A Preliminary Study on Text Mining–Based Human Resource Allocation in a Construction Project

Sangil Han [a] and Ghang Lee [a]

[a] Yonsei University, Republic of Korea
E-mail: hancowr@gmail.com, glee@yonsei.ac.kr

Abstract –

Even engineers who have the same job title and are working on the same project have different skills and backgrounds. However, despite the importance of assigning the right person to the right job to ensure successful project delivery, current human resource (HR) allocation practices are only concerned with the list of projects in which a candidate has been involved, years of experience in a discipline, and their previous job titles. As a result, some engineers lack the knowledge and experience to perform the tasks they are assigned on a project. This study demonstrates the need for a long, descriptive résumé rather than the commonly used short, brief résumé. When the job candidates who participated in this study were asked to describe their work experience in several sentences on their curriculum vitae (CV) instead of describing them using a few words composed of just the job title and roles, the contents of the CVs were clearer and showed less bias. This paper therefore presents a text-mining algorithm and a semantic résumé analysis system that we developed to automatically extract and analyze the work experience candidates write about on their résumés. This algorithm and analysis system can help HR managers control the significant amount of information found on prospective employees' long résumés. To validate the algorithm and the system, six sample résumés were collected anonymously. These résumés were then reviewed both manually and by the developed system. In a subsequent study, we expect to apply the text mining–based HR resource allocation algorithm to select construction engineers for a real project and measure the job-matching rate.

Keywords –

Construction engineer; text mining; human resource allocation; résumé; KNIME

1 Introduction

The purpose of this study is to develop an approach to resolve the problems occurred during the allocation of human resources (HR) in construction organizations by applying the text-mining method to long descriptive résumés. Searching for the best candidate for a specific job at a construction project is one of the main challenges of HR managers in the architecture, engineering, and construction industries. Although large companies have their own HR management systems that include information technology (IT) solutions (such as web-based management systems) to facilitate job allocation, their current usefulness and accuracy are questionable. The reason for this is that most HR management (HRM) systems only allow HR managers to search for appropriate engineers or managers for a new project based on previous team names or the names of project that the engineers/managers worked on, job titles, durations of each project, etc. Nevertheless, how candidates perform on a new project greatly differs depending on the specific tasks that each candidate performed for previous projects and the experience and skills that they gained through these tasks. For example, if a new project needs a site manager for a building construction project (referred to as an architectural site engineer, hereafter), an HR manager generally looks for potential candidates from an HR system by searching for employees who previously worked as an architectural site engineer. However, architectural site engineers perform highly diversified tasks throughout the entire period of a project and also work with different types of subcontractors and vendors and on different contractual systems, but the employee information in an HR system does not include these data, which therefore hinders the HR manager from finding the best person for the new project.

Similar problems exist when hiring new employees. Résumés or curriculum vitae (CVs) are often composed of discrete words or short sentences and thus are often vague and unclear. Some companies ask for a statement of purpose (SOP) in addition to a CV, but long SOPs are very time-consuming to analyze, especially when there are many applicants (for the more popular companies in South Korea, it is common to review hundreds of applicants for a single position). In addition, with new skill sets required due to the increasing globalization of the construction market and the rapid development of computer integrated construction and building information modeling, the need for an improved and detailed job classification system is increased [1]. To overcome these limitations, construction companies have tried to develop an HR system that can support

detailed job classifications. However, a drawback of the new systems is that they ask applicants and employees to provide too much information, sometimes by selecting an item from a multi-layered dropdown menu and sometimes by filling in blank text boxes. A more critical drawback of such systems is that they will always be limited in their ability to capture all the details of various types of construction tasks that the applicants and employees feel are important.

Therefore, this study has developed an improved HR solution for finding the best person for a job and for maintaining the records of project staff by collecting and analyzing long sentence-based job descriptions related to the construction industry using a text-mining method instead of keyword-based job descriptions. The text-mining system compensates for the natural shortcoming of long descriptions by refining and stemming key factors from the descriptions.

