Copyright 2019 by the Associated Schools of Construction

Semantic Analysis to Identify Expected Competencies of Construction Management Graduates

Sherif Attallah, Ph.D.; Tarek Mahfouz, Ph.D. and James W. Jones, Ed.D.
Ball State University
Muncie, Indiana

The construction industry is unique in terms of the diversity of calibers that form a team for a successful project and Construction Management (CM) programs strive to offer a pool for most of these calibers. To achieve this, CM programs must understand the requirements of companies and close any gaps between the current curriculums and the industry needs. One effective way to do this is by direct assessment of the recruiting efforts and exploring the details that companies include in search for certain competencies. Therefore, this paper presents a novel method to identify the competencies that contractors expect from CM graduates using semantic analysis. The proposed method utilizes collection of job descriptions published by contractors to recruit newly CM graduates. These job descriptions are analyzed through algorithms targeting semantic analysis of the wording and included in the contractors' job description to identify the most significant competencies expected for certain jobs. An example of a cost estimator being recruited by mechanical/electrical contractors is used to illustrate the results. The proposed method can be used as an effective way to understand contractors' expectations with the objective of further improvements of CM programs and holding more dialogues between the industry and academia.

Key Words: Construction Management – Competencies – Job Descriptions – Construction Education

Introduction

One of the main roles of construction management (CM) programs in the USA university is to prepare future CM professionals to understand the industry, gain skills that will enable them to contribute to the success of their companies and to have the competencies required to do specific jobs within these companies. To achieve this goal, the leaders of the CM programs who prepare and manage courses within the different curricula, must understand the needs of the industry and have a continuous dialogue with the construction companies to close any gaps between the industry needs and what is offered in universities and colleges. There are several means and methods including survey works and meetings between faculty members and companies. Advisory boards can be a very effective means to have this dialogue as well. Conferences and student competitions that involve companies as sponsors and student teams are also another platform for communicating these needs.

In additions to the above means, there have been great research efforts to develop more understanding of the competencies and skills companies go after. For example, Benhart and Shaurtte (2014) have explained the process used to obtain industry input in establishing undergraduate educational competencies for Purdue University and how these competencies were used to reform the Building Construction Management curriculum. One of the conclusions of for this study included lessons learned in managing curriculum change that could prove beneficial for other CM programs. From an industry perspective, Dainty et al. (2004) was able to identify 12 core behavioral competencies that support effective construction project management performance. Team-leadership and composure were identified among others as the most predictive for excellent performance. Also, in a study by Slattery and Summer (2011), an assessment of leadership characteristics of construction professionals who are viewed as rising team members within their respective organizations, was performed. Another study by Gunderson and Glockner (2011) focused on the competencies required for construction superintendents according to the management level in companies including presidents, vise-presidents, and project managers. Another approach was done by Rojas (2013) who used focus group discussions and interviews to analyze 12 beneficial characteristics of construction project managers in electrical companies.

One of the very comprehensive efforts in that direction is a study by Ahn, Annie, Kwon (2012) to examine U.S. construction industry perceptions about key competencies for construction graduates. This study utilized a survey of employers from over 100 construction companies in the eastern United States. In this study, the authors obtained rating from recruiters in construction companies for 14 different competencies including problem-solving skills, interpersonal skills, leadership, collaborative skills, safety issues, interdisciplinary application, technical skills, computer skills, estimating / scheduling skills, communication, and environmental awareness and used this data to rank these competencies and highlight their significance. In addition, Bhattacharjee et al. (2013) evaluated the skills and knowledge that are expected to be gained by CM graduates from both the by both the employers and students at the CM program of Ball State University. A step further was proposed by Also, Pellicer et al. (2012), which is in the form of a method to improve graduate degree programs in the CM area based on selected requirements and market demands collected by surveys. This method involved two indices, the completeness and adequacy indices. However, there is no enough research on collecting data directly from the job description announced by the companies and are based on discussions between the human resources of the companies and the project management teams.

Therefore, the objective of this paper is therefore to present research work aiming at identifying competencies that construction companies seek when recruiting CM graduates for certain jobs. The methodology adopts using semantic analysis or data mining of the wording in job descriptions used in the actual job advertisements used by construction companies. Machine learning tools were used to achieve the required semantic analysis and one example for recruitment of a cost estimator for mechanical and electrical specialty contractors is used to illustrate how the data were collected, processed and interpreted. The final section of the paper is a discussion on usability of this research work and suggestion for future work.

