

A network-based framework for dynamic linkage of unstructured data to BIM: supporting predictive analysis in work order management

Soroush Sobhkhiz¹, Tamer El-Diraby¹

¹ Centre for Information Systems in Infrastructure and Construction, Civil Engineering Department, University of Toronto, 35 St George St, Toronto, Canada

Abstract

Linking BIM to other data models is essential to establishing digital twins. Recently, ontologies have been used to establish the link between IFC and other data models. However, most of the data in digital twins are unstructured, for example, specifications, reports, and other communication. These types of data dynamically change and incorporate a wide range of concepts with complex relationships. It is difficult to develop and maintain an ontological representation for such forms of data. This research work explores the use of concept networks as means to link BIM to unstructured data. Using topic modeling tools, a set of key terms are extracted from documents. The relationships between the key terms are investigated and a network of these key terms is established. This approach is illustrated through a use-case example application to the work order management domain.

Keywords

BIM, Text, Unstructured Data, Network Analytics, Predictive Analysis, Work Order Management

1. Introduction

Unstructured data such as text and chats contain valuable knowledge. The processing and management of such data within a BIM environment is essential to developing digital twins. In this context, digital twin is not a simple digitization or a 3D model of a facility. It is a complete virtualization of facility data, work processes, stakeholder profiles that aim to create a digital replica to enable facility managers to study work and operations scenario in the virtual world before implementing them in the real world [1]. In the case of work orders, they contain valuable information regarding the overall building performance and occupant satisfaction. Studying patterns in work orders can enable the development of business intelligence tools, particularly predictive analysis [2]. We can study and model expected deteriorations, the timing of critical levels of performance, the expected work durations, and the expected maintenance costs.

However, currently, work orders are generated and managed inconsistently and it is difficult to derive the hidden and valuable knowledge within them. More importantly, a comprehensive analysis that can take work order data and other data sources into account, is practically infeasible. This is mainly because there is practically no link between work order data and structured data sources such as BIM.

We present here an approach to link BIM to work order data. Work order contents tend to be unstructured data describing the performance situation and the needed action [3]. They also contain structured data, particularly IFC-based data such as location and facility specifications. As a result, automating the analysis and processing of work orders requires finding pathways of linking structured and unstructured data domains.

LDAC 2022: 10th Linked Data in Architecture and Construction Workshop, May 29, 2022, Hersonissos, Greece

EMAIL: s.sobhkhiz@mail.utoronto.ca (S. Sobhkhiz); tamer@ecf.utoronto.ca (T. El-Diraby)

ORCID: 0000-0002-3462-1396 (S. Sobhkhiz); 0000-0001-6446-9199 (T. El-Diraby)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Typical linkage approaches in the BIM domain have focused on the use of ontologies. Such approach provides stable and reliable correspondence between ifcOWL, or other building ontologies, and established (and common) data models of other related domains, such as energy analysis, performance management, etc. [4]. While this top-down and structured approach provide reliable and scalable means for data linkage, it might not be flexible enough to capture specific contexts of work. Especially, when dealing with concepts that dynamically change over time. Although there has been some efforts to capture temporal changes, because of the needed overhead to update ontologies, it is hard to capture the evolution of knowledge in the domain [5].

We need to establish means to help formalize the conceptualization in a text source in a manner that enables clear linkages to IFC, and, at the same time, be model-independent to enable dynamic and contextualized capture of knowledge contained in unstructured data corpus. By formalizing both the structured (IFC-based data) and unstructured data of a work order, we can automate many of the processes in managing work orders. Comparing data of work orders can help us discover clusters or similarities between work orders, which can be used to predict future work orders. For instance, if we have historical records of a specific object, say an air handler, then we can perform predictive analysis to estimate the possibility of a new work order being submitted in close future. We can perform this analysis in such a way that provides planning insights for managers by informing them of the overall possible impacts (such as expected down time, or cost) of the work order.

In this paper, we aim to investigate an alternative bottom-up solution to the problem of linking work order text to IFC concepts. Specifically, the objective of this paper is to explore the potentials of linking IFC classes to network concepts obtained from textual data such as work orders. The motivation behind this approach is two-fold. First, modeling a text into a network is different from pushing it into pre-defined concept models (e.g., ontologies) in that the concepts are data-driven. Second, a network model is flexible enough to dynamically adapt to the changes in data and therefore, capture its evolution.

It should be noted that this paper is simply an exploration of an alternative possible solution, and that the proposed methodology has not yet matured for practical applications. Nevertheless, we provided a simple prototype use-case to showcase how the approach can potentially be used in practice. Future research will provide more details and implementations.