2 Previous studies

2.1 Literature Review

The importance of HRM has been emphasized by many researchers. As intellectual capital has become one of the most strategic assets of successful organizations, the ability to manage the expertise, skills, and experience of employees has become a key factor in overcoming the increasing competitiveness of the global market [2]. In today's competitive business environment, companies need to accurately grasp the competency of their HR in order to be successful [3]. However, technology companies often suffer from high turnover rates and find it hard to recruit the right talents [4]. Due to the current trend of pursuing better HRM options, an increasing number of publications concerning data mining in the context of HRM are providing new research data [5]. For example, one study investigated the relationship between web data mining and strategic HR planning by developing a practical model [6]. Another study focused on the HRM performance evaluation process by introducing a data-mining method termed knowledge discovery in databases [7].

Thus far, filtered résumés have been the most important resource for HR departments when hiring new employees [8]. One study introduced an approach to improve the matching of profiles by searching job descriptions and applicant profiles using filters that represent the relevant skills and competencies [9]. Nevertheless, several studies have found that résumés and work experience lists, which are composed of brief words or short sentences, are limited and have tried to improve HR solutions by adopting a semantic system approach, such as ontologies and text-mining methods. An example is the On-To-Knowledge project. The On-To-Knowledge project focuses on the application-driven development of ontologies during the introduction of ontology-based knowledge management systems [10]. One research study has suggested that a possible approach could be addressing an intelligent decision support system composed of case-based reasoning and ontology [11]. A case study focused on the electronic market introduced a unique configurable framework that was designed to be adapted for electronic marketplace applications [12]. Another research stressed the importance of HR recruiting, selecting individuals for teams based on different skills and qualifications, determining who to train and what training programs to offer, and recommending the right expert to individuals for acquiring information or learning from within the organization [3]. In addition, previous research has focused on the overall HR domain, whereas researchers studying the construction industry specifically have not attempted to actively apply text mining to HRM.

In addition, to fully understand the current status of HRM in the construction industry, we reviewed the relevant articles and papers relating to the inherent problems of construction HRM. One study pointed out the importance of an automated HR information system using a case study of Australian construction companies to prevent chronic problems like high turnover, lower job satisfaction, and motivation [13]. Another study stressed that a typical Korean construction organization has been simply classified by terms such as architect, safety, quality, administration, etc., even though it should be categorized with more detailed activities, such as claim/contract management, design management, HRM, document management, information management, risk management, etc. [14]. A report issued by a Korean research institute pointed out that construction companies' job classifications have not been set up systematically to trigger inefficient HRM and material resource management [15].

2.2 Existing Web-Based HR Systems

Probably the most widely well-known web-based HR systems are Linked-In and OilandGas. Their search systems are designed so that a recruiter can input a few words, and within milliseconds, the web page shows the search results of résumés through keyword matching, similar to portal web searching like Google. With the special services of Linked-In, recruiters can also search for specialists and candidates from various industries, not only for the construction industry. On the other hand, the OilandGas service is specialized for searching for experts in the oil and gas industry. These services use basically the same concept of discrete keywords matching. Hence, users can put in keywords, like a

position or job title, add some more information like nationality, and then search. The results can be anywhere from just a few to thousands of CVs. However, once this search process is completed, the recruiter has to download all the search results and résumés and then read through all the documents to pick the most suitable candidates.

3 Analysis of the Current Practices

To fully understand current practices, we conducted an interview with a HR manager at one of the largest construction companies in Korea. He has been working on the HR team at the company's headquarters for about 5 years and is in charge of recruiting experts for international construction projects. Due to the growing needs for the international construction market—especially for the oil and gas plants and the power plants, the HR overseas department target-recruits experts suited for those projects from outside the company. We asked the HR manager several questions related to our research subject:

1. How do you find a person who is suitable for a particular job?
2. How do you find new employees from outside the company?
3. Do you keep detailed records of each employee's job experience?
4. If so, how do you collect and manage detailed job experience information? Do you use long descriptions or the more traditional keyword-based HR database?
5. How important do you think are long descriptions about job experience for job allocation?
6. Do you use an automated HR resource allocation or recommendation system?

For the job allocation of employees, this particular company uses its own web-based HRM system that had fields to input detailed and descriptive records of employees' past job experiences, but these fields are optional. Thus, according to the HR manager, most staffs do not input these descriptions. In addition, the HR department does not consider these descriptions to be important, especially for low-profile staff. Internal job allocation of employees relies heavily on recommendations from previous project managers. For the recruitment of experienced employees from outside the company, the interviewee responded that résumés and SOPs were essential to filter out unqualified personnel and to allocate an individual most suitable for a particular job. Nevertheless, the analysis process is very time-consuming and challenging, as the number of applicants for each position has grown rapidly. According to the HR manager, the company uses its own HR system for controlling and managing all the employees. This system handles various aspects of individual staff, from private information like family members to involved projects of the company or of previous companies. But, like most Korean companies, work experience and job descriptions are expressed by a single word such as 'Architectural (construction)' or 'safety,' along with the involved project or team name.

