The published version is found in the ASCE Library here: **https://doi.org/10.1061/9780784482421.051**
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

Automated Item Matching and Pricing (IMP) for Wood Building Elements to Support BIM-Based Wood Construction Cost Estimation

Temitope Akanbi,[1] Jiansong Zhang, Ph.D., A.M.ASCE,[2] and Yong-Cheol Lee, Ph.D. A.M.ASCE. [3]

[1]Graduate Student, Automation and Intelligent Construction (AutoIC) Lab, Purdue University, West Lafayette, IN 47907; e-mail: takanbi@purdue.edu

[2]Assistant Professor, Automation and Intelligent Construction (AutoIC) Lab, Purdue University, West Lafayette, IN 47907; e-mail: zhan3062@purdue.edu

[3]Assistant Professor, Department of Construction Management, Louisiana State University, Baton Rouge, LA 70803; e-mail: yclee@lsu.edu

ABSTRACT

A major gap in the automation of construction cost estimation is the need of manual inputs to complete cost estimation processes. To address this gap, the authors propose a new method for matching wood building elements from a Building Information Modeling (BIM)-based design to cost data entries in a cost database. The proposed method uses a java constructor and HashMap to create objects, and store and retrieve the created values of the objects. Term matching and natural language processing (NLP) techniques are used in the method to match items from a design model and automatically extract their unit costs from a cost database. These unit costs retrieved are then used in generating the cost estimates. The proposed method was tested on estimating a wood construction model retrieved online. A cost estimate was successfully generated. Comparison of the experimental results with results from the state-of-the-art commercial software showed that the algorithms developed from the proposed method reduced the manual inputs required in generating wood construction cost estimates.

INTRODUCTION & BACKGROUND

Cost estimation is central to the realization of a successful construction project (Yu et al. 2006). According to Staub-French et al. (2003), one of the fundamental challenges in conducting cost estimation is the expertise required in selecting the appropriate cost parameters, which would ultimately affect the construction cost. Shane et al. (2009) argues that besides complexities in the engineering and construction design of a project, an estimator's bias can greatly influence a construction project cost. Besides this lack of consistency in the construction cost estimates,

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

manual cost estimation process is a tedious task subject to human errors (Samphaongoen 2009). Lee et al. (2014) stated that despite the automation of the quantity takeoff (QTO) processes, most commercial software programs still require estimators to manually match materials of building elements to work items to complete the cost estimation process. Elfaki et al. (2014) reviewed twenty-seven intelligent techniques in cost estimation over a ten-year period and concluded that there are still gaps in the automation of cost estimation in spite of the existing development, especially in the lack of an intelligent system that addresses the human dependability issues identified from the analysis of these techniques. In recent years, multiple research efforts have been devoted to successfully enable an automated QTO. For example, Akanbi and Zhang (2017) developed a method that automatically reads, and extracts quantities of wood building objects by leveraging the fundamental geometric representations of the components in an IFC model. Mandava and Zhang (2016) developed an automated IFC-based QTO method that successfully extracted the needed quantities of bridge components from IFC-based BIMs by leveraging the Cartesian points of the models. Choi et al. (2015) developed a method based on schematic QTO that extracts quantities from BIM architectural elements' data and utilizes ratio formulas to compute the quantity of materials. However, the matching of building elements with cost items are still mainly performed. To address this research gap in matching building elements with cost items, the authors proposed a new method for developing automated item matching and pricing (IMP) algorithm using natural language processing (NLP) techniques. The proposed method includes a series of four steps to develop an automated algorithm for IMP, to match building elements from a Building Information Modeling (BIM) design to cost data entries in a cost database. This reduces the need of manual inputs to complete cost estimation processes. The detailed steps are introduced in the next section.