Methodology

Figure 1 shows a diagram for the steps that are followed to meet the objective of the proposed research. The first step for this research work is data collection. Once the target job or designation and the type of companies under consideration are identified, the recruitment sites are explored to collect most recent job advertisements posted for the target job and the type of company identified. In the example used in the paper, the search data collected were for the job of a cost estimator with 1 to 3 years of experience that are recruited by electrical and mechanical specialty contractors. The search did not cover a specific region or state and was done over a period of 2 months. These data were organized data in spread sheet files that can be used for analysis and further processing.

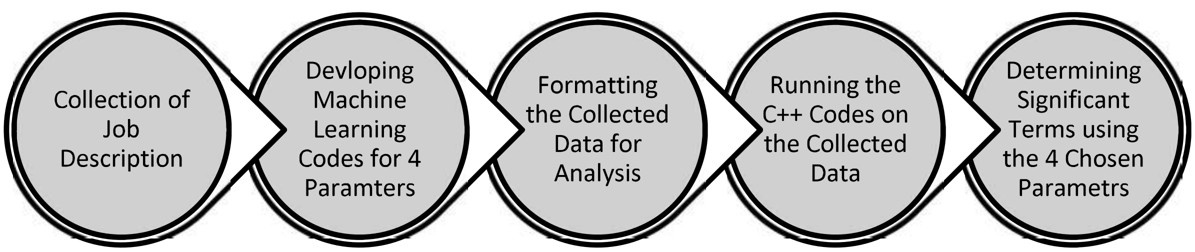

Figure 1: Research Process Diagram

The second step is to develop the machine learning codes that digs into the words and phrases, which is the main idea behind this research work. The codes were written to read all words included in the job descriptions and identify the most critical words and expressions that are significantly represented in the text. In order to achieve this goal and to increase the accuracy of the results, the text collected was separated into words and then returned to their roots. As a result, a C++ code was developed to do the following:

1. Separate the text included in the job description into a set of words,
2. Eliminate capitalization in the beginning of each word,
3. Return each word to its root by removing plurals, possessive characters, tenses, etc., and
4. Create a table in which each job description (meaning each company's ad) represents a row and the rooted words are expressed in the columns.

A similar code was used by the authors before for identifying significant words in construction legal cases (Mahfouz and Kandil 2010). At this point, it's important to refer to a mechanism to relate the significance of each word within the job description text. To explain this further, some words could be mentioned repeatedly but with no significant to the overall context. Literature in the areas of Natural Language Processing, and Text Mining highlights four weighing mechanisms that are used to achieve this target, namely Term Frequency (tf) Equation (1), Logarithmic Term Frequency (ltf) Equation (2), Augmented Term Frequency (atf) Equation (3), and Term Frequency Inverse Document Frequency (tf.idf) Equation (4). (Mahfouz and Kandil 2012).

$$tf_{i,d} = tf_{i,d} \quad (1)$$

Where tf is the term frequency or count of the number the term appears in the text. The term is denoted by i and the company is denoted by d

$$ltf_{i,d} = 1 + \log\left(tf_{i,d}\right) ;\ tf_{i,d} > 0 \quad (2)$$

Where ltf is the logarithm term frequency

$$atf_{i,d} = 0.5 + \frac{0.5 \times tf_{i,d}}{\max_t(tf_{i,d})} \quad (3)$$

Where atf is the augmented term frequency and $\max_t$ is the maximum number of term frequencies

$$tf.idf_{i,d} = (1 + \log\left(tf_{i,d}\right)) \times \log\left(\frac{N}{df_i}\right) \text{ if } tf_{i,d} > 0 \quad (4)$$

Where tf.idf is the term frequency invers document frequency, N is the number of samples and the df_i is the number if documents that the term is mentioned in

Therefore, the second component of the C++ algorithm is developed to calculate the mentioned four weighing mechanism. The Term Frequency (tf) is a count of the number of repetitions of each word within each jo description. However, this count might not be meaningful in some cases as some of the most repeated words may not provide insight or knowledge about the competencies that are the subject of this research. Therefore, the other three parameters are used to address this issue. The Logarithmic Term Frequency (ltf) uses (Log10) for every term frequency, which is greater than zero, thus lowering the significance of highly repeated words. For example (Log10) of 100 is 2 while (Log10) of 1 is 0. Also, the Augmented Term Frequency (atf) relates the frequency of a word repetition within the job description to the maximum repetition of that word in all job descriptions. Thus, if a word is repeated within all documents and in high frequencies, its weight is lowered, which means eliminating any bias to its significance. Moreover, Term Frequency Inverse Document Frequency (tf.idf) employs the logarithmic concept while normalizing the weight of each word to inverse of the number of job description in which it existed.

