www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

CONNECTING RESEARCH ON SEMANTIC ENRICHMENT OF BIM - REVIEW OF APPROACHES, METHODS AND POSSIBLE APPLICATIONS

SUBMITTED: October 2021
REVISED: December 2021
PUBLISHED: April 2022
GUEST EDITOR: Robert Amor
DOI: 10.36680/j.itcon.2022.020

Tanya Bloch
Faculty of Civil and Environmental Engineering, Technion
bloch@technion.ac.il

SUMMARY: *Semantic enrichment of BIM models is a process designed to add meaningful semantics to the information represented in a building model. Although semantic enrichment provides a valuable opportunity for BIM technology to reach its full potential, it is considered an emergent field of research. As such, the body of knowledge on the subject is incomplete and lacks formal definition of the process, possible applications, contributions, and computational approaches. In this work, an extensive literature review is performed to begin forming the body of knowledge in this field. A bibliometric analysis of relevant publications is implemented to identify previously explored approaches and methods for enrichment. Papers describing previous work in the field demonstrate the application of semantic enrichment to building information stored in accordance to the Industry Foundation Classes (IFC) schema as well as based on a web ontology. A detailed content analysis illustrates the benefits of semantic enrichment for various tasks in the BIM domain, including improvement of data exchange routines, design analysis and processing data obtained by remote sensing techniques. A formal definition for "semantic enrichment of BIM" is suggested based on the common features identified during the literature review. This work discusses the significance of semantic enrichment to a BIM workflow, pinpoints its current research gaps and describes direction for future research.*

KEYWORDS*: Building Information Modeling (BIM); Semantic enrichment; Industry Foundation Classes (IFC); Resource Description Framework (RDF); Interoperability, Symbolic AI, Machine Learning*

REFERENCE: *Tanya Bloch (2022). Connecting research on semantic enrichment of BIM - review of approaches, methods and possible applications. Journal of Information Technology in Construction (ITcon), Vol. 27, pg. 416-440, DOI: 10.36680/j.itcon.2022.020*

1 INTRODUCTION

Enrichment is defined as "The act or process of improving the quality or power of something by adding something else."[1] Semantic enrichment is a process associated and used for various operations in many domains. As such, different definitions, methods and possible applications of the process are provided in the literature. For example, in the medical field, semantic enrichment is used to analyze clinical trial data (Leroux et al. 2012). The process involves mapping the concepts within given data sets to anthologies and making implicit associations within the data explicit. The end goal of the proposed procedure is to enable the user to efficiently query the data. Also in the medical field, the use of semantic enrichment for data acquisition through semantic analysis is illustrated in Valêncio et al. (2013) to extract information from medical records.

In the computer science domain, semantic enrichment is presented as a method for providing interoperability between databases (Hohenstein and Plesser 1996). Development of the process was motivated by the technological advances and emergence of new applications with requirements to access data from different databases. In more recent work, the process of linking data sets to entities identified by the user on the web is also referred to as semantic enrichment. This process is designed to provide useful information about an entity without submitting new queries by the user (Fafalios and Tzitzikas 2013). Semantic enrichment has been used to classify Twitter posts by providing context to tweets through external data sources (Romero and Becker 2017). In the digital humanities field, semantic enrichment refers to the method of enhancing quality and discoverability of data provided by libraries, archives and museums (Zeng 2019). In this context, semantic enrichment encapsulates data mapping, alignment, matching, and merging, that result in the enrichment of existing metadata with more contextualized meanings.

In the engineering domain, semantic enrichment is often used for 3D reconstruction from Point Cloud Data (PCD) and for capturing existing conditions. For example, Fichtner et al.(2018) apply semantic enrichment to process 3D indoor point clouds to identify building elements that are important for indoor navigation in multi-story buildings. In this case, the point cloud is represented by an octree structure to identify empty and non-empty spaces. Sacks et al. (2018b) presented a complete workflow for acquisition of a semantically rich Building Information Model (BIM) based on data obtained from a laser scan. The workflow includes 3D geometry reconstruction, semantic enrichment and defect identification and assessment. The obtained model includes high levels of semantics, thus can support automated inspection of infrastructure.

Hichri et al.(2013) provide a review of approaches for creating an as-built BIM from PCD. They suggest the approaches can be arranged into four classes: heuristic approaches, approaches based on context, approaches based on prior knowledge and approaches based on ontologies. The heuristic approaches rely on encoding of human knowledge to distinguish between elements in a priory segmented scene. The context-based approaches are hinged on the same heuristic logic but they also leverage the relations between components. Approaches based on prior knowledge rely on the comparison between the scene and the 'as designed' models. And the ontology-based approaches rely on a priori knowledge of the geometry of the object structure, which can be extracted from various data bases (Hichri et al. 2013).

BIM models provide a rich representation of real world building concepts (Sacks et al. 2018a). Building elements, their geometry, attributes, properties and relationships are all given in a computer readable format. However, much of the information is not represented in an explicit form, making it inaccessible for other platforms (Krijnen and Tamke 2015). For example, topological relationships between building elements often remain implicit in the model which restricts information extraction from the BIM model for downstream analyses (Liu et al. 2016).

In addition, the available information is often not sufficient as input for various applications. Some applications require explicit representation of higher levels of semantic information than usually available in an authoring tool (Solihin et al. 2004). Integration of BIM with various applications is necessary to enable effective collaboration based on a digital workflow throughout different organizations, disciplines, and project phases (Panteli et al. 2020). BIM models serve multiple purposes thus requiring exchange of information across many disciplines and platforms. Every such platform is designed to perform a set of domain specific tasks which require an explicit representation of all concepts that are expected to play a role in the performance of those tasks.

[1] https://dictionary.cambridge.org/dictionary/english/enrichment

In practice, the use of the exchanged information in a receiving application remains a challenge as it requires human interpretation and manual work (Torma 2013). In other words, the exchange file needs to be manually pre-processed to fill all the information requirements of the target application (Eastman et al. 2009; Sacks et al. 2019). This problem triggered various initiatives aiming to reduce the amount of manual work by automatically inferring the required information, thereby semantically enriching the original building model (Beetz et al. 2006; Belsky et al. 2013; Bloch and Sacks 2018; Fahad et al. 2018; Werbrouck et al. 2020; Zhang et al. 2018; Zhang and Issa 2013). Yet, to date, the body of knowledge on this subject is incomplete and lacks an official definition of the process, main principles, goals, methods and associated concepts.

1.1 Preliminary work

Semantic enrichment of BIM, as presented in (Belsky et al. 2016), is a process designed to automate the manual preprocessing stage by using an inference engine to automatically add all missing information required by a receiving application. Given that semantic enrichment in general is not a new term, applications of semantic enrichment in the BIM environment that differ from the definition given by Belsky et al (2016) may be available in the literature. In the same manner, frameworks that describe the application of semantic enrichment of BIM may not be referred to as "semantic enrichment".

The term "semantic enrichment" appears in the scientific literature as early as 1991 (Castellanos and Saltor 1991) which supports the aforementioned statement that "semantic enrichment" per se is not a new process. To explore the possible application of semantic enrichment in the BIM environment, a preliminary search was performed in the Scopus database following the search code: TITLE-ABS-KEY ("semantic enrichment") AND [(TITLE-ABS-KEY ("BIM") OR (TITLE-ABS-KEY ("Building Information Model") OR (TITLE-ABS-KEY ("Building Information Modeling")]. Excluding papers from irrelevant fields such as medicine, business management, materials science, arts and humanities and social science, and other irrelevant publications the search resulted in only 25 papers. The earliest publication available in the Scopus data base with the terms "semantic enrichment" and "BIM" appearing in their title, keywords or abstract was published in 2013. The distribution of the documents in the search results over the years is presented in Fig. 1. The subject gained momentum since 2016 with the publication "Semantic Enrichment for Building Information Modeling" (Belsky et al. 2016) which was cited over 80 times.

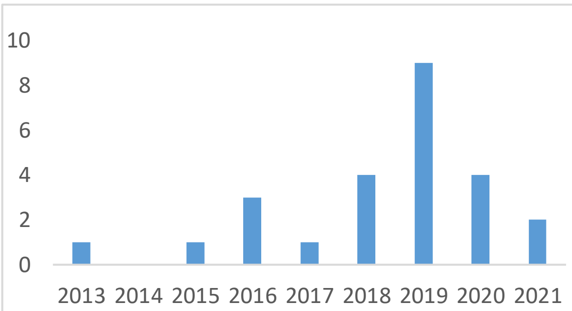


FIG.1: Distribution of number of published documents per year as identified in a perliminary search of publications

The significant number of citations of this paper alone suggests that the work done in the field of enriching BIM influenced many other researchers from possibly different domains. Yet, the keywords "semantic enrichment" and "BIM" appeared together in only 25 relevant publications. We can assume that some of the citing papers rely on the principles of semantic enrichment of BIM, or describe processes similar to semantic enrichment of BIM by other terms and keywords. To date, a structured set of concepts, principles and definitions that make up this domain is mostly lacking. The main objective of this study is to gain an in depth understanding of semantic enrichment of BIM, including the aims of the procedure, limitations and possible contributions. Through an extensive literature review and analysis, we can expect to shed light on the following questions:

a) What are the current challenges and problems that lead to efforts and initiatives targeting semantic enrichment of BIM?
b) What is the expected contribution of semantically enriching a BIM model?
c) What are the major developments in the field in terms of applicable computational methods?
d) What are possible future research directions?

The limited number of identified publications in the exploratory phase suggests that the term “semantic enrichment of BIM” is not well established in the scientific community and that there might be research groups working on similar subjects but not referring to them as “semantic enrichment of BIM”. Identifying and connecting the different efforts made towards enriching BIM models will contribute to establishing a firm foundation for advancing knowledge in the domain. Hence, the underlying goal of this work is to compile an official definition of the term "semantic enrichment of BIM" based on the common features extracted from publications describing existing efforts and initiatives, previously tested computational methods, and possible applications.

The rest of the paper is structured as follow: the next section of the paper describes the research methodology including data collection and analysis. Section 3 presents the results of a bibliometric analysis performed on the collected publications. Section 4 provides a review of previously explored approaches and computational methods for semantic enrichment of BIM based on content analysis of the most relevant publications. Section 5 provides a detailed description of possible applications identified through a keywords co-occurrence analysis. Discussion and conclusions are provided in sections 6 and 7 respectively.

2 METHODOLOGY

To answer the presented questions, this work relies on an a systematic literature review (Snyder 2019) based on documents available in the Scopus database. The Scopus data base provides wide coverage of journals in the engineering domain (Mongeon and Paul-Hus 2016), and it is also strongly correlated with other scientific data bases, such as Web of Science (Martín-Martín et al. 2018). Since the search performed in the exploratory stage yield only 25 relevant papers, the scope of the search was broadened to include all field codes instead of only title, abstract and key words. Also, the search was not restricted to the specific term “semantic enrichment”. Namely, data collection is performed following the search code: All (semantic AND enrichment AND BIM), resulting in a total of 462 publications available in the Scopus database. Excluding papers from irrelevant fields such as medicine, business management, materials science, arts and humanities and social science, and other irrelevant publications resulted in 234 papers including journal articles, review papers and conference papers.

We then explore the collected data set of publications through a bibliometric analysis performed in the VOSviewer platform, as described in Fig. 2. VOSviewer is a platform for generating and visualizing bibliographic networks that supports five types of bibliometric mappings: co-authorship analysis, co-occurrence of keywords analysis, citation analysis, co-citation analysis and bibliographic coupling analysis (van Eck and Waltman 2020). The collection of publications obtained from the search of Scopus data base serve as the input for a bibliometric coupling analysis, and a co-occurrence of key words analysis.