PROPOSED METHOD

The proposed item matching and pricing (IMP) algorithm development method consists of four steps for cost estimation (Figure 1): Step (1) - constructor and HashMap development – define a constructor and its arguments (i.e., parameters) to use in creating new objects, and create a HashMap (i.e., data framework) to store and retrieve values of targeted objects. The created objects represent materials and therefore will be referred to as material objects hereafter. For example, "*ProductsCatalogue (material, thickness, cost)*" is a java constructor for *ProductsCatalogue* with three arguments - *material, thickness and cost,* which are of *String, double,* and *double* data types in Java, respectively. To store the created materials, the *map.put* method of HashMap is utilized. For example, *map.put (1, material)* – add *"material"* to the HashMap at Index 1. Step (2) - item matching algorithm development – develop the algorithm for automatically matching items between building objects from the BIM design and cost items from the cost database, and for automatically extracting the unit cost of materials from the cost database for each building component (e.g., wall and floor). The item matching algorithm is developed using natural language processing (NLP) techniques; NLP techniques enable computers to understand and process natural language text (or speech) in a human-like manner

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

(Cherpas 1992). Step (3) - cost estimate computation – the retrieved unit cost from Step (2) is used to compute the cost estimate. Step (4) - evaluation – evaluating the proposed method by comparing a cost estimate based on the developed algorithm with manually created estimate using existing BIM software. The proposed method is expected to reduce the manual efforts needed to match materials from building design with the appropriate cost components. Therefore, this method helps address the human input issues pointed out by Elfaki et al. (2014) and Lee et al. (2014).

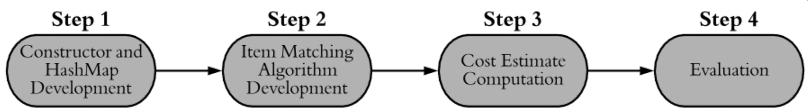


Figure 1. Proposed IMP algorithm development method.

EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method was tested in an experiment of estimating the cost of the floor and wall components of a wood structure. The implementation details are described as follows:

Step 1- Constructor and HashMap Development. In this step, a constructor and a HashMap were developed in Java. The constructor was used to create material objects. The HashMap was used to store or retrieve the newly created objects. A defined constructor has one or more parameters as its arguments. Figure 2. shows an example constructor named *"ProductsCatalogue"* that has three parameters - "material," "thickness," and "cost." A HashMap was then used to store the created material objects. In the HashMap, unique identifiers were assigned to each object. When the objects were accessed through the unique identifiers, the values associated with the objects were retrieved. For example, in Figure 2, *('Gypsum Wall Board, 0.05208 ->, $1.63/sf)* is a material object created in *"ProductsCatalogue,"* depicting a "gypsum wall board" with a thickness of 0.05208 (5/8"), the unit cost of which is $1.63/sf A HashMap uses the "*put*" and "*get*" methods to store or retrieve objects. Each material object was stored by calling the *"map.put()"* method and the values were retrieved by calling the *"map.getKey()"* method.

Step 2 - Item Matching Algorithm Development. In this step, an algorithm was developed for automated matching between extracted building elements from BIM and the cost items in the cost database. NLP techniques were used to support the matching, including tokenization and morphological analysis. Tokenization is a process of breaking a piece of text (e.g., the search string) into smaller units (i.e., words, symbols, or punctuations) which are referred to as tokens (Fares et al. 2013). Detailed steps of the developed algorithm are described as follows.

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

The algorithms includes 10 processes (Figure 3): *Process 1* initializes a search string (i.e., name of materials) based on material layer information extracted from the BIM. *Process 2* tokenizes the search string from *Process 1.* Tokenization helps enhance the efficiency of a search (Fares et al. 2013). For example, the text string 'Structure, Wood Joist/Rafter Layer' becomes five tokens after the tokenization: *Structure*, *Wood*, *Joist*, *Rafter* and *Layer*. This step helps improve the robustness of accommodating different BIM authoring platforms' proprietary naming conventions of building components. In *Process 3,* morphological analysis is conducted for the tokens to help match all forms of the token with the databases' lowercased names, e.g., if 'Joist,' or 'JOIST' is the token in the search string, the algorithm will execute a search for 'joist.' In *Process 4,* synonym tokens of the search token are generated; creation of synonym tokens ensures that while executing searches for a term, its synonyms are also searched. For example, while searching for 'Joist,' synonym terms such as 'beam' are also searched. *Process 5* uses the resulting terms from *Process 4* to select the appropriate material from the cost database, at this point, a conditional statement *Decision 1* is met.