Therefore, after collecting the data and organizing the data files in spread sheet formats, the codes developed are used to calculate the 4 parameters for the most significant words and expressions. The outcome is in the form of 4 tables where the words are organized in rows and the job descriptions are in columns. Each table is for one of the 4 calculated parameters. For the first table that includes number of repetitions, a total column is added, and the rows are sorted based on total repetitions. For the other 3 tables including the three other verification parameters, an average column is added, and rows are also sorted based on highest averages. The words that show as top list of the first spread sheets are then checked against their averages in the other 3 sheets that work as verification for the results obtained in the first one. If the averages of the calculated parameters in the three tables are also high, this indicates significance of the used term or word across all job descriptions analyzed.

The last step of this research is done through filtering the words in the columns to focus on words related to competencies of CM graduates like take-off, reading drawings, using scales, using estimating software, and so on. The tables can be exported to spread sheets where sorting can be used to rank the highest frequency of repetition along with the other 3 parameters to identify the competencies or skills that the companies are seeking for that specific job.

Results

As explained in the methodology section, the results of this research are in the form of tables that show the words in rows and the calculated the analyzed job descriptions in columns. For example, in the sample analysis presented in this paper, which looks for competencies required by MEP contractors for a recent graduate (1-3 years' experience) to fill an estimator position, covered 50 job descriptions, which means 50 columns and each column will include the calculated parameter for the term in the row. Each of the 4 table includes results for one of the calculated parameters. Samples of the tables described here are shown below in Tables 1, 2 and 3.

Table 1
Term Frequency Sample Results

Term	Company 1	Company 2	Company 3	Company 4	Company 5	Company 6	Company 7
bid	4	0	0	0	0	3	7
management	4	5	0	0	6	0	0
estimating	3	3	0	0	4	0	0
construction	3	0	4	4	7	5	0
skill	0	8	6	0	0	0	0
experience	0	6	3	0	4	4	0
project	0	5	5	0	15	8	3
required	0	4	0	0	3	3	0
must	0	3	0	0	0	0	0
excellent	0	3	0	0	0	0	0

Table 2
Weighted Term Frequency Sample Results

Term	Company 1	Company 2	Company 3	Company 4	Company 5	Company 6	Company 7
bid	1.60	0	0	0	0	1.47	1.84
management	1.60	1.69	0	0	1.77	0	0
estimating	1.47	1.47	0	0	1.60	0	0
construction	1.47	0	1.60	1.60	1.84	1.69	0
skill	0	1.90	1.77	0	0	0	0
experience	0	1.77	1.47	0	1.60	1.60	0
project	0	1.69	1.69	0	2.17	1.90	1.47
required	0	1.60	0	0	1.47	1.47	0
must	0	1.47	0	0	0	0	0
excellent	0	1.47	0	0	0	0	0

Table 3
tfid Frequency Sample Results

Term	Company 1	Company 2	Company 3	Company 4	Company 5	Company 6	Company 7
bid	0.96	0	0	0	0	0.88	1.11
management	1.16	1.23	0	0	1.29	0	0
estimating	0.56	0.56	0	0	0.60	0	0
construction	0.50	0	0.54	0.54	0.62	0.57	0
skill	0	1.48	1.38	0	0	0	0
experience	0	0.67	0.56	0	0.60	0.60	0
project	0	0.29	0.29	0	0.38	0.33	0.26
required	0	1.24	0	0	1.14	1.14	0
must	0	1.59	0	0	0	0	0
excellent	0	2.48	0	0	0	0	0

Tables 1, 2 and 3 show only sample of all the 50 job descriptions collected and analyzed. Also, the terms shown are samples where the full list included 180 terms that have shown in the various job descriptions.

