed into sections and split data.

3. *Compute matching score*

Step 1. For each CV, a score is computed for each section. A section score is calculated by the number of appearances of elements that define the competencies or other required parameters

$$score_{section}=scores_{ection}+n*weight \quad (1)$$

If the weight parameter is not customized, it will be considered equal to 1 by default.

Step 2. Compute score for CV

The score for the CV is calculated with the following formula:

$$score_{cv}=\Sigma score_{section} \quad (2)$$

4. Provide candidates list

Step 1. Sort the candidates resumes by score.

Step 2. Show the list of candidates based on the highest scores.

IV. Conclusion

The application has been tested using a set of 250 candidates profile and 15 job descriptions and the results were 97% positive. The scores were accurately computed and the recommendations fitted the required profile for specific jobs. The results were improved and the accuracy was higher when specific weight parameters were provided for specific skills. The semi-automatic tool for parsing resumes and identifying candidates' abilities and competencies leverages the selection process for recruitment specialists and reduces the time needed to manually process the information in the candidates CVs. Moreover, due to the integrated semi-automatic learning module, the competencies dictionaries are constantly updated and the selection model is improved. This tool can be a relevant decision support instrument for HR and recruitment specialists in identifying the best suited candidates for specific jobs, based on their profile and experience.

References

[1] Josh Bersin, Predictions for 2015 Redesigning the Organization for a Rapidly Changing World, Deloitte, 2015: http://www.cedma-europe.org/newsletter%20articles/misc/Predictions%20for%202015%20-%20Redesigning%20the%20Organization%20for%20a%20Rapidly%20Changing%20World%20(Jan%2015).pdf

[2] Hsiu-Fang Hsieh, Sarah E. Shannon, Three Approaches to Qualitative Content Analysis, QUALITATIVE HEALTH RESEARCH, Vol. 15 No. 9, November 2005 1277-1288, DOI: 10.1177/1049732305276687

[3] PDFBox: https://pdfbox.apache.org/index.html, last accessed July 2016

[4] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastian, SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining: http://sentiwordnet.isti.cnr.it/ , last accessed July 2016

[5] A. Cernian, D. Carstoiu, B. Martin, "Semi-automatic tool for parsing CVs and identifying candidates' abilities and competencies", International Conference on Education, Management and Applied Social Science – EMASS, 2016

ISBN: 978-988-14048-3-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)