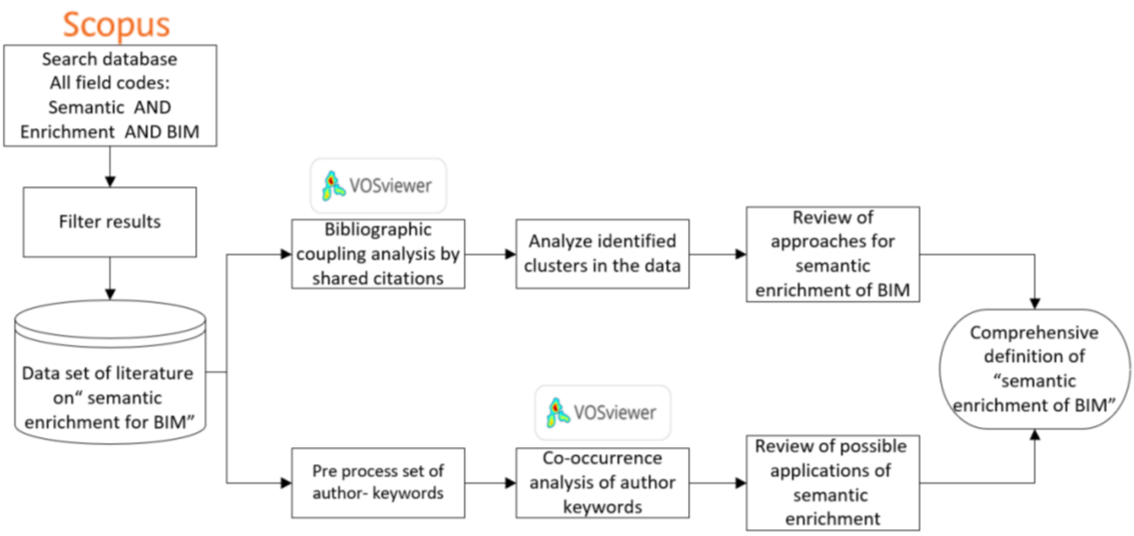


FIG.2: Description of the review process, including data collection, bibliographic coupling analysis, co-occurance of keywords analysis and expected conclusions

The assumption behind bibliographic coupling analysis is that scientific papers bear a meaningful relation to each other when they share references (Kessler 1962). Identifying clusters of papers that are closely related based on the amount of references they share can point to common areas of research described in the publications associated to each cluster. As presented in Fig. 2, the clusters obtained during bibliographic coupling analysis are further explored to identify the underlying research area in each cluster. A more detailed description of this process and the obtained results is provided in section 3.1 of this paper.

Author key words define the domain and the specific research niche that are covered in a given paper. Identifying key words that occur multiple times in a collection of papers can point to the most researched subjects, previously applied methods and previously explores problems within the domain. The author key words appearing in the collected data set of relevant publications are preprocessed and used as input for a co-occurrence analysis in the VOSviewer platforms. A review of possible applications and previously applied methods for semantic enrichment of BIM is then compiled based on the results of the analysis. A detailed description of this process and the results is given in section 3.2. Finally, as illustrated in Fig. 2, we rely on a review of identified research areas, research approaches, previously explored methods and applications to compile a comprehensive definition for the term “semantic enrichment of BIM”.

3 BIBLIOMETRIC ANALYSIS

Bibliometric analysis is a statistical tool used to identify the state of the art, research gaps, directions, and other essential information in a given scientific area. The use of bibliometric analysis provides researchers means to identify and support paths towards the development of future research (José de Oliveira et al. 2019). The goal of using bibliometric analysis in this work is twofold: a) to identify the general research directions and approaches for enriching a BIM model. b) to identify the previously explored methods for enrichment and the possible applications of the process. To achieve the set goals, we perform the bibliometric analysis in two stages, the first stage is based on a coupling analysis of documents, and the second stage relies on counting the co-occurrences of author keywords. A detailed description of the methods and results of each stage are presented below.

3.1 Bibliographic coupling analysis

The collection of relevant publications obtained from the Scopus database is imported to the VOSviewer platform to perform bibliometric analysis. Since the main goal of this stage is to identify and further explore papers that present a general approach to the application of semantic enrichment in the BIM environment, a coupling analysis was performed using documents as the units for analysis. In this method, the relatedness of items in the data set is based on the number of references they share. No threshold for minimum number of citations was set, to take into account recent publications that have not been cited yet. The strength of bibliographic coupling links is calculated for all documents. To compile a readable network, we choose to visualize only the 150 documents with the greatest total link strength. The results are visualized in a bibliographic network presented in FIG. 3.

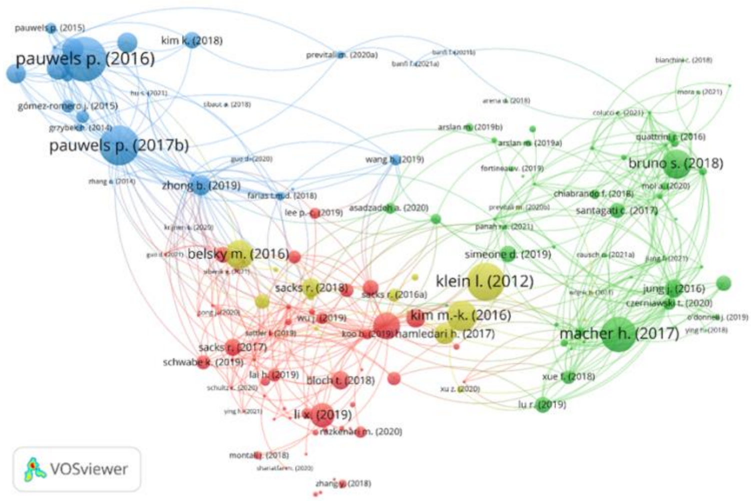


FIG. 3: Visualization of bibliographic analysis results, grouping the publications into four clusters. Each cluster is denoted by a different color in the presented network.

Based on the bibliographic coupling strength between documents, the obtained data set is arranged into four clusters, each denoted by the same color, as illustrated in the network in FIG. 3. The size of the nodes in the network is an indication to the significance of an item with respect to the other items in the same cluster (Van Eck and Waltman 2010). The abstracts of five most prominent papers in each cluster are further analyzed to identify the main theme of the publications, results are presented in Table 1.

Note that many of the prominent documents are review papers, understanding the main theme of these papers can indicate the underlying topics of the clusters. While it is difficult to pinpoint the common research issue in clusters number 1 and 3, it is clear that the publications in cluster number 2 focus on Semantic Web technologies and the conversion of IFC building representation to the OWL ontologies. The publications associated to cluster 4 deal with processing data obtained by imagery or point clouds and its integration with a BIM model. Reading the abstract of papers associated to cluster 1 and 3 revealed that the majority of publications exploit the data stored in the IFC file for various purposes.

Table 1: Most prominent publications in every identified cluster according to the coupling analysis results

Cluster 1		**Cluster 2**	
Reference	**Main theme**	**Reference**	**Main theme**
(Klein et al. 2012)	Scene reconstruction based on image processing to document and verify as-built condition	(Pauwels et al. 2017)	Literature overview of semantic web technologies in the AEC industry. (Review)
(Kim et al. 2016)	Quality assurance for full-scale precast concrete elements using 3D laser scanning and IFC	(Pauwels and Terkaj 2016)	Conversion of EXPRESS schemas into OWL ontologies.
(Belsky et al. 2016)	Semantic enrichment to supplement an IFC exchange file with semantically useful concepts inferred from the information contained in the building model	(Costa and Madrazo 2015)	Integration between product component catalogues and BIM through the Linked Data approach.
(Hamledari et al. 2017)	Incorporation of progress data into 4D BIM by updating the IFC file	(Terkaj and Šojić 2015)	An enhancement of the existing IFC to OWL conversion patterns.
(Sacks et al. 2018b)	Compiling information for bridge inspection and system management, using point cloud data processing and IFC	(Abanda et al. 2013)	Review of applications of semantic web in the built environment. (Review)
Cluster 3		**Cluster 4**	
Reference	**Main theme**	**Reference**	**Main theme**
(Yin et al. 2019)	Review of research on BIM for off-site construction. (Review)	(Bruno et al. 2018)	Historic Building Information Modeling. (Review)
(Li et al. 2019)	Integrating BIM and Prefabrication Housing Production. (Review)	(Macher et al. 2017)	Semi-automatic reconstruction of indoors of existing buildings from point clouds for their integration in BIM software.
(Hamledari et al. 2018)	IFC based development of As-is and As-built BIM.	(Jung et al. 2016)	Automated method for creating a 3D wire frame model of indoor surface structures with point-cloud data.
(Bloch and Sacks 2018)	Comparison of Machine Learning and rule-based inference techniques for semantic enrichment.	(Santagati and Lo Turco 2016)	A library of architectural elements to support Historical Building Information Modeling. (Review)
(Sacks et al. 2017)	Procedure for compiling inference rules for semantic enrichment of BIM models.	**(**Xue et al. 2018)	Automatic generation of semantically rich As-built Building Information Models using 2D images.

We can conclude that there are two major research directions regarding enrichment of BIM models, more specifically the representation of building data as the starting point of the process. One research direction is focused on the IFC representation of building data to manipulate the data stored in the building model, and another is focused on the OWL representation of the building as the source of information. These approaches are further explored through a more detailed review of relevant publications.

3.2 Co-occurrence of key words analysis

This stage is concerned with the identification of possible applications and previously explored computational methods for semantic enrichment of BIM as presented in previous research. Since author key words provide a meaningful insight to the main research domain as well as to the specific research subject described in a given paper, we use the co-occurrence of key words analysis to identify previously explored applications of semantic enrichment in the BIM environment. VOSviewer utilizes a text-mining technique to the content of titles, abstracts and author keywords to identify items where the same keywords appear. Due to syntactic differences, the data set was processed to ensure that keywords with the same meaning are grouped and their occurrence counted correctly. For example, 'BIM' appears in the obtained data set as Building Information Model, Building Information Modeling, and Building Information Modeling (BIM). The minimum number of occurrences for every keyword in the analysis was set to 3, which produced 660 keywords that meet the defined threshold.

Table 2: Most occuring keywords in the data set

Occurrences	Keywords
5 - 10	TLS, Semantic, Scan-to-BIM, RDF, Laser scanning, IDM, Deep Learning, MVD, Linked data, Artificial Intelligence, OWL, Semantic enrichment, Energy
10 - 20	Interoperability, Semantic web, Heritage, Point cloud, Ontology, HBIM
Above 20	BIM, IFC

The most occurring keywords according to the analysis are presented in Table 2. Results of the analysis are visualized and illustrated in Fig. 4. Based on the obtained network we can see four distinct groups in the list of keywords:

a) A cluster that contains generic terms such as "BIM", "IFC" and “semantic enrichment”. These are general keywords that do not provide any indication about possible applications of semantic enrichment of BIM. However, they do indicate that there is a correlation between semantic enrichment processes to the Industry Foundation Classes Schema (IFC), which reflects one of the research directions identified in the coupling analysis.
b) Key words such as “Ontology”, “Semantic Web”, “Interoperability”, “RDF”, “OWL” are grouped into a second cluster, which also reflects a research area identified during the coupling analysis.
c) The third cluster contains keywords such as “HBIM”, “Point cloud”, “Heritage” which indicate that one of the possible applications for enriching BIM is processing Point Cloud Data and generating BIM models of historical buildings.
d) The fourth cluster reflects the recent development in the AEC industry as it contains terms like “artificial Intelligence”, “Machine learning”, “deep learning” and “Industry 4.0”. These terms point to the possibility of applying learning methods for enrichment of BIM models.