For comparison, we collected other cases of Korean top-ranking construction companies and found that their HR systems were very similar. For companies A and B, their web-based HR systems showed only job position and its domain, such as architecture or infrastructure, and even such categories were generally left blank. This shows that most Korean construction companies do not seem to consider individuals' detailed experiences as important factors for HRM and that they prefer simple job categories due to ease of management. Also, we can assume that the Korean construction companies' HRM relies on simple facets of each employee, like the level of their position (staff, assistant manager, senior manager, etc.), their job title, and their reputation and references.

However, the interviewee's company has a unique system for collecting long descriptions of each staff. The staff members write their own job descriptions freely, using their own judgement to rank their skills and experience. It may seem like the most useful type of information for depicting each staff's career and abilities for certain tasks and positions; however, the company does not view this as a beneficial asset for HRM. The input guidelines consist of two sentences only, so the staffs also do not want to put their energy and time into filling out long descriptions to express their experience and major job skills.

From exploring the current status of HRM in common Korean construction companies, we can conclude that the HRM mainly relies on fragmentary factors of individuals or on their reputation that is transferred verbally. Hence, the purpose of this study to improve the current shortcomings of the HR systems of construction companies seems to be reasonable.

4 Introduction to the Text Mining–Based HR Allocation System

This section explains the text mining–based HR allocation system that we propose. It briefly explains the basic concept of the proposed text-mining method and then demonstrates the applicability of the proposed method using one of commonly used text-mining tools and a sample case.

4.1 Overview of the Proposed Method

The first step to accomplish the method is setting up the basic concept algorithm as follows:

1. The main objective is to use text mining to efficiently and effectively analyze the résumés of candidates for allocating positions for a construction project.
2. Setting the basic requirements of each position by searching for typical job requirements or applying an organization's own requirements.
3. The main usage of the method is (1) for in-house employees and (2) for new job seekers from outside the company.
4. The natural languages consist of long descriptive sentences in résumés that are modified and refined by the flow, such as 'input,' 'parsing,' 'filtering,' 'collecting,' 'matching,' and 'result showing'.

By using the concept of text mining to build the proposed system, we constructed a flow chart of the entire process for the final judgement of whether a candidate is the best fit for a designated position. The flow and its step-by-step explanation are shown in Figure 1.

- The construction domain must be selected before starting the text-mining process among industry divisions such as 'building,' 'infrastructure,' and 'plant' since each sector's expertise is defined specifically for its own purposes and characteristics.
- The input data is divided into two major categories: (1) for in-house use and (2) for hiring external candidates. The input data to be validated has to fully contain sufficient information of the personnel semantically before entering through the algorithm.
- The mining process consists of 'input,' 'filtering & parsing,' 'classification,' and 'matching.' This process enables the natural language used in the long descriptions in a CV to be converted into strings that are readable and comparable by a computing system. The last step 'matching,' is implemented after the requirements of the position are received, which was derived from a separate source. The requirements can be defined by any source, but they should cover the position as exhaustively as possible using essential keywords.
- The end of process is position matching; e.g., how well a candidate is suited for a position or a recommendable position for a particular candidate.

We would like to emphasize that this text-mining process requires that candidates' CVs are composed of long descriptive sentences that explain the candidates' work experience in detail. Otherwise, there is no reason to develop this much more complicated system as compared to the traditional HRM system that only uses short keywords.