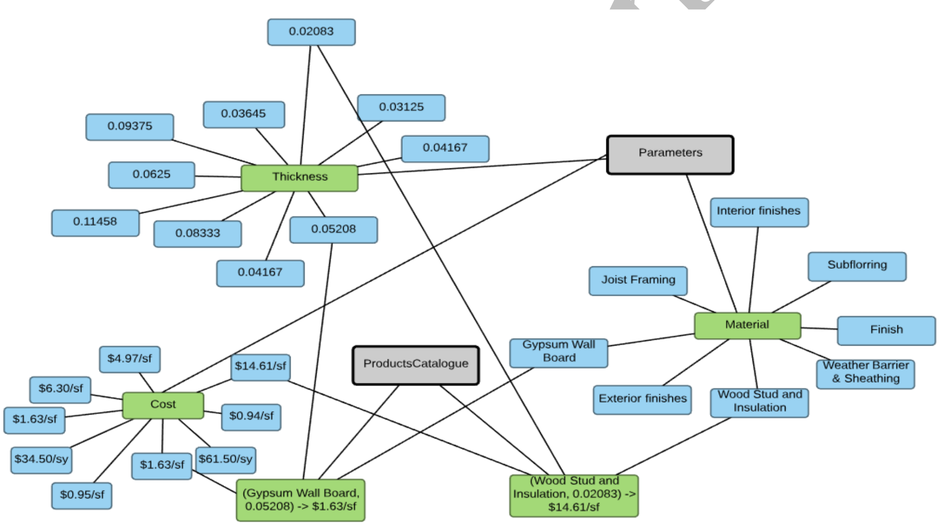


Figure 2. Map structure of a sample constructor and HashMap.

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

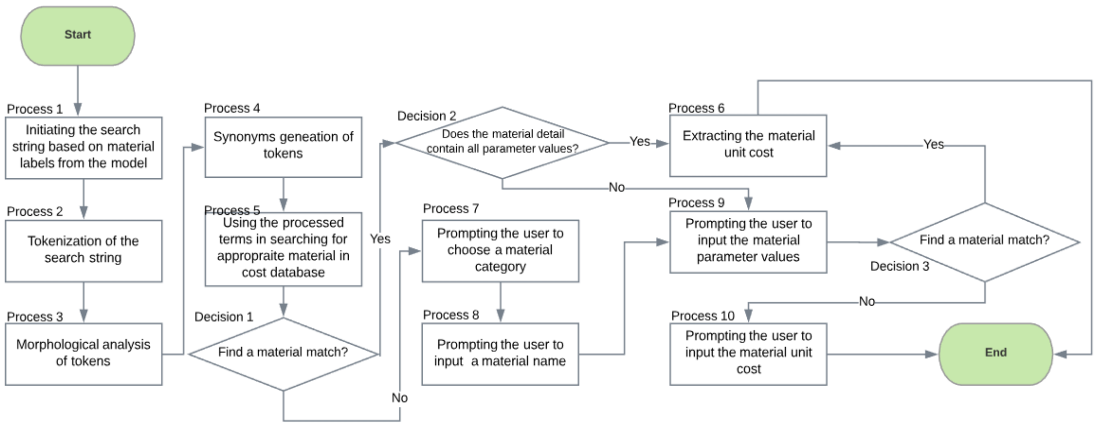


Figure 3. Flow chart of the developed item matching algorithm.

Decision 1 checks if there is a material match found in the cost database. If there is a material match, the algorithm proceeds to a new conditional statement (*Decision 2)*, otherwise, the algorithm prompts the user for information. *Decision 2* checks if the material detail contains all needed parameter values for picking a unit price. BIM has different level of development (LOD) specifications. LOD provides a reference that defines the level of details in BIM (Choi et al. 2015). According to the American Institute of Architects (AIA), LOD 400 BIM is suitable for construction, models at this level include assembly details such as size, shape, location, quantity, orientation, fabrication, and installation information (BIMForum 2013). If distinguishing parameter values exist, *Process 6* uses these parameter values to extract the unit cost of the material and the item matching and pricing process is finished. For example, the parameter used in selecting the unit cost of a gypsum board is the thickness of the gypsum board. For wall studs and insulation, the parameters are thickness, height and spacing. If, however, there is no material match found in the cost database in *Decision 1*, the algorithm proceeds to *Processes 7, 8,* and *9*, which prompts the user to input the material category (in the database, materials are categorized based on the component they belong to; e.g., "gypsum wall board" would be categorized under the wall component), material name, and the material parameters, respectively. If the BIM misses certain material details at *Decision 2*, the algorithm proceeds to *Process 9* as well to prompt the user to input the material parameter values. At this point the database is searched again, if a material match is still not found (*Decision 3*), *Process 10* will prompt the user to manually input the material unit costs. However, if sufficient design details exist in the input BIM, none of the manual processes would be activated and the item matching is fully automated.