2. Overview of the proposed approach

Conventional linked-data solutions are based on static representations and do not support capturing the evolution of the knowledge (dynamic changes in concepts and relationships) [6,7]. We argue that the solution should not be a top-down expert-based method where we define the standard/model for data correspondence. Rather, it should be a data-driven bottom-up method where the patterns are dynamically obtained and used for identifying relationships.

As for the case of work order data, we need an unsupervised method for extracting concepts and relationships from textual data. As a result, concept networks can potentially provide a solution. Here, a network refers to a directed graph of interlinked web of concepts that represent conceptual connectivity. With network analysis, we can extract the key concepts and relationships in a set of data, without imposing any restrictions (i.e., having to define the concepts or relationships). Further, the analysis can be dynamic, and the concepts and relationships can be updated over time when more data is available [8]. In the following section, we explain a proposed approach in more detail, take the following set of work order components as an example:

- “Door hardware issue”
- “Replace Door handle”
- “Need a locksmith in room GB329”
- “Rekey | Room 215”
- “Need fob access for contractor”
- “Cylinder issue in room 132, too much noise”

The examples above are extracted from the work order database of the University of Toronto Facilities. Conceptually, all of them relate to the object/concept ‘Door’. In order to make the link between these work orders and BIM objects, we propose the following steps:

1. Establish standardized links for obvious cases: develop correspondence lists for clearly linked terms and IFC concepts. For instance, the word ‘door’ always refers to the concept ‘ifcDoor’.
2. Transfer text corpus into concept networks.
3. Use data analytics to discover the cluster of other concepts that should be also linked to IFC concepts. For instance, we do not link concepts such as ‘lock’ or ‘cylinder’ to ifcDoor. Instead, we develop a concept network, where we find that these concepts are always associated with the key term ‘door’. With this, we use a bottom-up, no-model approach to deduce that a linkage should be made between ifcDoor and ‘lock’. In other words, we use text mining to enrich the simple, obvious, static relationship “door-corresponds-to-ifcDoor”. The anchor node “door” can be used to find additional semantically relevant key terms to be linked to the IFC concept.

The richness, diversity and the evolution of key terms (linked to an IFC) open the door for unsupervised learning, where we can discover clusters of work orders. This is the foundation for pattern discovery and predictive analysis.

3. Association Rule Mining and Network analysis

Figure 1 shows the process of establishing a concept network from a text corpus (work orders). The process initiates by pre-processing the text data so that it can be clean enough for analysis. For instance, punctuations and stop-words are removed and words are lowercased.

Next, we perform association rule mining to identify relationships between the concepts in the work order data. The lift value of an association rule defines how the presence of one word increases the probability of the appearance of the other, see Equation (1) [9]. As a result, this relationship can provide meaningful links between concepts which we can use in our study. Here, if there is a strong relationship between the word ‘door’ and any other concept, it will be discovered. We can use these links and develop a graph network where concepts are linked to each other based on their lift, the larger the lift value, the stronger the relationship.

$$Lift(A \Rightarrow B) = \frac{Confidence(A \Rightarrow B)}{P(B)} = \frac{P(A \cup B)}{P(A)P(B)} \quad (1)$$

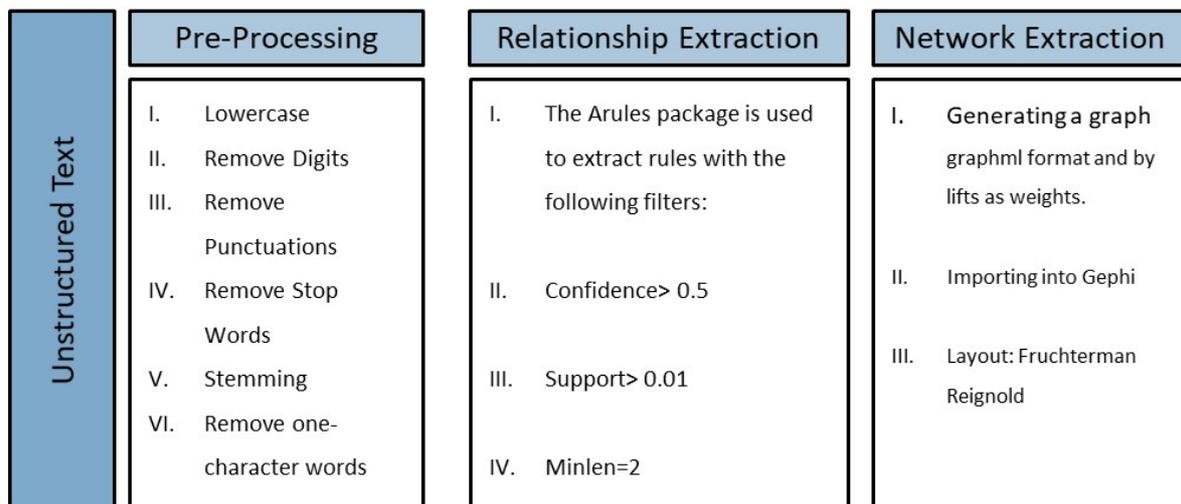


Figure 1: Network extraction workflow.