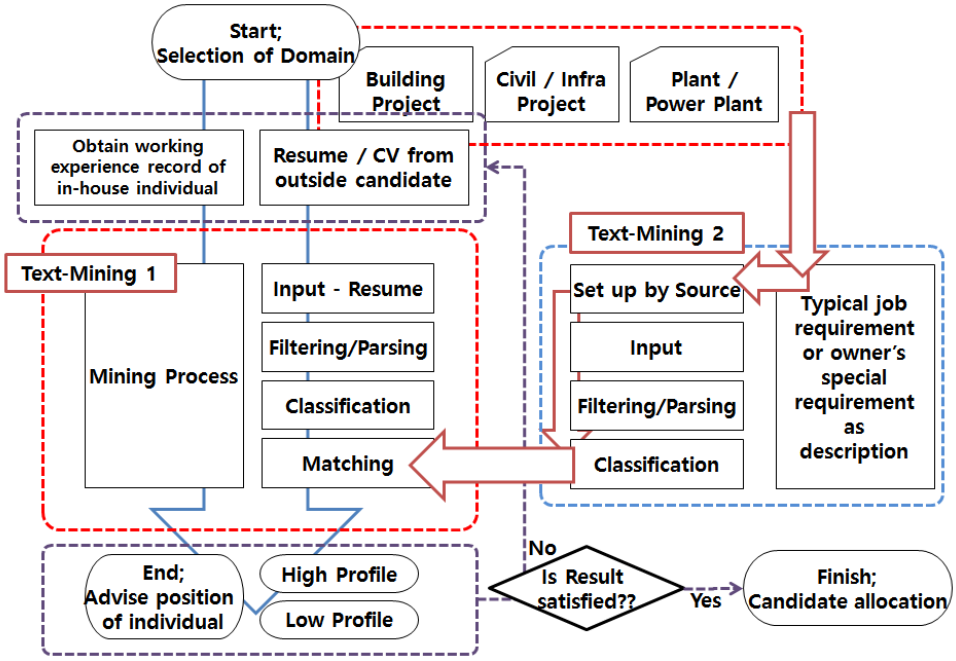

Figure 1. Text-mining flow

4.2 Applying the text-mining tool

The concept of text mining has been used since it was introduced, and many research teams and organizations have created text-mining platforms for various purposes. The text-mining tool employed for this study was KNIME, which is one of the more widely used open platforms (Figure 2). Since 2006, KNIME has been used in pharmaceutical research, but it is also used in other areas like CRM customer data analysis, business intelligence, and financial data analysis. This platform has intuitive visual, so users can build the process with a visualized work flow.

4.2.1 Main Functional Nodes Set-Up

In the introduced text-mining platform, 9 functional nodes were dragged and linked to conduct the mining process for résumé and requirement in accordance with the framework. In KNIME, we numbered the nodes from 1 to 9 so that the input resources—résumés and job requirements—were transformed when they went through the nodes in the system.

- #1. Document parser: This node reads the documents and creates a document for each file. The full text is extracted.
- #2. POS tagger: This node assigns a part of speech to each term in the document (tag).
- #3. Bag of words (BoW) creator: This node creates a BoW for a set of documents.
- #4. Number filter: This node filters all terms consisting of numbers only.
- #5. Punctuation erasure: This node erases the punctuation characters from terms.
- #6. Stop word filter: This node filters all terms contained in the specified stop word file (such as 'a,' 'about,' 'but,' 'could,' 'the,' 'from,' etc.).
- #7. Stemmer: This node stems terms by stemmer algorithms.
- #.8. Keyword extractor: This node extracts relevant keywords based on frequency and its own scoring method.
- #9. String matcher: This node compares two lists of strings, computes the distance between the strings, and lists the most similar strings. (These functional node descriptions are cited from the KNIME platform.)

The functional nodes were extracted from the Node Repository sector of the KNIME platform categorized as 'Text Processing.' Each node has 3 color signals similar to traffic lights so a user could easily recognize when an error occurs and amend the process effectively. In addition, a user can adjust the options of each node, which affects the level of function, to control the outcome level of the process.

5 Test Implementation of Text Mining

5.1 Collection of Test Résumés and Job Requirements

To assess the usability and reliability of the structured text-mining process, we collected six résumés of different candidates who were willing to be assigned a construction manager construction manager (CMr) position. We also obtained the job requirements of the CMr position from the Civil Service Commission of State of New Jersey, USA, website

