2017

A review of approaches to solving the problem of BIM search: towards intelligence-assisted design

Hamed Khademi
d16128679@mydit.ie

Avril Behan
Technological University Dublin, avril.behan@tudublin.ie

Recommended Citation

Khademi, H. & Behan, A. 2017 "A review of approaches to solving the problem of BIM search: towards intelligence-assisted design". *Proceedings CitA BIM Gathering 2017*, Croke Park, Dublin, Ireland, 23rd-24th November 2017. doi:10.21427/sged-qg40

A review of approaches to solving the problem of BIM search: towards intelligence-assisted design

Hamed Khademi[1] and Avril Behan[2]

School of Multidisciplinary Technologies

Dublin Institute of Technology, Dublin

E-mail: [1]hamed.khademi@mydit.ie [2]avril.behan@dit.ie

Abstract − Due to the growing adoption of BIM and the rising popularity of cloud computing, BIM models are increasingly stored in central cloud repositories or Common Data Environments. Effective management and exploitation of these models creates the requirement for BIM retrieval systems. Thus far, the BIM industry has utilized general-purpose, text-based search techniques that operate on BIM metadata. This paper highlights the need for a domain-specific BIM search engine and reviews various approaches to address the problem of BIM search. Three main approaches were identified as context-, geometry-, and content-based BIM retrieval. For a comprehensive BIM retrieval system, all three approaches need to be utilized. Literature about geometry- and content-based retrieval was scarce, and about context-based retrieval was almost non-existent. Context-based retrieval is a special approach that is relevant here due to the project-based and goal-oriented nature of architectural design and needs support from stakeholders in the AECO industry.

Keywords − Building Information Modelling (BIM), Information Retrieval, Machine Learning, Information Seeking

I Background

The number of BIM models created worldwide is growing rapidly due to increases in BIM adoption in recent years. With easy and economic availability of cloud storage and with the rising popularity of cloud computing, these models are increasingly being stored in either private or public central repositories, which in turn creates the requirement for BIM retrieval systems. Thus far, the BIM industry has utilized general-purpose, keyword-based search techniques for BIM search [1]. To develop high performance Information Retrieval (IR) systems, general-purpose IR approaches should be appropriately adapted for the BIM domain [1]. Domain-specific search engines are technically referred to as 'vertical search engines'. They focus on one area of knowledge which gives them some advantages including: greater precision due to limited scope, leveraging domain knowledge including taxonomies and ontologies, and support for specific unique user tasks [2].

Most publicly available BIM retrieval approaches rely on metadata (e.g. keywords, tags, descriptions) which in turn are dependent on annotation quality and completeness [1]. Moreover, manual annotation of BIM models is not practical for large databases and may not capture the correct keywords to describe the models for the diverse types of user queries. Similar issues existed in metadata-based image search (a.k.a. text / concept / description-based image indexing / retrieval) which, in time, was solved by Content-Based Image Retrieval (CBIR) approaches [3]. The authors believe that the same philosophy can be applied to improve BIM retrieval quality by going beyond metadata and taking into account data/contents stored in BIM models during the indexing process (indexing is a process in information retrieval systems that includes collecting, parsing, and storing data to facilitate fast and accurate information retrieval). There is some research that considered this approach, however, most focused on BIM 'products/objects' rather than viewing building models as a whole/system [1], [4]. In contrast, the goal of this research is to retrieve BIM models based on e.g. site location (latitude, longitude, elevation), space functions, building envelop shape and properties, aggregate quantities of elements, etc. In addition to content-based BIM retrieval, this research also tries to answer the question of whether it is possible to retrieve BIM models based on the context for which they were originally designed.

This paper aims 1) to demonstrate the industry-wide need for a domain-specific BIM retrieval engine and

2) to investigate various approaches to address the problem of BIM search.

II Why Do We Need BIM retrieval system?

In the following, three major applications of BIM retrieval systems are presented which we hope will attract the attention of stakeholders to support the growth of this research area.