A lift of 1 for a rule, means that the probability of occurrence of the antecedent and consequent are independent of each other and no rule can be drawn from the two events (here the appearance of each word). If the lift value is greater than 1, it indicates the degree to which the two appearances are dependents on each other. Such a rule can be potentially useful for predicting the appearance of the consequent whenever the antecedent appears. For instance, consider the words “Cylinder”, “Crossbar”, “Hardware”, and “Handle”. The below relationships are identified between these words:

- {Hardware} => {Cylinder}: Lift value = 78

'ifcDoor' and other concepts as shown in Figure 3. Following step 1, assume that we can link the 'ifcDoor' to the concept 'door' in the work order network. This is the static obvious relationship. Step 2 is already performed with the results shown in Figure 2. We now perform step 3 and link the IFC data schema to the work order network. The result is shown in Figure 4.

Please note that this is a high-level prototype example simply to provide a generic understanding of the idea of a network-based solution. Detailed implementations will be provided in future research.

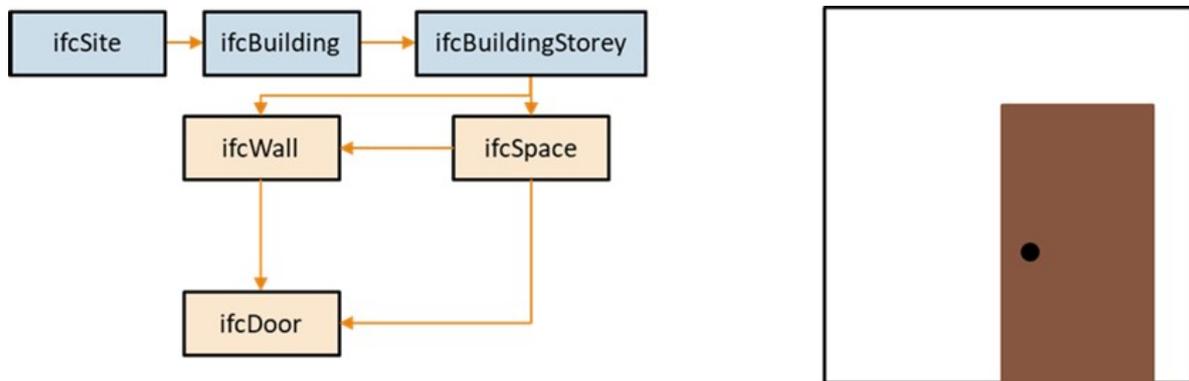


Figure 3: IFC relationships for representing a 'door'.