The bibliometric analysis outlines the different research areas and efforts that need to be reviewed and connected to understand what is semantic enrichment of BIM. Among those research efforts is creation of historical BIM, point cloud processing, interoperability, buildings energy performance, the capabilities of artificial intelligence, machine learning and deep learning in the context of enriching a BIM model, all considering different approaches for representing building information. Review of relevant publications from each identified research direction contributes to a deep understanding of semantic enrichment of BIM in general, the possible approaches for enriching a BIM model, as well as the aims, contributions, and possible applications of such a process.

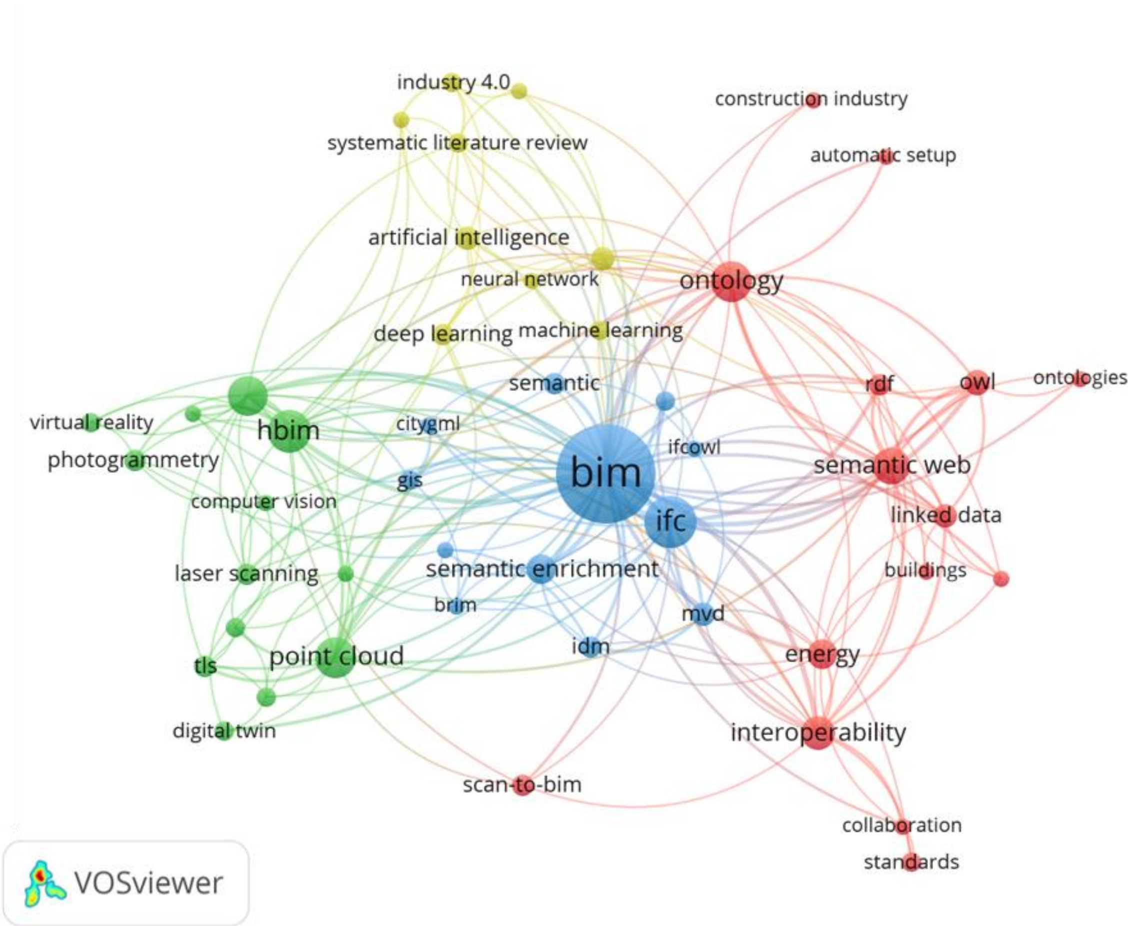


FIG. 4: A network visualizing the results of co-occurrence of author key words analysis. Results indicate four underlying themes in the collection of publications denoted by different colors in the network.

4 APPROACHES FOR SEMANTIC ENRICHMENT OF BIM MODELS

Results of the coupling analysis point to two distinct approaches for representing building information, previously used as a starting point for enriching a model. While one research direction is focused on the most common format for data exchange in the industry, i.e. the Industry Foundation Classes (IFC), the other assumes that in order to extract meaningful semantics the model has to be represented using the WEB ontology. The methods for enrichment slightly vary based on the initial representation of the model.

4.1 IFC representation of Building Information

Most documents associated to clusters 1 and 3 (as listed in Table 1), describe processes that rely on an IFC representation of building information. A deep review of these documents indicates that there are two main approaches for enrichment in this case. The identified approaches are listed below:

4.1.1 Inference based enrichment

Some of the identified publications describe the use of symbolic AI techniques to exploit the information implied but not stored explicitly in BIM models. As stated in Bloch and Sacks (2018), the backbone of semantic enrichment is the ability to query information stored in the BIM model. Domain specific query languages for building information models such as BIMQL (Mazairac and Beetz 2013) and QL4BIM (Daum and Borrmann 2013) allow the application of high level spatial operators on the geometric representations of individual objects to exploit the meaningful information constructs implied in the models. Similarly, the proposed methods for enrichment focus on manipulating the information explicitly stored in the IFC to intelligently retrieve new facts, which is the basis for creation of an expert system for semantic enrichment.

Such inference methods can be referred to as deductive reasoning. Deductive reasoning is based on logical propositions that require people to draw reasoned conclusions (Sternberg et al. 2012). One of the primary types of deductive reasoning is conditional reasoning, in which conclusions are drawn based on if-then propositions (logical rules). Deductive reasoning is the foundation of expert systems designed to mimic the decision-making ability of a human expert in a specific domain (Todd 1992). Rule based reasoning systems are not developed to model human intelligence, but to simulate performance in just one narrow field (Sternberg et al. 2012). The SeeBIM engine (Belsky et al. 2016) for example, is a data driven, rule based system for semantic enrichment of BIM models. It encapsulates procedural knowledge acquired from domain experts in a form of logical statements.

The platform receives an IFC file as input and deploys a predefined set of inference rules to retrieve new information about the model. The platform then writes a new IFC file that contains both representation of the original building model as well as representation of the new inferred facts. The ultimate goal of the described process is to provide an explicit representation of all information required by a receiving application within the exchange file. The rule sets represent domain expert knowledge in the form of logical IF-THEN statements manifested by elements' geometry, functional properties and relationships to other elements, much like in the previously explored domain specific query languages. These statements are then used in a forward chaining strategy to infer new facts and integrate them into the growing database of established facts, until the fact of interest is finally derived (Todd 1992).

Sacks et al. (2017) suggested a procedure that relies on rigorous rule sets for semantic enrichment. The procedure is composed of seven steps and requires representation of expert knowledge in a form of matrices. This was demonstrated for BIM objects classification using a synthetic model of a concrete bridge designed to illustrate the process of obtaining a semantically rich model from PCD. Namely, the initial model contained only 3D geometry without any semantic information, including object types, and was enriched using rigorous rule sets (Sacks et al. 2018b). Although this led to 100% accurate classification results, such a procedure can be tedious when dealing with multiple classes, and may be less effective when the geometry of the classified elements is not as distinctive. This issue was explored by implementing the procedure for classification of spaces in a BIM model, and indeed proved to be less useful for classification of abstract spatial elements (Bloch and Sacks 2018). This demonstrates that such rigorous rules are not effective when there are not enough distinctive geometric features to the classified elements.

Wu and Zhang (2019) suggested another rule based, iterative method to classify BIM objects in IFC. The proposed algorithm relies on a data-driven and pattern matching rule-based approach. The rules were compiled based on a collected and labeled data set of IFC entities. The geometric features of a single object were analyzed and used to compose patterns and rules for identifying objects with a similar geometric representation. The method was implemented to classify building elements into five predefined types and the obtained results achieved 100% precision and recall. The results indicate that geometric features are useful for compiling rules for classification of elements in the AEC domain. However, geometric features alone are not sufficient to provide classification of objects with similar geometric representations. This aligns with the conclusions of the room type classification experiment presented in Bloch and Sacks (2018).

One of the major advantages of deductive reasoning-based systems is that they provide conclusions that will always be unique and reliable, unless of course the rules were not correctly defined or there were errors in the input data. In other words, if you start with a truth you can infer only new truths (King 2011). Another benefit of a rule based approach is that the propositions are readable to the user thus the conclusions can be interpreted and validated by human experts (Todd 1992). Unfortunately, deductive methods are often insufficient to reach a conclusion as they rely only on what is already known (King 2011). In other words, they can only reach a conclusion in cases where fixed, predefined outcomes can be determined based on a set of well-known facts. Rigorous rules limit the achieved intelligence since they are restricted to do only what they have been written to do.

Application of intuitive logical rules that can be easily defined and understood by experts proved to be useful for inferring several types of information implicitly represented in BIM models. However, we cannot represent everything in the design of our buildings and facilities as rigorous rules. A human expert can easily reason over the implicit information in a BIM model by relying on various types of data. This data is not limited to the geometric representation of model elements and their spatial placement and relationships. Annotations, text, images and quantity values all contribute to our understanding of a design. To automate the reasoning process, Songa et al. (2019) suggest application of Natural Language Processing (NLP) technique to leverage the information represented by text in a building model. They specifically tackle the problem of spatial elements classification in educational facilities in Korea and illustrate a process for using a deep learning model to classify spaces in a school building. As different naming conventions may be used in every model, taxonomic problems and lexical problems affect the interoperability adversely.

Although trained professionals from the construction industry can easily apprehend the functions of a space even from a two-dimensional representation, this information is currently not computer readable. Bloch and Sacks (2018) also addressed the problem of space usage classification by using an iterative machine learning based technique that relies on non-textual data as features of each space. Their work illustrates the application of a machine learning approach through the use of Artificial Neural Network (ANN) algorithm based on the geometric features of a space as well as the relationships between the spaces. However, the achieved accuracy in the presented test case is only 82% as it relies on a very limited dataset.

ANN have also been used for classification of building stories by their function and intended use (Krijnen and Tamke 2015). The same work also demonstrates the use of unsupervised learning technique for detecting misclassifications in the IFC file. Similarly, Koo and Shin (2018) present a novelty detection approach to check the integrity of IFC classes to BIM element mappings. Application of Support Vector Machine (SVM) for classifying BIM elements to their corresponding IFC classes has been demonstrated in Koo et al. (2019). Algorithms for image recognition also proved to be useful for classification of structure to three building classes: apartment buildings, industrial buildings and other building types (Lomio et al. 2018).