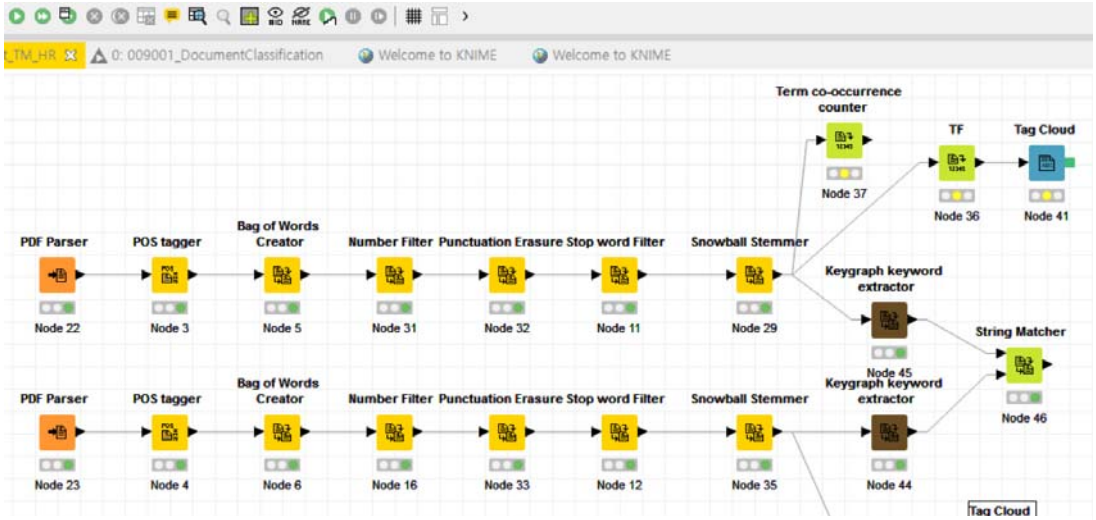

Figure 2. Main window of KNIME.

(http://www.state.nj.us), which provides official descriptions of construction job positions. Job requirements can be obtained from various channels, so we have searched many sources through web searching and our literature review. We selected the aforementioned website from this search process because it offered sufficient descriptions of construction job positions for our text-mining test. The candidates' CVs for our test were collected anonymously from a commercial service provider. The résumés and requirements were written in English and in long sentences. All of these input resources were in a PDF format, and no artificial amendments were added. We gathered samples from only six individuals, although we could have collected more CVs from various sources because we needed to verify each résumé carefully and compare it with the text-mining results within the limited research time, which is a potential limitation of this study.

5.2 Conceptual Design of Weight by Distance

To measure the degree of accuracy that the candidates were matched to the job position of construction project manager, we used the matched results of the 'String Matcher' node, which showed how many keywords were coincidental or related between two input resources—the job requirements and candidates' résumés—as a result of the text-mining process. Two important indices were extracted for assessment, one was the number of matched strings and the other was the distance between each string. The distance indicated the accuracy of two matched keywords; for instance, '0' was perfectly coincidental and '1' showed a small, but ignorable, difference between two keywords. Two strings with a distance of less than '1' (e.g., 0 and 1) were considered matched strings, while strings with a distance of more than '2' were assumed to not be paired words. Thus, in the formula for calculating the matching score, a distance of more than 2 was not considered.

Hence the six different results were compared by the number of matched strings under a distance of '1.' To make this comparison, we decided to input the weight for '0' distance and '1' distance only. '0' had a weight of 1, and '1' had a weight of 0.9.

$$\text{Score} = W0 \times n(D0) + W1 \times n(D1),$$

where
D0(Distance 0) has weight W0 = 1,
D1(Distance 1) has weight W1 = 0.9,
Distance (≥2): Weight = 0 (excluded in formula),
n(D0): number of DO, and
n(D1): number of D1.

5.3 Running and Validating the Results of the Text Mining

Candidate A's score was 3 matched keywords, 'respons', 'project,' and 'manag,' with a distance of 0. Candidate B's score was 4 matched keywords with distances of 0 and 1. Candidate C and candidate D had a score of 5 keywords within less than a distance of 1. Candidate E had 4.9 points. Finally, candidate F scored 8 matching keywords with a distance of 0 and 1 (Figure 3). In terms of the validation rule we set up, candidate F received the best score and was therefore the most qualified for the CMr position (Table 1).