Step 3- Cost Estimate Computation. Similar to industry practice in construction cost estimation, the total cost of each wood building element assembly is made up of its cost components (i.e., the materials that make up the assembly grouped for cost purposes). For

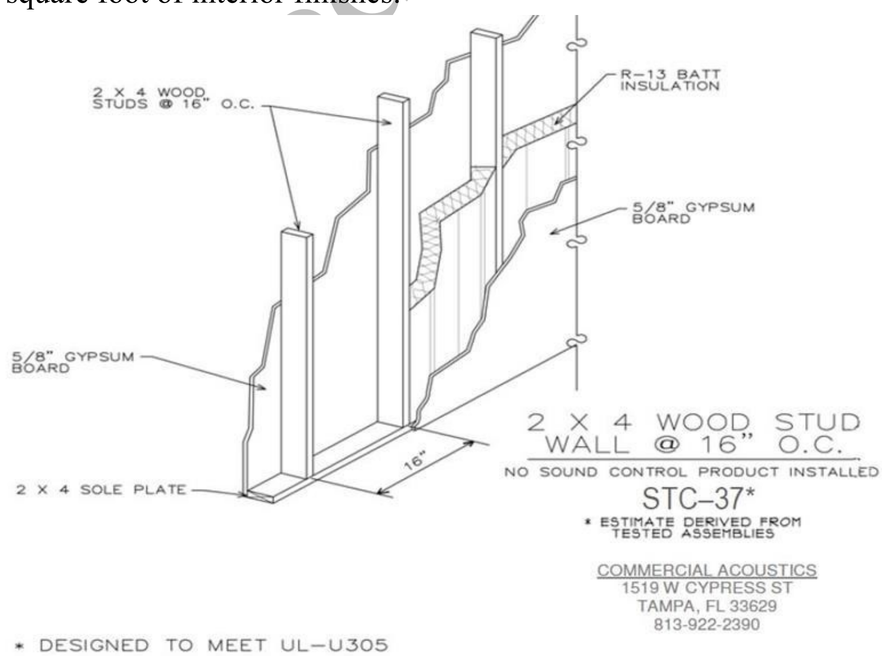
The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

example, a wall assembly is made up of five cost components following a similar naming convention as in the Industry Foundation Classes (IFC) model: (1) gypsum board, (2) wood stud and insulation, (3) weather barrier & sheathing, (4) exterior finishes, and (5) interior finishes. Whereas the floor components are grouped into three cost components: (1) joist framing, (2) subflooring, and (3) finish.

In the database, costs were stored as unit costs. Each unit cost represents the cost to install one unit of the component (i.e., including all labor, material, and equipment costs). Components may have different units of measures, which dictate the quantity to be multiplied with unit cost for computing the cost estimate. For example, gypsum board uses a unit of measure of square foot (S.F.) whereas carpet uses a unit of measure of square yard (S.Y.). Therefore, to compute the cost estimates of gypsum board and carpet, the unit cost per S.F. of gypsum board and the unit cost per S.Y. of carpet were used to multiply the corresponding quantities, namely, net area of the wall in S.F. and net area of the floor in S.Y., respectively.

To illustrate these processes in computing the cost estimate, a subcomponent of wall (wood studs and insulation) is used as an example for detailed explanations below.

Figure 4. shows an example wall with "2x4 wood stud" including details about its subcomponents. The "*MaterialLayerSet"* that follows an IFC naming convention of a typical wall consists of two layers of 'Gypsum Wall Board,' one layer of 'Structure, Wood Joist/Rafter Layer, Batt Insulation,' and another two layers of 'Gypsum Wall Board.' Hence, the following cost variables would be utilized in estimating the costs of the components: (1) unit cost per square foot of gypsum board; (2) unit cost per square foot of wood studs and insulation; and (3) unit cost per square foot of interior finishes.



The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

Figure 4. Material layers of a wall (Commercial Acoustics 2017).