The approaches described above can be referred to as inductive reasoning techniques. Inductive reasoning is a way of logical thinking, in which generalizations are made based on previous experience. That is, reached conclusions are predictions about new instances based on previous observations. We use inductive reasoning in our everyday life, to predict the weather for example, however our predictions are not guaranteed to be correct (King 2011). In order to assess the degree of belief in a reached conclusion, we quantify it in a probabilistic form (Flach and Kakas 2000). Machine learning algorithms are inductive by nature as they are trained to provide predictions based on a set of examples. The obtained results are probabilistic and are not guaranteed to always be correct. Previous work demonstrates good performance of various machine learning algorithms for classification of building elements based on data obtained from BIM models (Bloch and Sacks 2018; Koo et al. 2021).

The accuracy of the obtained results is reliant on the size of the data set used for training (the number of provided examples), and the quality of selected features (Kotsiantis et al. 2007), and a 100% accuracy cannot be achieved. Nevertheless, Bloch and Sacks (2018) demonstrate that not all semantic information can be inferred by deduction and that some cases require a more adaptive and flexible approach that allows representation of irregularities and exceptions. In these cases, we must consider the chances of receiving false conclusions and their possible effect on the information flow and on the processes that make use of this information. This raises questions about the scope of semantic enrichment and the impact of the process on receiving applications within the exchange workflow.

4.1.2 Integrating IFC with external data sources

Research efforts have been made to enhance information stored in a BIM model without using inference engines or query languages. This is possible by linking the information stored in the model with an external data source. Hamledary et al. (2018) proposed a method for automatic integration of site inspection results into the model's IFC-based data. Site inspection data can include identified defects, elements as-built type, as-designed type, time/date of the inspection, inspector's notes, and captured images. This information is linked to each inspected element to be visualized in a BIM environment. A similar work process was presented in Hamledary et al. (2017) for integration of progress information collected during site visits to update a four-dimensional (4D) building information model in terms of schedule and progress.

Xue et al. (2018) focused on generating a semantically rich as-built BIM using 2-D images. In this work, information available in the original as-built model, which usually lacks semantic information, is integrated with open-access BIM components that contain high level of semantics. This approach takes measurement data (images) and a library of semantically rich components and uses them as input for a constrained optimization problem. They use a similarity measure between the digitally represented components and the obtained measurements as the objective function and derive the constraints from the topological requirements between components. The semantically rich BIM is obtained by linking the recognized geometries with existing BIM component libraries. Although this work illustrates the general idea of translating the as-built BIM generation problem into a constrained optimization problem, areas such as the required input/output formats, variable settings, and performance measurements need further development.

Obtaining an accurate representation of the current state of a building is one of the key challenges of modern Facility Management (FM). A 3D point cloud can be used to capture the current physical state of a building, however it consists of raw unstructured data with no associated semantic information which is required by most applications (Sadeghineko and Kumar 2020). Generation of a semantically rich as-built BIM is an ongoing challenge which usually consists of two stages: segmentation of raw data (such as 2D images or 3D PCD), and enrichment of the resulting geometry with semantic information. Semantic registration tools are the most commonly used methods for adding semantics to the model. Xue et al.(2019) represented a semantic registration approach, enhanced by architectural domain knowledge to address the problem of noisy data and occlusion. As the described test case was a University lecture hall, the elements were assigned with general semantic information (such as parent object), but also with case specific information such as row number and sequence number. In the work of Xue et al. (2018), semantic information was registered from an external semantic resource (a Revit family in this case) to produce a semantically rich BIM.

Sadeghineko and Kumar (2020) present an approach to generate a semantically rich parametric model for existing assets, based on geometric information obtained from PCD. In this work the needed information was manually captured in comma-separated values (CSV) files and then converted to a Resource Description Framework (RDF) and ultimately to Industry Foundation Classes (IFC). An enriched model obtained from a laser scan was also the basis for production quality evaluation of prefabricated components (Xu et al. 2020). In this case, the obtained production quality information should be extended to the IFC standard to provide detailed parameters for onsite construction. This extension is either an addition of a new entity or a proxy entity, or an extension of the existing elements' property set.

The overall idea of adding semantics to a model that contains building elements with no additional information, is one of the purposes of semantic enrichment of BIM. However, semantic enrichment per se, is designed to automate not only the process of integrating information into the models, but also the process of retrieving the information, not relying on manually created data sources like implemented in the work described in this section. Evidently, the existing literature refers to two distinct procedures in the same manner. One of these procedures is semantic enrichment of BIM which differs from integration of IFC with external data sources by relying on inference engines to retrieve new information. We therefore refer to integration of IFC with external data sources as semantic enhancement.

4.2 Semantic Web technologies for building information

Semantic Web technologies allow to represent information in structured graphs, Resource Description Framework (RDF) (Arenas et al. 2009). The Web Ontology Language (OWL) adds constructs of description logic to the RDF schema thus extending its expressive capabilities. Allowing representation of the semantic meaning of concepts, OWL enhances the RDF making it useful for inference of new semantic information that is implicitly available in the original RDF constructs (Pauwels et al. 2017).

As stated in (Werbrouck et al. 2020) "*Ontologies do not only organize concepts on a specific domain, but also aid in negotiating between contradictory information on the web and <u>the process of making implicit information explicit</u> through analysis of existing relationships and definitions in a graph, called 'inference'.*" Namely, an automatic procedure can generate new information based on the existing data and additional rules. The use of ontology based applications in the construction industry have been explored and illustrated for implementation of code compliance checking (Yurchyshyna and Zarli 2009; Zhong et al. 2012), cost estimation (Liu et al. 2016), construction planning (Zhong et al. 2015) and energy analysis (Baumgärtel and Scherer 2014).

The motivation for using Semantic Web technologies in the AEC industry is presented in Pauwels et al. (Pauwels et al. 2017). This work points to three main reasons for using semantic web technologies for design and construction: interoperability, linking data across domains and logical inference and proofs. The last aim relates to the use of generic inference engines based on First Order Logic (FOL) to infer new information from the original building model. Combining the inferred information with the information originally stored in the model results in an enriched building model, which is the ultimate goal of a semantic enrichment process.

The applicability of ontologies to problems from the BIM domain has been discussed in the work of (Beetz et al. 2009). In their work, an ontology for the building and construction sector based on the industry foundation classes is developed to enhance the machine readability and interpretability of information. A method for lifting EXPRESS schemas onto an ontological level is presented resulting in the ifcOWL ontology, and a practical example for the use of OWL and RDF for partial model extraction is illustrated. This work demonstrates the ability to apply generic query languages and reasoning algorithms to domain specific problems that otherwise have to be manually hard-coded into applications for processing building information.

The connection between semantic web technologies and the IFC standard (namely, the agreed Web Ontology Language for IFC - ifcOWL) was further investigated in (Pauwels and Oraskari 2016; Terkaj and Šojić 2015). All efforts aim to increase machine readability and eventually provide a semantically rich platform to support the integration of software tools and data exchange (Pauwels and Terkaj 2016). However, due to the complexity of the resulting ontology, RDF graphs that are based on ifcOWL are difficult to query efficiently (Werbrouck et al. 2020) which is a drawback in terms of semantic enrichment applications.

Although the use of semantic web applications in the AEC domain has been extensively investigated, and the ability to exploit explicitly represented information to derive new facts about the building model has been previously illustrated, the term “semantic enrichment” is rarely mentioned in the publications describing reasoning over ontology-based building representation. In fact, searching the Scopus engine with the field code TITLE-ABS-KEY (“Ontology” AND” semantic enrichment” AND” BIM”) yield only four relevant publications. In the same manner, searching for the key words ("semantic web" AND "semantic enrichment" AND "BIM") yield only three publications. This demonstrates that research groups are indeed investigating processes for enriching BIM models but not referring to them as “semantic enrichment of BIM”. This strengthens the need for a comprehensive definition of the process and its goals.

5 POSSIBLE APPLICATIONS OF SEMANTIC ENRICHMENT OF BIM

The main themes and research areas in the obtained collection of papers are identified through the bibliometric analysis. Results of co-occurrence of key words analysis indicate that semantic enrichment of BIM is useful for generating models of existing buildings, building performance analysis (such as energy analysis) and for interoperability. To further investigate how semantic enrichment of BIM is applicable in these areas a content analysis of the relevant papers is presented below.

5.1 Scan to BIM

A semantically rich as-built BIM can be valuable for facility management, energy consumption simulations, building renovation etc. Laser scanning is a widely adopted technique for surveying existing facilities, however PCD can only represent geometric data that needs to be parsed and reassembled as BIM components. The result of this process is a simplified model of the sensed facility that includes the geometric representation of the identified components but no other information is associated to them. Previous research addresses the issue of adding the semantic information to the geometric model for various purposes.

The use of BIM technology for storing and managing information about historic buildings is known as Historic Building Information Modeling (HBIM). The models should include representation of semantic information, fundamental for the maintenance and conservation of the building, such as the relations between 3D building objects, definition of materials or degradation processes. Currently, HBIM models usually remain poor in terms of semantics related to relevant features of the building. The relevant semantic information can be supplemented through various external data sources integrated into the BIM environment. For example, in Mol et al. (2020), geometric data obtained from PCD was supplemented with data from drilling resistance tests for representing information regarding decay and damage in existing timber structures. In this work, the information is linked to each element by including a uniform resource locator (URL) link, which leads to an external database containing raw and analyzed data. In this case, the data is collected, analyzed and stored in an external environment and it is not based on the information stored in the building model.

Creation of a semantically rich HBIM often includes association of non-geometrical data to specific points belonging to physical components of the building through Geographical Information Systems (GIS). Since this approach is only efficient when the models are obtained from point-based survey techniques, many researchers addressed the issue of integrating semantic information to object oriented representations of historic buildings

(Simeone et al. 2019). For example, Mora et al. (2021) derived the relevant semantic information from various related assets and integrated it in a BIM environment by using a specially designed plugin for Revit that allows assigning different families with all their associated parameters to the 3D elements. This is of course limited to a specific platform.

In the work of Werbrouck et al. (2020) a framework called 'scan-to-graph' (STG) is proposed to support the current scan-to-BIM processes by integrating the concepts of scan-to-BIM with Semantic Web technologies. The authors present a plugin for the Rhinoceros software, which provides an RDF representation of multiple interlinked topologies and other metadata about the building. Enrichment of the existing geometries, which is the core of the plugin, can be used for location specifications, relationships with other elements, product identification and metadata remarks. Additionally, SPARQL queries can be executed on the building model to visualize specific geometries. The flexibility that this method provides in identifying and classifying non-common elements that may exist in historic building is a strong benefit of the approach but it may also be seen as a drawback. This is due to the fact that it is possible for the users to create their own classes, even if such classes exist in standardized taxonomies.

Previous research suggests that information retrieved from a laser scanner can also support bridge management systems, provided that it is appropriately processed to contain bridge elements representation (Sacks et al. 2018b; Yan and Hajjar 2021). To facilitate the applications of laser scanning in bridge inspection and management, Yan and Hajjar (2021) propose an automated approach for detection and extraction of main structural elements of steel girder bridges from point clouds. In the presented test case, the girders have a distinctive I-shape section thus they are recognized first. Other structural elements can then be recognized by leveraging their spatial relationships with respect to the detected steel girders.

Isailović et al. (2020) propose a framework for generating a semantically rich as-is BIM model of bridges with structural damage features obtained from PCD. They explore image-based multiview classification to detect and extract concrete macro damage features from 3D point clouds. The input for the algorithm is an as-designed model of the bridge and a point cloud describing its current state. A CNN is then applied for image-based damage detection and analysis by training on numerous photos of damaged and non-damaged reinforced concrete elements. Observations made in course of inspections are depicted to textual description and 3D geometric Representation. Once the damage elements are detected their geometry is extracted as triangular meshes, and inserted in the as-built model. This results in a semantically rich as-is model that reflects the current state of the bridge and includes information like damage type, severity, and extent.