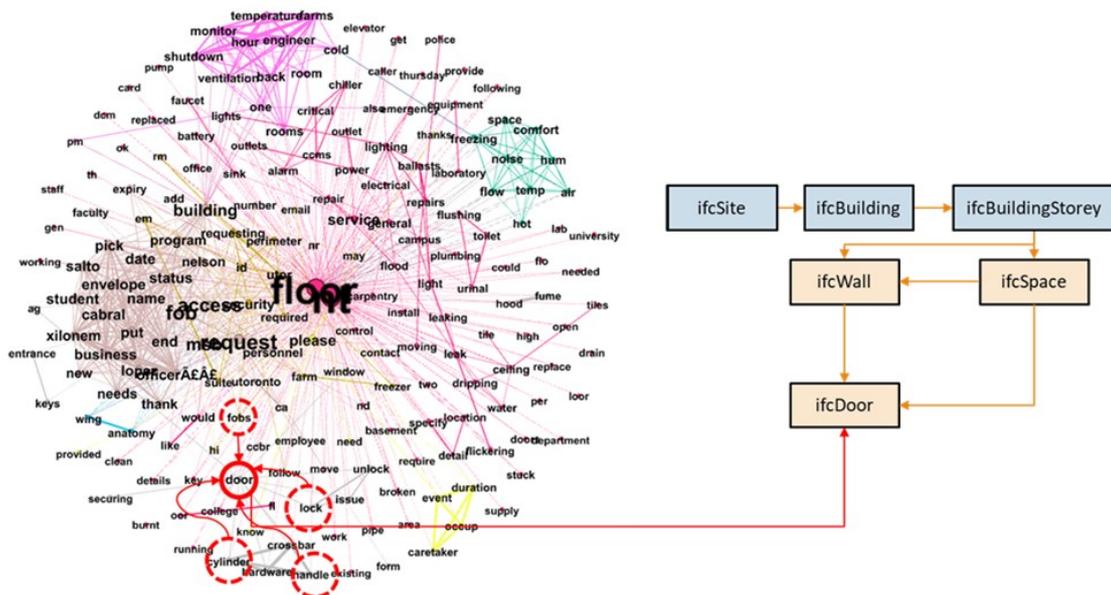


Figure 4: Linking work order concepts to the IFC data schema.

5. Discussion and Conclusion

The proposed approach combines the advantages of structured, expert-based analysis with bottom-up, data-driven analysis. A set of anchor key terms are linked to corresponding IFC concept at the start. The list of these linkages will tend to be stable and static. Parallel to this, text mining is used to find other key terms that are typically connected to the anchor node. Indirectly, these additional key terms are linked to the IFC concept. As the concept network evolves, the extended word cloud around an IFC concept morphs.

This process is more capable of capturing the changes in the relationships. If a new relationship emerges, the process can capture that because it is not limited to the expert intuition and relies on the actual patterns in the data. We can leverage the weights of the relationships to analyze how strong

relationships are. Therefore, the proposed approach does not result in firm connections and each connection comes with a certainty. For instance, a work order with the concept ‘cylinder’ and ‘hardware’ has a strong connection to the concept ‘door’ and therefore to the concept ‘ifcDoor’. But the connection of the same workorder to the concept ‘HVAC’ is weak. So, there is a higher probability that the work order relates to the door object in a BIM model.

The fundamental contribution of this approach is not linking a set of key terms to IFC concept. It is in doing so dynamically, and in an unsupervised manner that learns from data. The enrichment of IFC concept with a dynamically generated cloud of key terms enables the implementation of machine learning on work order data that are both structured (extracted from BIM) and unstructured (extracted from free text).

6. References

- [1] El-Diraby, Tamer, and Soroush Sobhkhiz, “The Building as a Platform: Predictive Digital Twinning”, *Buildings and Semantics: Data Models and Web Technologies for the Built Environment*, edited by Pieter Pauwels, CRC Press, 2022. To appear.
- [2] Lavy, Sarel, Nishaant Saxena, and Manish Dixit. "Effects of BIM and COBie database facility management on work order processing times: Case study." *Journal of Performance of Constructed Facilities* 33, no. 6 (2019): 04019069. doi: 10.1061/(ASCE)CF.1943-5509.0001333.
- [3] Dutta, Saptak, H. Burak Gunay, and Scott Bucking. "A method for extracting performance metrics using work-order data." *Science and Technology for the Built Environment* 26, no. 3 (2020): 414-425. doi: 10.1080/23744731.2019.1693208.
- [4] Lee, Do-Yeop, Hung-lin Chi, Jun Wang, Xiangyu Wang, and Chan-Sik Park. "A linked data system framework for sharing construction defect information using ontologies and BIM environments." *Automation in Construction* 68 (2016): 102-113. doi: 10.1016/j.autcon.2016.05.003.
- [5] Turk, Žiga. "Interoperability in construction–Mission impossible?." *Developments in the Built Environment* 4 (2020): 100018. doi: 10.1016/j.dibe.2020.100018.
- [6] Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data: The story so far." In *Semantic services, interoperability and web applications: emerging concepts*, pp. 205-227. IGI global, 2011. doi: 10.4018/978-1-60960-593-3.ch008.
- [7] Sobhkhiz, Soroush, Hossein Taghaddos, Mojtaba Rezvani, and Amir Mohammad Ramezaniapour. "Utilization of semantic web technologies to improve BIM-LCA applications." *Automation in Construction* 130 (2021): 103842. doi: 10.1016/j.autcon.2021.103842.
- [8] Aragao, Rodrigo, and Tamer E. El-Diraby. "Network analytics and social BIM for managing project unstructured data." *Automation in Construction* 122 (2021): 103512. doi: 10.1016/j.autcon.2020.103512.
- [9] McNicholas, Paul David, Thomas Brendan Murphy, and M. O'Regan. "Standardising the lift of an association rule." *Computational Statistics & Data Analysis* 52, no. 10 (2008): 4712-4721. doi: 10.1016/j.csda.2008.03.013.