Design Recycling

Building models designed for similar contexts and requirements as those of the project at hand can serve as a knowledge repository and a source of inspiration and solution for current design problems [5], [6]. This is especially the case in early stages of design [7]. Design reuse is a topic that has been discussed in several papers and textbooks. Reusing design helps to find solutions quicker and avoid 're-inventing the wheel' which can save project resources and, in turn, project costs [8]. Here, we use the term 'design recycle' to emphasise that the process of reusing previous designs, in most cases, is partially possible. The goal is not to copy previous designs identically, but rather to seek inspirations, to reuse design intentions, or to reuse a subsystem (of a whole building system), etc.

One of the arguments for the possibility of design recycling is based on Case-Based Reasoning (CBR). The basic premise in CBR is that similar problems have similar solutions. In CBR, a case consists of a problem description and a solution description. Aamodt and Plaza described the whole process within the CBR cycle using the verbs retrieve, reuse, revise and retain [9]. The approach of applying CBR to design and architectural tasks is known as Case-Based Design (CBD) and has been extensively researched both within and outside the domain of Architecture [10]

Knowledge Management

Organizations have recognized that knowledge constitutes a valuable intangible asset for creating and sustaining competitive advantage [11]. Creation of knowledge in great volumes and their storage as information, in turn, creates the problem of 'information overload', which is increasingly recognized and documented [12]. Knowledge management systems and automated information retrieval systems have been developed to address this issue. Knowledge management (KM) is the process of capturing, sharing, reusing and maintaining the knowledge and information of an organization [13]. KM is a multidisciplinary approach to achieving organizational objectives by making the best use of accumulated knowledge [14].

The UK Department of the Environment (DOE) has funded a feasibility study into the concept of a knowledge base for the construction industry to achieve multiple objectives. Of these, the objectives relevant to this study are: 1) improving the quality and efficiency of buildings and building projects by sharing information on standards and best practice; 2) improving the efficiency of the construction market by facilitating market communications; 3) reducing the cost and improving the quality of building design by sharing design knowledge [15].

In one view, BIM models can be seen as knowledge repositories as it takes experience, knowledge, research and skills to create them [16]. This argument is strengthened with the widening scope of BIM beyond design stage to address the information needs in construction (e.g. 4D and 5D BIM) and operations phase (e.g. more Ds [17]) [18]. Significantly, efforts have also been made to capture energy-related knowledge into BIM [19].

Research shows that engineers spend a great deal of time in searching for information [20]–[22]. In doing so, accessibility of information is one of the most influential factors in choosing information sources [23]–[27]. When BIM models produced in an organization are stored centrally and made accessible to other members of the organization, they turn into organizational assets and create a 'corporate memory'; and when these knowledge repositories grow in size, searching for information efficiently becomes an extremely critical function [28].

Kaizen (Continuous Improvement)

Nearly half-a-century ago, in her study of playgrounds, Lady Allen of Hurtwood wrote: "Why so many expensive mistakes ... made over and over again? One reason may be that there is no central body whose job it is to collect experience and research throughout the world, digest it, and make it readily available to architects and planners" [cited in [29]]. A casual observer may think that this problem has already been solved with today's technology since all the hardware ingredients to realize this idea has been available for at least two decades. Yet, this simple idea remains unrealized. Some have suggested to create a central repository for building models and complement it with Post-Occupancy Evaluations (POE) [30]. These central repositories do exist today for BIM models as a result of BIM collaboration cloud services offered by BIM collaboration software developers such as Autodesk Inc, Graphisoft and Trimble Inc. However, POE studies are rare. There are alternative methods for measuring the quality of BIM models, e.g. certificates and energy ratings [31], user feedback from Facility Management (FM) systems, data from emerging IoT systems, building data exhaust [32], expert user ratings, performance analysis results, etc. Once, successes and failures of buildings from various aspects are identified using the aforementioned methods, this data can be associated with BIM, either embedded in the model or as linked data. The next step is to provide an information retrieval system based on the quality of building projects so that the future designers can build on successes and avoid mistakes. In the long run, this practice can contribute to the generation of architectural knowledge and speed up the cycle of innovation.