Sacks et al. (2018b) suggest the use of the SeeBIM engine (Belsky et al. 2016) for semantic enrichment of concrete bridges. In this case, the 3D geometry of a bridge is obtained from point cloud data and supplemented with semantic information about the objects classification, numbering, and aggregations by application of forward chaining rule sets based on distinctive geometry representation of the elements and the spatial relationships among them (Sacks et al. 2017). Note that in this case no external data sources are involved as all additional information is inferred based on the geometric information stored in the initial model. However, as the backbone of the developed framework are rigorous rules that leverage the geometric features of concrete bridges, the method is limited to a specific bridge type and cannot be easily adapted.

5.2 Interoperability

A smooth data exchange between different platforms and stakeholders in the AEC industry is vital in all stages of a construction projects (Panteli et al. 2020). The Industry Foundation Classes (IFC) schema ("BuildingSMART" 2021; Liebich and Wix 1999) was created to enable an exchange of the available information. An alternative to sharing information through an IFCSTEP physical file is by converting it into the extensible markup language (XML), resulting in the domain specific ifcXML format. Similarly, building models can also be represented as RDF graphs following the previously discussed ifcOWL ontology.

Currently, the IFC format is the most widespread and commonly used standard for data exchange in the AEC industry ("BuildingSMART" 2021). The IFC is a rich data schema that includes representation of entities from various disciplines, making it very complex and redundant. This may cause inconsistencies and incompatibilities between data representation in the exported files to the receiving platforms, thus hindering the ability for seamless exchange of information with unambiguous meaning. Venugopal et al. (2015)states that IFC schema is necessary but not sufficient for achieving robust data exchanges as it lacks semantic clarity in mapping entities and relationships. Semantic enrichment of BIM models may automatically correct the inconsistencies and supplement the models with information required by the receiving application.

Some researchers focused on ensuring BIM elements are mapped correctly to their corresponding IFC class entities to improve data exchange routines. For example, Wu and Zhang (2018) suggest a six-step process for automatic classification of IFC entities into pre-defined categories using a rule based algorithm. This process contains many sub-algorithms each designed to recognize a specific category based on the intrinsic properties of BIM objects. By studying the shape representation of a given object, a set of geometric properties is defined and translated to corresponding pattern-based rules for classification of similar entities. This approach is limited to dealing with classification of objects with distinct geometric representations.

The work of Koo et al.(2021) investigates the use of deep learning to add information about subtypes of building elements. Providing detailed categories for typical building elements is more challenging then classifying element types because they share similar geometry representation and require other details for differentiation. The use of deep learning algorithms such as Multiview Convolutional Neural Network (MVCNN) and PointNet were explored to automate the classification of door and wall subtypes. The obtained results were analyzed and compared to results obtained by the Support Vector Machine (SVM) algorithm. The proposed classification process is meant to ensure integrity of the IFC file and to enable reuse of information in applications that require a higher level of semantics then usually exported in the commonly used MVDs (Koo et al. 2021). Evaluation of performance of SVMs for classification of subtypes of three IFC entities, IfcDoor, IfcWindow, and IfcColumns is provided in Koo et al. (2019). These processes focused on identification of misclassifications of existing building elements to ensure integrity of mapping BIM elements to IFC classes. Information that was not explicitly provided was not supplemented and identified misclassifications were not corrected in the BIM model.

Every design analysis tool requires tailored information to be present in the BIM model, including information about elements' properties, attributes and relationships. To promote collaboration between multiple disciplines, the Model view definition (MVD) has been developed under the IFC schema as a subset standard for domain-specific data exchanges. However, the information defined in each MVD is often insufficient to conduct an advanced design analysis. For example, exchange of information between architectural design and structural analysis platforms requires extensive geometry interpretations. Sibenik and Kovacic (2021) suggest a framework for an automated interpretation of an IFC building data model to a structural analysis model. The framework consists of three parts: classification, interpretation and automation. Interpretation in this work comprises multiple procedures required for preparing the model for import to the structural analysis platform. This includes filtering, non-geometrical interpretations and geometrical operations. Essentially, interpretation is the process of reasoning over the information that is explicitly represented in the native model. The automation stage is then used to put all the information in context by using a mapping process between a domain-specific model and the native building model.

An alternative to using IFC for data mapping was illustrated in Costa and Sicilia (2020). Their work identifies fourteen types of possible data mapping patterns that enable transformation of one data model into another. Some of the identified mapping patterns require application of additional conditions to the existing data, to derive the required data representation. In this work, Semantic Web query languages were explored as a tool to facilitate an automated data transformation. The process was demonstrated by application of SPARQL based queries for transformation of BIM models to energy buildings simulation models in an automated way. The transformation of data from a source domain into a target domain requires the knowledge of a domain expert to formalize the data mapping patterns for every element. This knowledge is embedded into the target model through semantic queries to represent all information required by the receiving platform in the correct format.

Beetz et al (2006) addressed the issue of translating different representations and mappings for various scenarios. In their work, spatial relationships between building elements are inferred using OWL to transform 'hidden' implicit knowledge to an explicit representation. They propose the Semantic Web Rule Language (SWRL) rules and SPARQL queries for semantically enhanced reasoning. This indicates that interoperability issues of building information modeling can be addressed using existing, generic tools tailored to a domain specific applications or data exchange requirements. The output of SWRL rules and SPARQL queries is inferred semantics that can be generated to automatically match the information structure of a target application.

Unfortunately, despite the efforts made in different research directions, a seamless exchange of information between different platforms and stakeholders in the AEC industry has not been achieved yet.

5.3 Building design and performance evaluation

Retrieval of all required information in its correct representation is one of the challenges in all existing applications for automated code compliance checking (Eastman et al. 2009; Preidel and Borrmann 2015). Existing applications for automated code checking require semantic information to be manually supplemented by the users directly in the code checking platform. Hence, all applications are limited both in scope and in the amount of automation that can be achieved due to the labour intensive, manual pre-processing stage. For example, in Solibri Model Checker (Solibri 2017) platform, which is one of the more advanced applications for compliance checking, information such as space classification, number of exits from a building, definitions of fire separations and fire compartments is required to be manually specified by the user. Much of this information has been successfully inferred automatically through a semantic enrichment process (Bloch et al. 2019; Bloch and Sacks 2020) thus making it useful for reducing the manual preprocessing of model prior to checking.

Bloch et al. (2019) proposed the use of the SeeBIM engine for semantic enrichment to prepare models for checking against a local regulatory code clause for security rooms in residential buildings. The described process includes analysis of the specific code requirement to identify all information that needs to be supplemented, application of computational methods for information retrieval and integration of the retrieved information into the BIM model. Once the model was enriched, a rule set for compliance checking was compiled and deployed in the SeeBIM engine to check compliance of the given model to the regulatory requirement. The implemented procedure facilitates a completely automated compliance check with no user intervention. Various computational methods for semantic enrichment to support automated code checking are analyzed and compared in Bloch and Sacks (2020). This research defined four groups of semantic enrichment tasks that are useful for preparing a BIM model for compliance checking: classification tasks, association tasks, creation tasks and classification tasks. The computational approach for inferencing information in each of the groups determines the overall approach for semantic enrichment.

Semantic Web technologies can also be used to facilitate compliance checking. In this case the building model is represented by an RDF graph following the OWL ontology, and the rules are logical conjunctions of declarative IF-THEN statements. Once all the data and all the rules are available in a complete and consistent form, the rule checking process is straightforward (Pauwels and Zhang 2015). To represent the regulatory requirements, Zhong et al. (2015) explored an ontology-based semantic modeling approach of regulation constraints for construction quality compliance checking. They suggest to model regulation constraints into OWL axioms and SWRL rules and present a method for quality checking in parallel to the construction process. However, their work is limited to constraints only about underground-diaphragm-wall and still needs to be validated for other constraint types. Similarly, Bouzidi et al. (2012) designed a database of rules by manually extracting regulation constraints and transformed them into SPARQL by domain experts. However, their experiment showed that some constraints remain non-transformable into SPARQL which limits the applicability of the proposed approach.

Guo et al. (2021) also tackled the problem of rule extraction from regulatory documents by combining Natural Language Processing (NLP) with automatic generation of SPARQL queries. As part of their effort to elevate the manual work required for code compliance checking, the authors also address the need to supplement information required for checking. They suggest enriching the objects represented in the model by extending functions and data properties with SPARQL. The overall approach for code checking presented in this work contains extracting rule information from regulatory documents, semantic enrichment of BIM, mapping rule keywords to BIM RDF data, automatically generating SPARQL query and compliance results. This work demonstrates, once again, the need for semantic enrichment procedures to achieve higher levels of automation in various workflows.

Making a BIM model more processable for specific applications such as compliance checking or performance analysis, is currently a manual task performed by domain experts. Semantic enrichment aims to elevate the need for such manual processing by automatically adding required objects and attributes as specified in the receiving application. For example, Beetz et al (2006) illustrated the use of Semantic Web Rule Language (SWRL) and SPARQL queries to convert IFC geometry to the input for energy simulation platform. The semantics of a given model is thus enhanced with the results of the queries.

Baumgärtel and Scherer (2014) created an extended energy BIM (eeBIM) by enriching the original building model through RDF ontology. SPARQL was used to query existing models and infer new information through semantic rules. Green building design parameters were then integrated as additional ontologies, thus creating a new RDF representation of the building model that includes higher level of detail than the original model. The final RDF graph contained all necessary information to run thermal energy simulations. However, in the presented case the user is required to be heavily involved in the creation of the enriched model. Key performance indicator target values, such as required u-value ranges for windows, must be specified by the user manually in a graphical user interface.

Pauwels et al. (2011) demonstrate that other building performance evaluations such as acoustic performance can also benefit from semantic enrichment. They investigate the possibility of using an information description language and a rule language from the semantic web field to enhance the IFC for building performance checking. The suggested checking environment in this work consists of explicit building information, implicit building information and an inference engine. The inference engine is designed to leverage the existing (explicit and implicit) information to retrieve information required by the performance simulation platform. The documented test case in this work illustrates the implementation of semantic web technologies for an acoustic performance checking environment. This work provides an initial proof of concept with this test case, however further evaluation of the suggested approach is necessary.

Liu et al.(2016) explore an ontology-based semantic approach for quantity takeoff to enable construction practitioners to semantically query BIM design models and retrieve building quantity information from a construction perspective. Extracting construction-oriented quantity takeoff requires representation of measures taken off for construction activities based on activity definition and detailed specifications of construction methods and materials that are not usually specified in the BIM design model. In this work BIM models are enriched with distinguishable domain terms or classes to represent features of this type. They demonstrate that those implicit design features can be detected and then explicitly stored into an enhanced BIM model through topological analysis and ontology reasoning on a given BIM model. However, the ontology formalized in this work is in the particular context of light-frame building construction and still needs to explored and validated for other construction types.

6 DISCUSSION

Existing BIM platforms are able to maintain design integrity due to the use of parametric modeling and design intent behavior of the modeled objects (Sacks et al. 2004). However, BIM platforms cannot interpret information that is not explicitly represented in the model, as they lack the intelligence a domain expert has about the meaning of the represented topology, function and behavior. For example, dealing with inconsistencies in the mapping of BIM elements to IFC classes, Koo et al. (2019) illustrate the problem by describing a situation where a steel bearing plate is exported as an *IfcBeam* instead of the more accurate *IfcPlate*.