To validate the results of the text-mining process, we asked an HR manager who was not informed of the results of our text mining to independently assess the six résumés. Similar to our results, the HR manager judged that candidate F was the best option for the CMr position among candidates, and he mentioned he would disqualify candidates A and B. He judged candidates C and D to be acceptable but not fully qualified and candidate E to be relatively qualified (he didn't have experience as a project manager). We had a short discussion after the manger's appraisal and concluded that candidates A and B did not have enough experience to qualify for the CMr position. Candidate A had experience as a quality control manager, and candidate B had worked as a labor control manager at the

Table 1. Text mining result

Candidate sample	Main Experience	Match Result Score
A	Quality Control Manager	3 D0 x W0 + 0 D1 x W1 = 3 x 1.0 + 0 x 0.9 = 3.0
B	Resource Management	2 D0 x W0 + 2 D1 x W1 = 2 x 1.0 + 2 x 0.9 = 3.8
C	QM & CM	2 D0 x W0 + 3 D1 x W1 = 2 x 1.0 + 3 x 0.9 = 4.7
D	Construction Manager	3 D0 x W0 + 2 D1 x W1 = 3 x 1.0 + 2 x 0.9 = 4.8
E	Construction Manager	4 D0 x W0 + 1 D1 x W1 = 4 x 1.0 + 1 x 0.9 = 4.9
F	Construction Manager	6 D0 x W0 + 2 D1 x W1 = 6 x 1.0 + 2 x 0.9 = 7.8

subcontract level. Candidates C and D were relatively better than A and B, but both résumés did not contain enough description of roles and duties relevant to the CMr position. Candidate E was also better than candidates A and B and a little bit above candidates C and D, but his experiences were mainly related to contract and QS part. The last candidate F, who achieved the best score through text mining, had more experience as a CMr listed on his résumé than the other candidates.

Based on these results, we tentatively concluded that the deployed text-mining method has potential to be used for HRM of the construction industry, especially for allocating candidates to proper positions when the HR manager needs to screen a large amount of résumés in a short amount of time. If the algorism of text mining is amended and elaborated further, the results can be even more accurate and satisfactory.

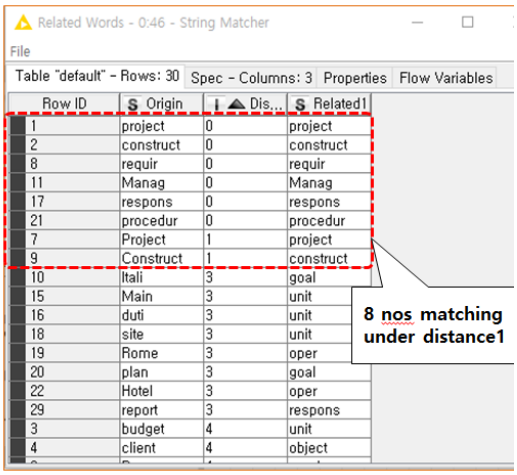

Figure 3. Matching results of candidate D in KNIME.

6 Conclusion

This study started with a question regarding the current methods that HR managers in construction companies use. It was also motivated by the struggle to find a better way to manage HR, especially position designation for construction projects for those with different, unique experiences and training. We based our study on the premise that long descriptive résumés or CVs are essential for exactly figuring out a candidates' capabilities and qualifications in view of the roles and responsibilities from his or her career history, but this information has not been used well by construction companies due to the difficulties of management. This study does not claim that a long descriptive résumé is superior to a set of short keywords for HRM, but this was the premise of this study. From this premise, in order to overcome the drawback that the long descriptive résumé cannot be easily controlled and classified in practical level, we introduced the text-mining method that is commonly used to treat large amounts of semantic data and also is used by the HRMs of other industry domains. The purpose of the text-mining process was to compare and match the keywords taken from a candidate's résumé and the job requirements, and the process was implemented by the widely used platform, KNIME. Six candidates' résumés and one document listing the CMr position requirements were used as input data for the text mining. Based on the results, we tentatively conclude that the text-mining method has potential for screening and selecting proper candidates for a certain position by comparing the results with an HR manager's manual judgement, which was similar. The text-mining method, however, only took less than a minute for inputting the data and receiving results, while the HR manager took around 30 minutes—5 to 7 minutes for each resume—for screening and 1 hour to fully review the six résumés. This means that this kind of system could improve current HRM of construction companies when searching for an appropriate person to assign to a position by automatically screening out many inappropriate candidates, at least in regard to time. If an HR manager needs to review a bunch of CV documents, then the time gap between the manual and automatic methods will be far more than even we have expected during the screening stage.