Wood studs and Insulation – Three parameters were used in selecting the unit costs of wood studs and insulation: (1) the thickness of the layer, (2) the height of the wall, and (3) the spacing of the studs. The first and second parameters, thickness of the material layer and the height of the wall were extracted during the QTO process. The third parameter, the spacing of the wood studs could only be retrieved from a Level of Detail (LOD) 400 BIM and above. In this paper, the BIM used was at LOD 300 – the model elements were represented in terms of quantity, size, shape and orientation within the model. Therefore, in retrieving the spacing of wood studs, *Decision 1* in the developed algorithm (Figure 3) did not find a material match. The algorithm then proceeded to *Process 7* as illustrated in Figure 3. The system prompted the user to choose a material category (category 1- all, category 2 - wall, category 3 - floor). Next, *Process 8* prompted the user to input a material name (wood). Next, *Process 9* prompted the user to input a value (16" O.C. - sixteen inches on center). All other needed parameters were automatically found. At this point, all parameters to retrieve the unit cost of wood studs and insulation had been completed. The systems found a material match (*Decision 3)* from the database. The unit cost of the material was retrieved. The unit of measure for wood studs and insulation was square foot (S.F.). Hence, the retrieved unit cost per square foot of wood studs and insulation coupled with square foot of the area covered by the wood studs and insulation were utilized in computing the cost estimate (Figure 5).

Step 4- Evaluation. A comparison was made between the cost estimate using the authors' method and the cost estimates by a professional estimator based in Detroit, Michigan. The comparison were conducted in two dimensions: (1) estimation results, and (2) needed manual inputs.

For Estimation Results: There was a 13% difference in cost estimates between the experimental results using the proposed method and that prepared by an estimator, which was found to be caused by the different cost data source used. While the unit cost in the authors' database were based on U.S. national averages, the professional estimator's prices were based on their own historical costs data, which were affected by several factors such as availability of material, availability of labor, and labor productivity.

For Manual Inputs: The processes based on the state-of-the-art commercial software required the estimator to classify cost items manually, whereas the developed IMP method and algorithm extracted the cost items by leveraging the material characteristics of each component in an IFC file, automatically.

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

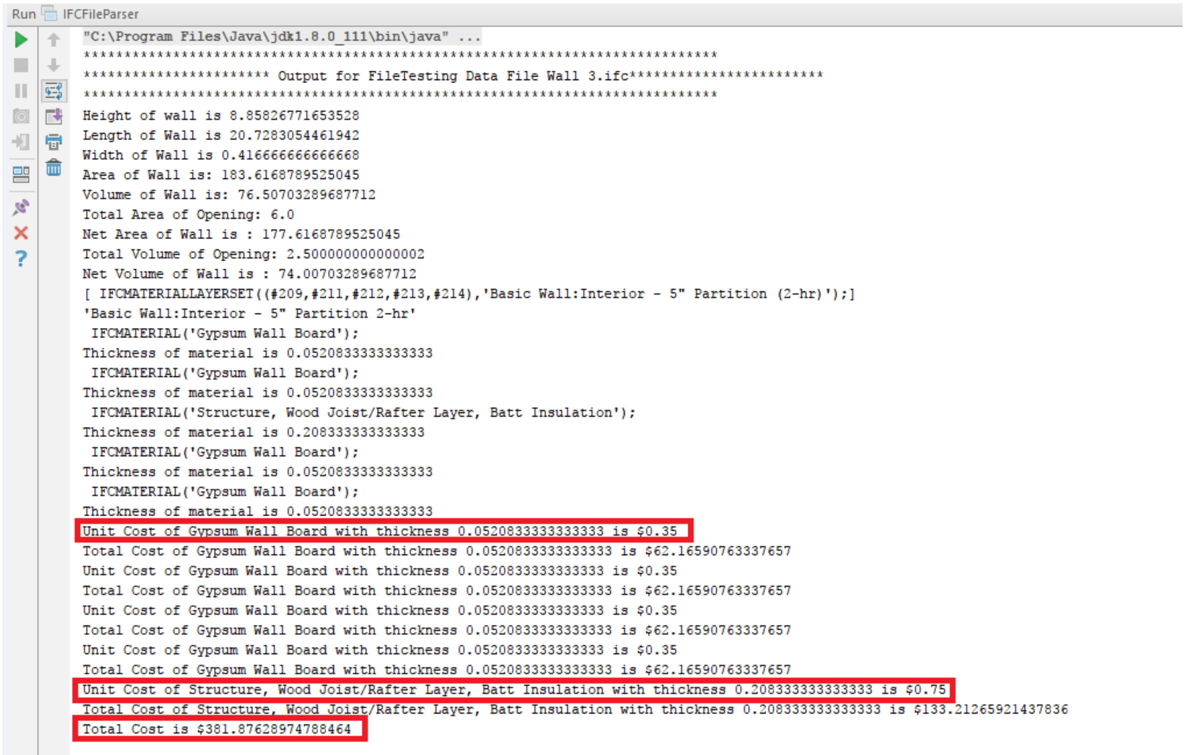


Figure 5. Experimental cost estimating results (partial) using proposed method and corresponding algorithm.