The underlying goal of semantic enrichment for BIM is to infuse the models with a higher level of "intelligence" than achieved through the use of parametric modeling. The ability to reason over the explicitly represented data to derive new facts about the model is equivalent to embedding the BIM platforms with knowledge about the physical world, thus making them more "intelligent". This cannot be achieved by simply linking data from external sources to model elements. In this case, although the semantics of the model at hand will be elevated, it will lack the ability to adapt and infer additional information.

Many research efforts focused on querying BIM to exploit the implicit information stored in a model (Borrmann et al. 2006; Mazairac and Beetz 2013; Wülfing et al. 2014). However, in previous work, inference results were reported to the user but not represented as part of the building information which prevents it from being used in downstream procedures. Although the ability to query models and infer new information is the backbone of semantic enrichment, to complete the process an explicit representation of the inferred information must be provided.

Based on these premises, we can define semantic enrichment of BIM as ***a process designed to automatically infer new information based on the information stored in a BIM model, and provide an explicit representation of that information to facilitate its use in a target application or procedure.***

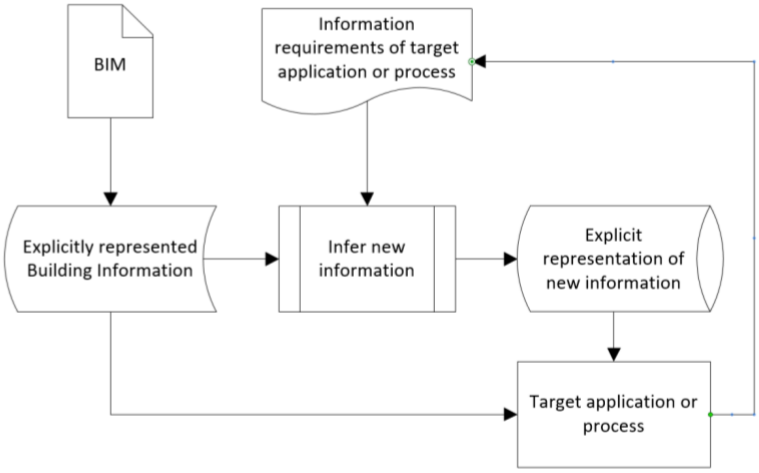


FIG. 5: The generic process of sematic enrichment of BIM models includes inference and explicit representation of new information based on IFC or RDF representation of the original information

The process of semantic enrichment of BIM based on the proposed definition is illustrated in Fig. 5. Two distinct approaches for enriching a BIM model were identified in this work. Although one approach aims at enriching a model in accordance to the IFC schema and the other explores the use of semantic web technology for representation of building information, both rely on an enrichment process described in Fig. 5. Regardless of the method for representing building information, the analyzed documents indicate that semantic enrichment of BIM models is a mean to provide easy communication of construction-related information between various environments to facilitate various procedures. Thus, semantic enrichment has been previously explored using both approaches for several possible applications. An overview of the possible applications identified in the collected documents is provided in Fig.6.

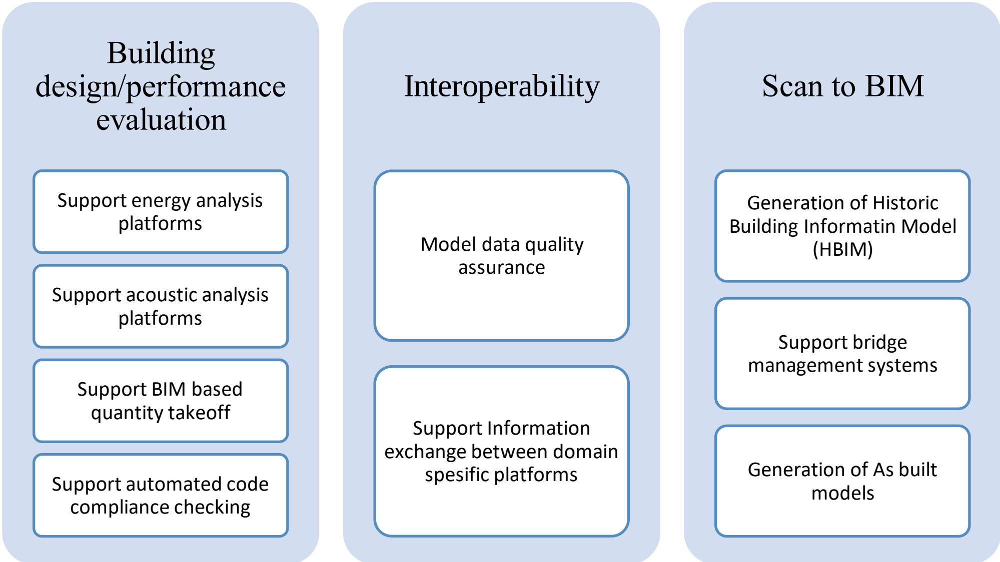


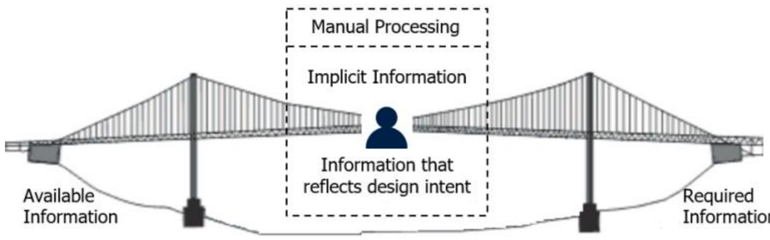
FIG.6: An overview of possible applications of semantic enrichment of BIM.

We can conclude that semantic enrichment is goal-driven process as the requirements for enrichment are defined by the intended use of the information. An important characteristic of semantic enrichment is that it consists of two, fully automated stages. The inference stage is designed to receive information from the source application, parse it using various computational methods to derive new facts about the model. In the second stage an explicit representation of the derived information is generated, creating a new and enriched BIM. The requirements for the inference engine are determined by the gap between the explicit information needed for the target application and the information provided in a source application.

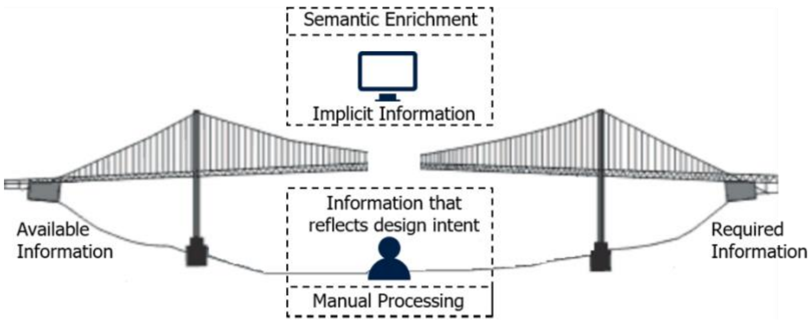
It is important to note that semantic enrichment of BIM is not a design tool and should not be used to add information that was not implied in the original models. As described in the work of Bloch and Sacks (2020), the gap between the information required by a target application and the information provided in the authoring tool consists of two information types: a) information that is implied in the model and can be inferred and supplemented in a sufficiently precise form to enable down stream applications, b) information that cannot be supplemented without making assumptions as to the modelers design intent. For example, we can rely on the spatial relationships between the model elements to determine which door in the building is the exit. This means that although the door may not be explicitly labeled as "Exit" in the model, this information exists in an implicit form. On the other hand, if the model contains a generic wall with no layers and material types explicitly assigned to it, we cannot infer any information about this walls' fire rating without making assumptions as to the design intent thus compromising the integrity of the design.

As illustrated in Fig. 7, in current practice, all information types are supplemented manually by a domain expert to enable the use of the information stored in the BIM model in various applications. Although the feasibility of partially automating this stage has been illustrated many times in previous research, an in-depth analysis and characterization of the information types to determine which are candidates for semantic enrichment and which must be supplemented by the domain expert is not available yet. Bloch and Sacks (2020) provide an initial framework for such an analysis and illustrate the procedure on information requirements drawn from applications for automated code compliance checking. However, their work is limited to a single possible application of semantic enrichment (automated code compliance checking) and requires further validation.

Evidently, the term semantic enrichment is used in the literature to refer to processes that also strive to enhance the semantics of a model, but greatly differ in the procedure to achieving set goal. Examples of such applications are provided in section 4.1.2 of this paper. Unlike semantic enrichment of BIM, these semantic enhancement procedures rely on external data sources instead of queries or inference engines. Based on the literature review, such enhancement procedures have been explored mostly for generation of semantically rich BIM models that capture the existing conditions of a building. Linking BIM elements with external data is fundamentally different from inferring new information based on the information originally available in the model. Although the two share a similar goal, semantic enrichment and semantic enhancement should be referred to as two distinct processes.



a) Current common practice of supplementing all required information by manual processing



b) The contribution of semantic enrichment to partially automate the processing stage

FIG. 7: Two approaches to dealing with the existing gap between the information available in the BIM generated in an authoring tool and the information required by receiving applications or procedures.

6.1 Limitations

Since an official and accepted definition of "semantic enrichment of BIM" is not available yet, we assume that some researchers refer to similar processes by terms other than “semantic enrichment of BIM”. Thus, we cannot guarantee that all relevant publications have been identified in this work. Also, due to the inaccurate terminology we cannot guarantee that all possible applications of semantic enrichment for BIM are identified. There might be additional domains that benefit from application of semantic enrichment for BIM that are not described in this work. Nevertheless, this review identifies and provides insight into the prominent and influential pieces of work that present and explore a variety of methods and possible applications of the process. The common features extracted from the collected publications serve as a firm base for compiling a comprehensive definition of “semantic enrichment of BIM”. Based on this work, we define semantic enrichment of BIM as a process designed to automatically infer new information based on the information stored in a BIM model, and provide an explicit representation of that information to facilitate its use in a target application or procedure.

6.2 Future research directions

Although a formal definition is fundamental for advancing in the field, many other concepts that make up this domain have not been defined and explored yet and should be considered as directions for future research. As mentioned in (Bloch and Sacks 2020) the characteristics of information that can be supplemented through semantic enrichment has not been thoroughly explored yet. Furthermore, the selection of appropriate tools and methods for enrichment should depend on the nature of the problem at hand, however only an initial framework for classifying enrichment tasks and selecting the most appropriate computational method is currently available (Bloch and Sacks 2020). This framework is both initial and limited in scope as it is focused on supporting only code compliance checking platforms.

Another issue that needs to be further explored is the accuracy and reliability of results obtained through semantic enrichment. Since semantic enrichment can be useful for pre-processing of the BIM models for various applications, such as code compliance checking for example, inaccurate enrichment results may lead to high-risk outcomes. Note that in any event, the possibility of human error always exists in manual processes. Further research is therefore needed to determine the overall reliability of semantic enrichment as a pre-processing tool for various applications, comparing the nature of errors made by human experts with possible errors in semantic enrichment and exploring the effect of the errors for downstream applications.

As there are currently no "of the shelf" platforms for semantic enrichment for BIM, this raises questions about the approach for development of such applications. One approach is to develop a stand-alone semantic enrichment engine able to produce an exchange file tailored for a group of domain specific tasks. For example, a semantic enrichment engine for energy analysis applications can be designed to fill the requirements of several platforms for energy analysis. The feasibility of such an approach needs to be further investigated as various applications may require the same concepts to be mapped in different ways to match their native data format. Another approach would be to limit the scope for enrichment to a very narrow domain by integrating the semantic enrichment engine within the receiving applications. Although development of such applications might be costly, software vendors should be motivated to provide users with more automated workflows.