We did encounter several limitations as the research proceeded. The text-mining algorithm was not built perfectly with full consideration of various alternatives and conditions because it was challenging to gather enough résumés that contained long-enough descriptions. In addition, the algorithm was not designed to consider unique characteristics of projects and the importance of a candidate's recent project experience, but rather was focused on key word parsing, extracting, and matching from typical requirements using only the basic algorithm of text mining. However, it has led to the question of whether the text-mining method can be practical in real situations of job positioning and target recruiting of experts. Methods to effectively gauge the level of accuracy of text mining and to determine the proper keyword extraction and matching ratio must be addressed in future studies. In addition, the introduced text-mining platform needs to be validated by comparing it to other platforms. Moreover, this study has only concentrated on manager positions that are directly involved in construction management in a construction company. Thus, it has not been shown to be applicable to the entire domain of HRM in the construction industry. And lastly, this text-mining method might not seem different from existing

web-based recruiting systems based on keyword input and matching mechanisms in terms of aim and function, even though the basic concept and the process itself are different.

Although these shortcomings and limitations should be taken into account, this study has made a meaningful step forward as a preliminary study to reach practical use and to extend various domains for improved HRM by applying a text-mining methodology to the construction industry, which encompasses relatively diversified positions that require various training and experiences compared to other industries. Specially, the text-mining method provides an opportunity to dramatically save time compared with the traditional method of reviewing CVs. We hope that this text-mining methodology can be used to more effectively match candidates with the best positions as a very helpful tool for HR managers in construction projects in a shorter amount of time.

References

[1] A. Ekholm, S. Fridqvist, A conceptual framework for classification of construction works, Electronic Journal of Information Technology in Construction, 1 (1996) 1-25.

[2] S. Colucci, T. Di Noia, E. Di Sciascio, F.M. Donini, M. Mongiello, M. Mottola, A formal approach to ontology-based semantic match of skills descriptions, Journal of Universal Computer Science, 9 (2003) 1437-1454.

[3] M. Fazel-Zarandi, M.S. Fox, Semantic matchmaking for job recruitment: an ontology-based hybrid approach, in: Proceedings of the 8th International Semantic Web Conference, 2009.

[4] C.-F. Chien, L.-F. Chen, Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, Expert Systems with applications, 34 (2008) 280-290.

[5] S. Strohmeier, F. Piazza, Domain driven data mining in human resource management: A review of current research, Expert Systems with Applications, 40 (2013) 2410-2420.

[6] H. Al-Zegaier, H.A. Al-Zu'bi, S. Barakat, Investigating the link between web data mining and strategic human resources planning, Computer and Information Science, 4 (2011) 67.

[7] X. Zhao, An empirical study of data mining in performance evaluation of HRM, in: International Symposium on Intelligent Information Technology Application Workshops, IEEE, 2008, pp. 82-85.

[8] D. Çelik, A. Elçi, An ontology-based information extraction approach for résumés, Pervasive Computing and the Networked World, 7719 (2013) 165-179.

[9] A.L. Paoletti, J. Martinez-Gil, K.-D. Schewe, Extending Knowledge-Based Profile Matching in the Human Resources Domain, in: Database and Expert Systems Applications, Springer, 2015, pp. 21-35.

[10] T. Lau, Y. Sure, Introducing ontology-based skills management at a large insurance company, in: Proceedings of the Modellierung, Citeseer, 2002, pp. 123-134.

[11] I. Zhukova, M. Kultsova, M. Navrotsky, A. Dvoryankin, Intelligent support of decision making in human resource management using case-based reasoning and ontology, in: Knowledge-Based Software Engineering, Springer, 2014, pp. 172-184.

[12] D. Veit, J.P. Müller, M. Schneider, B. Fiehn, Matchmaking for autonomous agents in electronic marketplaces, in: Proceedings of the fifth international conference on Autonomous agents, ACM, 2001, pp. 65-66.

[13] S.T. Ng, R.M. Skitmore, T. Sharma, Towards a human resource information system for Australian construction companies, Engineering Construction and Architectural Management, 8 (2001) 238-249.

[14] JY. Cho, YS. Kim, Improving Construction Management Functions of the Korean Construction Companies through the Reformation of the On-site Organization , Educational Program and Career Management Systems JOURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Structure & Construction, 17 (2001) 87-94.

[15] J.S.K. T.H. Kim, Improvement way for job classification of construction company, in, Construction Economy Research Institute of Korea, Korea, Republic of, 2001, pp. 86.