CONCLUSIONS, LIMITATIONS AND FUTURE WORK

In this study, the authors developed an automated item matching and pricing method to reduce manual inputs needed from estimators in BIM-based cost estimation. The proposed method computes the cost estimate by automatically retrieving units costs from a linked cost database, using an algorithm based on term-based match and natural language processing (NLP) techniques. The proposed method was tested on a wood construction model retrieved online. The experimental results showed the proposed method successfully computed the cost estimates of the wood components and reduced the need of manual input in matching building components with cost items, when comparing to an estimates generated using the state-of-the-art commercial software by a professional estimator. The proposed method provides a foundation for automatically matching design elements with cost items in a broad range of construction types (e.g., wood, steel, concrete) using IFC-based BIMs.

One main limitation is acknowledged: the search strings were developed using the known naming conventions of few selected BIM authoring platforms, which may encounter problems when used with unseen BIM authoring platforms. In the future work, the item matching algorithms will be expanded to support a more robust search by incorporating a more powerful matching mechanism and more search strings compatible with a variety of BIM authoring tools.

ACKNOWLEDGMENTS

The published version is found in the ASCE Library here: https://doi.org/10.1061/9780784482421.051
Akanbi, T., Zhang, J., and Lee, Y.C. (2019). "Automated item matching and pricing (IMP) for wood building elements to support bim-based wood construction cost estimation." Proc., 2019 ASCE International Conference on Computing in Civil Engineering, ASCE, Reston, VA, 402-409.

The author would like to thank the National Science Foundation (NSF). This material is based on work supported by the NSF under Grant No. 1745374. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

REFERENCES

Akanbi, T., and Zhang, J. (2017). "Automated wood construction cost estimation" *Proc., 2017 ASCE Intl. Workshop on Comput. in Civ. Eng.*, ASCE, Reston, VA., 141-148.

Cherpas, C. (1992). "Natural language processing, pragmatics, and verbal behavior." *Analysis of Verbal Behavior*, 10,135-147.

Choi, J., Kim, H., and Kim, I. (2015). "Open BIM-based quantity takeoff system for schematic estimation of building frame in early design stage." *Journal of Computational Design and Engineering*, 2(2015), 16-25.

Elfaki, A. O., Alatawi, S., and Abushandi, E. (2014). "Using intelligent techniques in Construction Project Cost Estimation." *Advances in Civil Engineering*, 2014(107926).

Fares, M., Oepen, S., and Zhang, Y. (2015). "Machine learning for high-quality tokenization replicating variable tokenization schemes." *Proc., 14th International Conference, CICLing*, Samos, Greece, 231-244.

Lee, S., Kim, K., and Yu, J. (2014). "BIM and ontology-based approach for building cost estimation." *Automation in Construction,* 41(2014), 96-105.

Mandava, B., and Zhang, J. (2016). "A new automated quantity takeoff method for BIM-based bridge designs." *Proc., CIB W78*, Conseil International du Bâtiment (CIB), Rotterdam, The Netherlands.

Samphaongoen P. (2010). "A visual approach to construction cost estimating." Master's Theses. Marquette University, Milwaukee, Wisconsin. Paper 28 <https://epublications.marquette.edu/theses_open/28> (October 30, 2018).

Shane, J., Molenaar, K., Anderson, S., and Schexnayder, C. (2009). "Construction project cost escalation factors." *Journal of Management in Engineering*, 25(4), 221-229.

Staub-French, S., Fischer, M., Kunz, J., and Paulson, B. (2003). "A generic feature-driven activity based cost estimation process." *Advanced Engineering Informatics,* 17(2003), 96-105.

Yu, W., Lai, C., and Lee, W. (2006). "A WICE approach to real-time construction cost estimation." *Automation in Construction,* 15(2006), 12-19.