III Related Previous Work

Previous research on the issues core to the problem of BIM search, such as context-based or content-based retrieval over a large collection of BIM models, is limited. Some of the previous works utilize natural language processing or other machine learning and information retrieval techniques to improve querying and searching within a single BIM model or across a collection of BIM 'objects/products' rather than over a collection of BIM 'models'. Lin et al. [33] have proposed a method for querying information from within a single BIM model using natural language to make BIM querying more user-friendly for non-experts. Other researchers worked on improving information retrieval on BIM 'objects' by 1) enhancing semantic annotation of documents [4]; and 2) enhancing the user query mechanism [1], [34].

Regarding design reuse, information retrieval techniques have been used on (mostly CAD-based or text-based) civil engineering documents [35]. A line of research on design knowledge management conducted at the Project-Based Learning Lab at Stanford University is of particular interest here. This line of research was later continued at Loughborough University [36], [37]. These projects were based on Schön's reflective practitioner paradigm of design [38]. First in line was a Semantic Modeling Engine (SME) which is a framework that enables designers to map objects from a shared CAD product model to multiple semantic representations and to other shared project knowledge [39]. Then, the ProMem (Project Memory) system was developed which complements SME by adding a time dimension using a version control system [40]. ProMem captures the evolution of the project at three levels of granularity identified by SME as emulating the structure of project knowledge: project, discipline, and component. In turn, ProMem was extended in two ways to develop CoMem (Corporate Memory), which aims at 1) grouping an accumulated set of project memories/knowledge into a corporate memory/knowledge, and 2) supporting the designer in reusing design knowledge from this corporate memory in new design projects [37]. The approach used in CoMem for information retrieval was visualization of an entire repository of design content. Later, CoMem was upgraded to CoMem XML to provide support for query-driven search in addition to visual exploration and discovery [41]. In CoMem XML, the focus was on information and documents linked to buildings and building models, rather than the knowledge embedded in the design and construction processes and the professionals performing those processes.

Another interesting line of research was conducted by a team at TU Munich that investigate the retrieval of BIM models from a repository based on topological relationships between spaces [7], [42]. The authors argued that such a retrieval system could be used for finding design inspiration in early design stages. It can be argued that this type of retrieval system can also be beneficial for space planning in later stages of design.

IV Indexing Depth

Before going into the subject of 'how' to develop and implement a BIM retrieval system, we first discuss 'what' it is that need to be indexed. The concept of 'indexing depth' is introduced at four levels of 1) metadata, 2) data 3) information extraction, and 4) domain knowledge incorporation. To go further in 'depth', more data processing is required; however, this effort can pay off by improving information retrieval performance (see Figure 1).

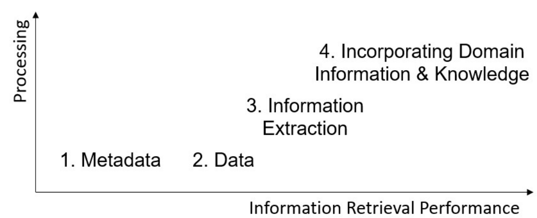

Figure 1: Four levels of indexing depth shown against 'data processing' and 'information retrieval performance'

Metadata

Metadata is "data that provides information about other data" [43]. In the case of BIM models, metadata may refer to file title and description, owner of BIM file, creation and modification dates, or any other associated information that is provided for document management purposes. BIM retrieval systems use this data to index them and later on to retrieve them by matching user queries against them. Since metadata is poor in content that matters for designers and engineers, BIM authors may be asked to fill in description attributes for BIM files with potentially useful information. This is, of course, not a practical solution. Nevertheless, this is representative of the current state of BIM search in the industry; generic text-based search on BIM metadata.