7 CONCLUSIONS

Semantic enrichment has been defined in the context of various domains and used to support a variety of activities. Semantic enrichment of BIM models is a field of research that still lacks formal definition, a formalized process and scope of possible applications. In this work previous research related to semantic enrichment of BIM available on the SCOPUS database is identified and analyzed. This study relies on a bibliometric analysis of relevant publications to provides a description of the semantic enrichment process, possible approaches and previously explored computational methods for semantic enrichment of BIM.

Two types of bibliometric analysis were performed using the VOSviewer platform; coupling analysis and co-occurrence of keywords analysis. Results of the coupling analysis indicate that there are two major approaches for enrichment of BIM model that are defined by different representations of building information. One research direction is based on manipulating information stored in accordance to the IFC schema to enrich a model. Another approach is based on the use of Semantic Web technologies. Both approaches have been previously explored, demonstrated and applied for semantic enrichment of BIM models.

Results of the co-occurrence of keywords analysis point to the domains that may benefit from application of semantic enrichment for BIM. This includes capturing existing conditions of buildings and infrastructure, improving interoperability in multidisciplinary workflows and evaluation of building design and performance. Further analysis of the papers shows that semantic enrichment is applicable for generation of HBIM, as-built BIM and support of bridge management systems. In the context of improving interoperability, semantic enrichment of BIM is useful to ensure integrity of the data stored in the models and to support information exchange between domain specific platforms. It can also support design analysis routines such as energy analysis, acoustic analysis, quantity takeoff and automated code compliance checking.

With the advent and wide acceptance of the BIM technology in the AEC industry (Sacks et al. 2018a), there is a growing need for integration of BIM with various applications to enable effective collaboration based on a digital workflow throughout different organizations, disciplines, and project phases (Panteli et al., 2020). The use of BIM holds the potential to automate many processes thought different stages of construction projects such as code compliance checking, energy analysis, inspection of existing infrastructure, facility management. However, this requires extensive information exchange between various sources, some of which expect higher levels of semantic information to be explicitly represented in the models (Solihin et al. 2004; Torma 2013).

This led to efforts and initiatives targeting enhancement of semantic information represented in BIM. While the most primitive way of enhancing the semantics of a model is by manual processing, we should strive to reduce the manual time-consuming work, and achieve higher levels of automation. This work differentiates semantic enhancement of BIM and semantic enrichment of BIM. While the former relies on linking BIM elements to information stored in external data sources, the latter is concerned with retrieving new information from the original model through inferencing engines or query languages, and representing it in an explicit form.

This paper provides an in-depth description of the entire process of enriching BIM models, including previously explored computational methods, to facilitate progress toward industrial applications of semantic enrichment of BIM. The presented possible applications and contributions of the process to a BIM workflow provide motivation to engage in development of generic semantic enrichment routines for BIM models. The goal of such routines should not only be to enhance the semantic richness of a model, but also to keep the entire workflow on the highest possible level of automation.

This work illustrates the inconsistencies in the existing literature in terms of terminology. While some researchers describe semantic enrichment of BIM as a fully automated process (Belsky et al. 2016), others refer to linking models with external data bases as semantic enrichment (Mol et al. 2020; Mora et al. 2021). And most importantly, research aiming to enrich BIM models is not referred to as semantic enrichment. These inconsistencies hinder further development of this research area. Hence, the main contribution of this study is the definition of "semantic enrichment of BIM" as a process designed to automatically infer new information based on the information stored in a BIM model, and provide an explicit representation of that information to facilitate its use in a target application or procedure. The suggested definition is expected to formalize and establish the basis for the body of knowledge in this domain.

Finally, based on this review we can see common interests and goals between different research groups from two research areas. Although efforts have been made towards semantic enrichment of BIM using both research approaches (enrichment based on the IFC and enrichment based on semantic web technology), there are currently no research efforts to leverage the contributions of each research area towards these common goals. The expected contribution of semantic enrichment of BIM for many possible applications, as identified and described in this work, justify future multidisciplinary efforts for further development of semantic enrichment tools.

REFERENCES

Abanda, F. H., Tah, J. H. M., and Keivani, R. (2013). “Trends in built environment semantic Web applications: Where are we today?” *Expert Systems with Applications*, 40(14), 5563–5577.

Arenas, M., Gutierrez, C., and Pérez, J. (2009). “Foundations of RDF databases.” *Reasoning Web International Summer School*, Springer, 158–204.

Baumgärtel, K., and Scherer, R. J. (2014). “Automatic ontology-based green building design parameter variation and evaluation in thermal energy building performance analyses.” *Simulation*, 2012(27).

Beetz, J., Van Leeuwen, J., and De Vries, B. (2006). "Towards a topological reasoning service for IFC-based building information models in a semantic web context." *Proceedings of the Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, 3426–3435.

Beetz, J., Van Leeuwen, J., and De Vries, B. (2009). "IfcOWL: A case of transforming EXPRESS schemas into ontologies." *Ai Edam*, Cambridge University Press, 23(1), 89–101.

Belsky, M., Sacks, R., and Brilakis, I. (2013). "A framework for semantic enrichment of IFC building models." *Proceedings of 30th International Conference on Applications of IT in the AEC Industry CIB-W78*, 514–523.

Belsky, M., Sacks, R., and Brilakis, I. (2016). "Semantic enrichment for building information modeling." *Computer-Aided Civil and Infrastructure Engineering*, 31(4), 261–274.

Bloch, T., Katz, M., Yosef, R., and Sacks, R. (2019). "Automated model checking for topologically complex code requirements – security room case study." *Proceedings of the 2019 European Conference for Computing in Construction*, University College Dublin.

Bloch, T., and Sacks, R. (2018). "Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models." *Automation in Construction*, 91, 256–272.

Bloch, T., and Sacks, R. (2020). "Clustering Information Types for Semantic Enrichment of Building Information Models to Support Automated Code Compliance Checking." *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 34(6), 04020040.

Borrmann, A., Van Treeck, C., and Rank, E. (2006). "Towards a 3D spatial query language for building information models." *Proc. Joint Int. Conf. of Computing and Decision Making in Civil and Building Engineering (ICCCBE-XI)*.

Bouzidi, K. R., Fies, B., Faron-Zucker, C., Zarli, A., and Thanh, N. L. (2012). "Semantic web approach to ease regulation compliance checking in construction industry." *Future Internet*, Molecular Diversity Preservation International, 4(3), 830–851.

Bruno, S., De Fino, M., and Fatiguso, F. (2018). "Historic Building Information Modelling: performance assessment for diagnosis-aided information modelling and management." *Automation in Construction*, Elsevier, 86, 256–276.

"BuildingSMART." (2021). <http://www.buildingsmart-tech.org/> (Jul. 14, 2021).

Castellanos, M., and Saltor, F. (1991). "Semantic enrichment of database schemes: an object oriented approach." *Proceedings First International Workshop on Interoperability in Multidatabase Systems*, IEEE Computer Society, 71–72.

Costa, G., and Madrazo, L. (2015). "Connecting building component catalogues with BIM models using semantic technologies: an application for precast concrete components." *Automation in Construction*, 57, 239–248.

Costa, G., and Sicilia, A. (2020). "Alternatives for facilitating automatic transformation of BIM data using semantic query languages." *Automation in Construction*, Elsevier, 120, 103384.

Daum, S., and Borrmann, A. (2013). "Checking spatio-semantic consistency of building information models by means of a query language." *Proc. of theIntl Conference on Construction Applications of Virtual Reality*.

Eastman, C., Lee, J., Jeong, Y., and Lee, J. (2009). "Automatic rule-based checking of building designs." *Automation in construction*, 18(8), 1011–1033.

van Eck, N. J., and Waltman, L. (2020). *VOSviewer*. Leiden University's Centre for Science and Technology Studies (CWTS).

Fafalios, P., and Tzitzikas, Y. (2013). "X-ENS: semantic enrichment of web search results at real-time." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 1089–1090.

Fahad, M., Bus, N., and Fies, B. (2018). "Semantic BIM reasoner for the verification of IFC Models." *eWork and eBusiness in Architecture, Engineering and Construction*, CRC Press, 361–368.

Fichtner, F. W., Diakité, A. A., Zlatanova, S., and Voûte, R. (2018). “Semantic enrichment of octree structured point clouds for multi-story 3D pathfinding.” *Transactions in GIS*, Wiley Online Library, 22(1), 233–248.

Flach, P. A., and Kakas, A. C. (2000). “Abductive and Inductive Reasoning: Background and Issues.” *Abduction and Induction: Essays on their Relation and Integration*, P. A. Flach and A. C. Kakas, eds., Springer Netherlands, Dordrecht, 1–27.

Guo, D., Onstein, E., and Rosa, A. D. L. (2021). “A Semantic Approach for Automated Rule Compliance Checking in Construction Industry.” *IEEE Access*, 9, 129648–129660.

Hamledari, H., McCabe, B., Davari, S., and Shahi, A. (2017). “Automated Schedule and Progress Updating of IFC-Based 4D BIMs.” *Journal of Computing in Civil Engineering*, 31(4), 04017012.

Hamledari, H., Rezazadeh Azar, E., and McCabe, B. (2018). “IFC-Based Development of As-Built and As-Is BIMs Using Construction and Facility Inspection Data: Site-to-BIM Data Transfer Automation.” *Journal of Computing in Civil Engineering*, 32(2), 04017075.

Hichri, N., Stefani, C., De Luca, L., Veron, P., and Hamon, G. (2013). “From point cloud to BIM: a survey of existing approaches.” *XXIV International CIPA Symposium*, Proceedings of the XXIV International CIPA Symposium, na.

Hohenstein, U., and Plesser, V. (1996). “Semantic enrichment: A first step to provide database interoperability.” *Workshop Föderierte Datenbanken, Magdeburg*, 3–17.

Isailović, D., Stojanovic, V., Trapp, M., Richter, R., Hajdin, R., and Döllner, J. (2020). “Bridge damage: Detection, IFC-based semantic enrichment and visualization.” *Automation in Construction*, Elsevier, 112, 103088.

José de Oliveira, O., Francisco da Silva, F., Juliani, F., César Ferreira Motta Barbosa, L., and Vieira Nunhes, T. (2019). “Bibliometric Method for Mapping the State-of-the-Art and Identifying Research Gaps and Trends in Literature: An Essential Instrument to Support the Development of Scientific Projects.” *Scientometrics Recent Advances*, S. Kunosic and E. Zerem, eds., IntechOpen.

Jung, J., Hong, S., Yoon, S., Kim, J., and Heo, J. (2016). “Automated 3D Wireframe Modeling of Indoor Structures from Point Clouds Using Constrained Least-Squares Adjustment for As-Built BIM.” *Journal of Computing in Civil Engineering*, 30(4), 04015074.

Kessler, M. M. (1962). *An experimental study of bibliographic coupling between technical papers*. MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB.

Kim, M.-K., Wang, Q., Park, J.-W., Cheng, J. C. P., Sohn, H., and Chang, C.-C. (2016). “Automated dimensional quality assurance of full-scale precast concrete elements using laser scanning and BIM.” *Automation in Construction*, 72, 102–114.