As explained before, some of these terms are not significant for the context of competencies like “must”, “required” that show in Table 3 above. In the interpretation step, these rows are removed to focus on the terms relevant to this research. After deleting the irrelevant rows, an additional column can be generated to calculate the average for the whole row and then sorting can be done to identify the most significant terms.

For the estimator position taken as an example, the results show the various terms that are most significant and were also supported by high averages for the verification parameters shown in equations 2 to 4. These terms include review drawing, construction experience, bid, both mechanical and electrical systems pricing followed by another set showing a smaller number of repetitions including contract, team, specification, software. These results show that the mechanical and electrical subcontractors that have been included in the survey emphasize the competencies associated with these terms which are ability to review drawings, having field experience, preparation of bids and familiarity with pricing electrical and mechanical components. The results did not show significance for competencies associated with team work or communication with suppliers and vendors, which were expected by the authors for an estimator.

Discussion and Recommendations

This paper presents a new tool to identify the industry needs in a more systematic way through direct assessment of the job descriptions included in the advertisement for requirement. The significance of this research is determining the competencies that are sought by construction companies through an unbiased quantitative methodology that analyze the job descriptions by companies. However, the following points should be considered when dealing with the results presented above and for future work:

- The data collection for this case was done for both electrical and mechanical specialty contractor. Limiting the data collection to one company type can be more meaningful.
- The data collection for this case was done for specialty contractors and across USA. Data collection can be done for specific regions in USA or other countries to identify trends and hold comparisons between the expectations in different geographic areas.
- The sample presented in this paper is based on 50 job descriptions, which is an arbitrary number to test the work of the code on the collected data. More representative samples can help to provide more credibility to the results and the outcome of the analysis.

The results of this kind of analysis can be used in three ways. First, the academic construction management programs can understand more what the companies are looking for when recruiting and consider improvements or changes in the CM curriculum to reflect this understanding. Second, it can be used by the construction companies as references for the standard or most common skills and competencies that other companies look for when recruiting certain professionals at certain levels including the entry level. Third, it draws the attention of some construction companies to some of the competencies that are not showing in their job descriptions while being required or anticipated due to that nature of these jobs.

References

Ahn, Y. H., Annie, R. P., & Kwon, H. (2012). Key competencies for US construction graduates: Industry perspective. *Journal of Professional Issues in Engineering Education & Practice, 138*(2), 123-130.

Benhart, B. L., & Shaurette, M. (2014). Establishing new graduate competencies: Purdue University's construction management curriculum restructuring. *International Journal of Construction Education and Research, 10*(1), 19-38.

Bhattacharjee, S., Ghosh, S., Young-Corbett, D. E., & Fiori, C. M. (2013). Comparison of industry expectations and student perceptions of knowledge and skills required for construction career success. *International Journal of Construction Education and Research, 9*(1), 19-38.

Dainty, A. R., Cheng, M. I., & Moore, D. R. (2004). A competency-based performance model for construction project managers. *Construction Management and Economics, 22*(8), 877-886.

Gunderson, D. E., & Gloeckner, G. W. (2011). Superintendent competencies and attributes required for success: A national study comparing construction professionals' opinions. *International Journal of Construction Education and Research, 7*(4), 294-311.

Mahfouz, T. and Kandil, A. (2012). Litigation outcome prediction of differing site condition disputes through machine learning models. *Journal of Computing in Civil Engineering*, *26*(3), 298–308.

Mahfouz, T., Kandil, A. (2010). Construction lLegal decision support using support vector machine (SVM) (pp. 879-888). Banff, Canada: Proceedings of 2010 Construction Research Congress.

Pellicer, E., Yepes, V., & Ortega, A. J. (2012). Method for planning graduate programs in construction management. *Journal of Professional Issues in Engineering Education and Practice, 139*(1), 33-41.

Rojas, E. M. (2013). Identifying, recruiting, and retaining quality field supervisors and project managers in the electrical construction industry. *Journal of Management in Engineering, 29*(4), 424-434.

Rojas, W. J., & Kazanas, H. C. (2011). *Mastering the instructional design process: A systematic approach*. John Wiley & Sons.

Slattery, D. K., & Sumner, M. R. (2011). Leadership characteristics of rising stars in construction project management. *International Journal of Construction Education and Research, 7*(3), 159-174.