Data

Although, by definition, all data stored in Building Information Models (BIM) is 'information', here, the term 'data' is used to refer to the contents in BIM models in raw form. To search on BIM data, only selectors and filters would be used without complex processing. Data in BIM models can be extracted using BIM query languages and indexed for later information retrieval needs. This data can be as simple as 'site location', which exists as an explicit attribute in most BIM formats, or it can be a little more complex such as the total number of floors or rooms. In either case, BIM query languages that function similar to database query languages suffice to extract this data. This has been done at BIM 'object' level to some extent in bimobject.com's object search. [1], [4], [34] are some of the research projects on improving BIM 'object' search.

Information Extraction

In this level, some processing is required to extract information that is not readily retrievable using BIM query languages; e.g. geometrical shape of the building envelop or total surface area of the envelope. These examples may seem simple, however, at least in the case of IFC BIM format, they can become tricky to achieve with good accuracy. The underlying problem is that there is no built-in mechanism to validate the information entered by the BIM author [44]. For example, a wall can be an external wall yet it may have been flagged as internal. Over recent years, high-level BIM query languages have been emerging that can facilitate information extraction more reliably. For example, QL4BIM can facilitate extraction of spatial information based on geometrical information rather than semantic information [45].

Incorporating Domain Information and Knowledge

If the domain information and knowledge is incorporated properly, it can improve search performance [34]. Incorporating domain knowledge for the purpose of information retrieval requires significant inputs beyond the contents embedded in BIM models. In a hypothetical and likely-to-happen search query scenario, a designer may seek BIM models that are designed for similar climate profiles as the climate profile of the project at hand to potentially recycle previous designs or to get some inspiration. The traditional way to achieve this would consist of four steps: 1) extracting site location; 2) finding the climate profile of the nearest weather station to the site location; 3) finding all locations (weather stations) in the world that have similar climate profiles to that of the project in question (which is not a trivial problem); and 4) finding a BIM retrieval system that can filter models based on multiple locations. It is evident that while these steps are doable, they are not practical for repeated use due to their complexity and time-intensiveness. To abstract away the intermediary steps, BIM retrieval engines should have climate profile 'information' for various locations in the world, and should have the 'knowledge' to calculate similarity between these climate profiles. In other words, meteorological information and knowledge from an architectural engineering point of view should be incorporated into the BIM retrieval engine. Similar examples could be given for other design purposes, e.g. solar radiance intensity and angles for lighting design, soil properties for foundation engineering, seismic activities and wind speed for structural engineering, and urban and cultural context for façade design. Incorporation of domain information and knowledge related to architecture in a BIM retrieval engine may be done simply with arithmetic calculations or it may require complex machine learning algorithms depending on the area of interest.

Linked External Data

The scope of the previously discussed four levels of BIM retrieval was limited to the BIM models alone. BIM models can also be linked to external data whether systematically [19], [46], [47] or using general-purpose document or knowledge management systems. Some have proposed using the Sematic Web/Linked Data platform as an alternative to IFC to provide a ubiquitous machine-readable format [48]. If associating BIM data to other data is done properly, the process of indexing them for IR purposes would be possible in an automatic manner (this would be the case if Linked Data was the BIM platform since this platform is made to be machine-readable). For other formats like IFC, machine learning techniques need to be utilized for extracting information from associated documents [49]–[51]. These documents could be project goals and requirements/constraints (such as Employers and Organizational Information Requirements as per the PAS1192 suite of standards), engineering calculations that support design decisions, legal and regulatory documents, certificates and approvals, etc. In recent years, COBie is gaining traction in the AECO industry as a standard for delivery of facility management information. If COBie data is associated with (or embedded in) BIM models, it can be used for indexing and retrieval of BIM models.

As discussed earlier, one of the purposes of BIM search is to enable continuous improvement by building on best practice. To identify best practice, a means to judge the quality of design is needed. One approach is to source this information from BIM model-associated documents such as instrument measurements and user feedback gathered via POE studies, simulation results, or building certifications. Search users can exploit such data to judge the quality of BIM models by filtering or prioritizing models meeting minimum performance criteria. Figure 2 illustrates a comprehensive design-oriented BIM search based on combined BIM and POE data.