King, R. D. (2011). “Rise of the robo scientists.” *Scientific American*, JSTOR, 304(1), 72–77.

Klein, L., Li, N., and Becerik-Gerber, B. (2012). “Imaged-based verification of as-built documentation of operational buildings.” *Automation in Construction*, 21, 161–171.

Koo, B., Jung, R., and Yu, Y. (2021). “Automatic classification of wall and door BIM element subtypes using 3D geometric deep neural networks.” *Advanced Engineering Informatics*, Elsevier, 47, 101200.

Koo, B., La, S., Cho, N.-W., and Yu, Y. (2019). “Using support vector machines to classify building elements for checking the semantic integrity of building information models.” *Automation in Construction*, 98, 183–194.

Koo, B., and Shin, B. (2018). “Applying novelty detection to identify model element to IFC class misclassifications on architectural and infrastructure Building Information Models.” *Journal of Computational Design and Engineering*, 5(4), 391–400.

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). “Supervised machine learning: A review of classification techniques.” *Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*, IOS Press.

Krijnen, T., and Tamke, M. (2015). “Assessing Implicit Knowledge in BIM Models with Machine Learning.” *Modelling Behaviour*, M. R. Thomsen, M. Tamke, C. Gengnagel, B. Faircloth, and F. Scheurer, eds., Springer International Publishing, Cham, 397–406.

Leroux, H., McBride, S., Lefort, L., Kemp, M., and Gibson, S. (2012). “A method for the semantic enrichment of clinical trial data.” *Health Informatics: Building a Healthcare Future Through Trusted Information: Selected Papers from the 20th Australian National Health Informatics Conference (HIC 2012)*, IOS Press, 111.

Li, X., Shen, G. Q., Wu, P., and Yue, T. (2019). “Integrating Building Information Modeling and Prefabrication Housing Production.” *Automation in Construction*, 100, 46–60.

Liebich, T., and Wix, J. (1999). “Highlights of the development process of industry foundation classes.” *Proceedings of the 1999 CIB W78 Conference*, 18.

Liu, H., Lu, M., and Al-Hussein, M. (2016). “Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry.” *Advanced Engineering Informatics*, 30(2), 190–207.

Lomio, F., Farinha, R., Laasonen, M., and Huttunen, H. (2018). “Classification of building information model (BIM) structures with deep learning.” *2018 7th European Workshop on Visual Information Processing (EUVIP)*, IEEE, 1–6.

Macher, H., Landes, T., and Grussenmeyer, P. (2017). “From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings.” *Applied Sciences*, 7(10), 1030.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., and López-Cózar, E. D. (2018). “Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories.” *Journal of informetrics*, Elsevier, 12(4), 1160–1177.

Mazairac, W., and Beetz, J. (2013). “BIMQL An open query language for building information models.” *Advanced Engineering Informatics*, 27(4), 444–456.

Mol, A., Cabaleiro, M., Sousa, H. S., and Branco, J. M. (2020). “HBIM for storing life-cycle data regarding decay and damage in existing timber structures.” *Automation in Construction*, Elsevier, 117, 103262.

Mongeon, P., and Paul-Hus, A. (2016). “The journal coverage of Web of Science and Scopus: a comparative analysis.” *Scientometrics*, Springer, 106(1), 213–228.

Mora, R., Sánchez-Aparicio, L. J., Maté-González, M. Á., García-Álvarez, J., Sanchez-Aparicio, M., and Gonzalez-Aguilera, D. (2021). “An historical building information modelling approach for the preventive conservation of historical constructions: Application to the Historical Library of Salamanca.” *Automation in Construction*, Elsevier, 121, 103449.

Panteli, C., Kylili, A., and Fokaides, P. A. (2020). “Building information modelling applications in smart buildings: From design to commissioning and beyond A critical review.” *Journal of Cleaner Production*, 265, 121766.

Pauwels, P., and Oraskari, J. (2016). “IFC-to-RDF-converter, 2016.” *URL https://github. com/mmlab/IFC-to-RDF-converter. Last update on*, 30.

Pauwels, P., and Terkaj, W. (2016). “EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology.” *Automation in Construction*, 63, 100–133.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., De Meyer, R., Van de Walle, R., and Van Campenhout, J. (2011). “A semantic rule checking environment for building performance checking.” *Automation in Construction*, 20(5), 506–518.

Pauwels, P., and Zhang, S. (2015). “Semantic rule-checking for regulation compliance checking: An overview of strategies and approaches.” *32rd international CIB W78 conference*.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). “Semantic web technologies in AEC industry: A literature overview.” *Automation in Construction*, 73, 145–165.

Preidel, C., and Borrmann, A. (2015). “Automated Code Compliance Checking Based on a Visual Language and Building Information Modeling.” *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, Vilnius Gediminas Technical University, Department of Construction Economics & Property, 1.

Romero, S., and Becker, K. (2017). “Improving the classification of events in tweets using semantic enrichment.” *Proceedings of the International Conference on Web Intelligence*, 581–588.

Sacks, R., Bloch, T., Katz, M., and Yosef, R. (2019). “Automating Design Review with Artificial Intelligence and BIM: State of the Art and Research Framework.” *Computing in Civil Engineering 2019*, American Society of Civil Engineers, Atlanta, Georgia, 353–360.

Sacks, R., Eastman, C., Lee, G., and Teicholz, P. (2018a). *BIM handbook: A guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*. John Wiley & Sons, Hoboken, NJ.

Sacks, R., Eastman, C. M., and Lee, G. (2004). “Parametric 3D modeling in building construction with examples from precast concrete.” *Automation in construction*, 13(3), 291–312.

Sacks, R., Kedar, A., Borrmann, A., Ma, L., Brilakis, I., Hüthwohl, P., Daum, S., Kattel, U., Yosef, R., Liebich, T., Barutcu, B. E., and Muhic, S. (2018b). “SeeBridge as next generation bridge inspection: Overview, Information Delivery Manual and Model View Definition.” *Automation in Construction*, 90, 134–145.

Sacks, R., Ma, L., Yosef, R., Borrmann, A., Daum, S., and Kattel, U. (2017). “Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry.” *Journal of Computing in Civil Engineering*, 31(6), 04017062.

Sadeghineko, F., and Kumar, B. (2020). “Development of Semantically Rich 3D Retrofit Models.” *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 34(6), 04020039.

Santagati, C., and Lo Turco, M. (2016). “From structure from motion to historical building information modeling: populating a semantic-aware library of architectural elements.” *Journal of Electronic Imaging*, 26(1), 011007.

Sibenik, G., and Kovacic, I. (2021). “Interpreted open data exchange between architectural design and structural analysis models.” *Journal of Information Technology in Construction (ITcon)*, 26(4), 39–57.

Simeone, D., Cursi, S., and Acierno, M. (2019). “BIM semantic-enrichment for built heritage representation.” *Automation in Construction*, Elsevier, 97, 122–137.

Snyder, H. (2019). “Literature review as a research methodology: An overview and guidelines.” *Journal of Business Research*, 104, 333–339.

Solibri. (2017). “Solibri Model Checker (SMC).” <https://www.solibri.com/> (Mar. 13, 2017).

Solihin, W., Shaikh, N., Rong, X., and Poh, K. L. (2004). “Beyond interoperatibility of building model: A case for code compliance checking.” *BP-CAD Workshop*, Carnegie Mellon University.

Songa, J., Kima, J., and Leea, J.-K. (2019). “Spatial Information Enrichment using NLP-based Classification of Space Objects for School Bldgs. in Korea.” *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, IAARC Publications, 415–420.

Sternberg, R. J., Sternberg, K., and Mio, J. S. (2012). *Cognitive psychology*. Wadsworth/Cengage Learning, Belmont, CA.

Terkaj, W., and Šojić, A. (2015). “Ontology-based representation of IFC EXPRESS rules: An enhancement of the ifcOWL ontology.” *Automation in Construction*, 57, 188–201.

Todd, B. S. (1992). *An introduction to expert systems*. Technical monograph PRG, Oxford Univ. Computing Laboratory, Programming Research Group, Oxford.

Torma, S. (2013). “Semantic Linking of Building Information Models.” *2013 IEEE Seventh International Conference on Semantic Computing*, IEEE, Irvine, CA, USA, 412–419.

Valêncio, C. R., Martins, R. D., Marioto, M. H., Corrêa, P. L. P., and Babini, M. (2013). "Automatic knowledge extraction supported by semantic enrichment in medical records." *2013 International Conference on Parallel and Distributed Computing, Applications and Technologies*, Ieee, 79–83.

Van Eck, N. J., and Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping." *scientometrics*, Springer, 84(2), 523–538.

Venugopal, M., Eastman, C. M., and Teizer, J. (2015). "An ontology-based analysis of the industry foundation class schema for building information model exchanges." *Advanced Engineering Informatics*, 29(4), 940–957.

Werbrouck, J., Pauwels, P., Bonduel, M., Beetz, J., and Bekers, W. (2020). "Scan-to-graph: Semantic enrichment of existing building geometry." *Automation in Construction*, Elsevier, 119, 103286.

Wu, J., and Zhang, J. (2018). "Automated BIM object classification to support BIM interoperability." *Construction Research Congress 2018: Sustainable Design and Construction and Education*.

Wu, J., and Zhang, J. (2019). "New automated BIM object classification method to support BIM interoperability." *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 33(5), 04019033.

Wülfing, A., Windisch, R., and Scherer, R. (2014). "A visual BIM query language." *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2014*, 157.

Xu, Z., Kang, R., and Lu, R. (2020). "3D reconstruction and measurement of surface defects in prefabricated elements using point clouds." *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 34(5), 04020033.

Xue, F., Lu, W., and Chen, K. (2018). "Automatic Generation of Semantically Rich As-Built Building Information Models Using 2D Images: A Derivative-Free Optimization Approach: Automatic generation of semantically rich as-built building information models using 2D images." *Computer-Aided Civil and Infrastructure Engineering*, 33(11), 926–942.

Xue, F., Lu, W., Chen, K., and Webster, C. J. (2019). "BIM reconstruction from 3D point clouds: A semantic registration approach based on multimodal optimization and architectural design knowledge." *Advanced Engineering Informatics*, Elsevier, 42, 100965.

Yan, Y., and Hajjar, J. F. (2021). "Automated extraction of structural elements in steel girder bridges from laser point clouds." *Automation in Construction*, Elsevier, 125, 103582.

Yin, X., Liu, H., Chen, Y., and Al-Hussein, M. (2019). "Building information modelling for off-site construction: Review and future directions." *Automation in Construction*, 101, 72–91.

Yurchyshyna, A., and Zarli, A. (2009). "An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction." *Automation in Construction*, Elsevier, 18(8), 1084–1098.

Zeng, M. L. (2019). "Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article." *El profesional de la información*, Ediciones Profesionales de la Información SL, 28(1).

Zhang, C., Beetz, J., and de Vries, B. (2018). "BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data." *Semantic Web*, IOS Press, 9(6), 829–855.

Zhang, L., and Issa, R. R. A. (2013). "Ontology-Based Partial Building Information Model Extraction." *Journal of Computing in Civil Engineering*, 27(6), 576–584.

Zhong, B., Ding, L., Luo, H., Zhou, Y., Hu, Y., and Hu, H. (2012). "Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking." *Automation in Construction*, Elsevier, 28, 58–70.

Zhong, B. T., Ding, L. Y., Love, P. E. D., and Luo, H. B. (2015). "An ontological approach for technical plan definition and verification in construction." *Automation in Construction*, 55, 47–57.