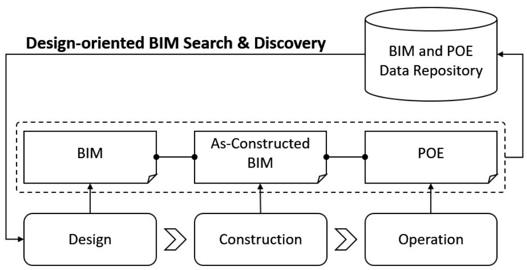

Figure 2: The proposed design-oriented BIM search and discovery based on BIM and POE data repository

V Approaches to BIM Search

To give structure to this paper as well as the wider BIM search project being carried out by the authors, the natural evolutionary design process was considered in categorizing and prioritizing the problem areas.

In the conceptual design stage, before creation of any models, requirements and constraints (such as functions, site location and climatic parameters, budget limits, regulations [52]) that the designer needs to meet are at their minimum and the nature of these constraints are mostly contextual to the building itself. The purpose of BIM search at this stage may be for finding inspiration rather than solutions. At this point, the designer needs to formulate the queries based on these contextual constraints. We call this context-based BIM retrieval. After the creation of a conceptual model, approximate geometrical boundaries (e.g. envelope shape, space adjacency and accessibility relationship as discussed by [53], etc.) of the building are determined. From this point on, the designer may need to add this approximate geometry to the pool of contextual constraints during search operation. We call this geometry-based BIM retrieval. In the design development stage, as the design progresses, more details of building elements are determined such as façade system, materials, and HVAC systems. At this stage, it may be necessary to add constraints to the search query based on these partially-determined design details. We denote this as content-based BIM retrieval.

In a search operation instance, one or any combination of these three types of search (context-, geometry- and context-based retrieval) may be used depending on the specificities of the design problem at hand.

Context-Based BIM Retrieval

Research shows that engineers working with BIM models place particular importance on understanding of retrieved content before using it or applying it, and that exploration of context is essential for this understanding [41]. This is logical because buildings are designed 'for' satisfying the contextual constraints of projects [52]. Hence, if a BIM retrieval engine can utilize 'contextual information' in indexing BIM models, it would allow designers and engineers to perform searches based on high-level project goals/intentions and constraints rather than merely matching the low-level non-intentional contents stored in BIM models against non-intentional user queries (refer to [54] for further clarity on terminology). To the best of our knowledge, there is only a limited amount of research on the subject of context-based retrieval in other domains and even less so on BIM search. In Web search, Strohmaier et al. argue that due to the lack of explicit intentional structures and representations on the Web, search engines cannot associate users' goals with the Web contents [54]. This incurs a cognitive load on users to translate their high-level goals into the non-intentional structure of the Web in the forms of specific search queries, tag concepts, classification terms or ontological vocabulary. For product search, some websites provide goal-oriented exploration of their catalogs e.g. when shopping for a laptop, it is possible to search based on application of the laptop (gaming, design, office, etc.) which frees the users from figuring out what laptop specifications (processing power, memory, display quality, etc.) would satisfy their computing needs. Similar to the Web, BIM models do not capture goals, intentions or constraints for which the building is designed, although, there has been some efforts at research level to extend BIM to include contextual 'information' for design automation and reuse purposes [55]. Until such efforts find their way into practice, alternative solutions need to be sought. One solution is extracting contextual information from associated project documents (see "Linked External Data" subsection above). The disadvantages of this solution are that 1) it relies on additional project documents which may not always be available; and 2) it is error prone and would require manual supervision to achieve acceptable accuracy. The second solution is following the same path that the designer/engineer may have taken throughout the design process. This can be done automatically in some cases. For an example structural design project, given a project site location, wind speed or precipitation, information can be found from national or worldwide databases (domain information). In turn, BIM models in the repository can be indexed according to this contextual domain information using simple or sophisticated algorithms depending on the complexity of the information.

We identified the main contextual constraints of building projects as described in Table 1. Some of the items in the table can only be obtained from associated project documents (e.g. most items in the project category), some others can be found from national or worldwide databases (e.g. items in the climate category), and some others can be inferred from other known parameters (e.g. mandatory standards are defined by the state, which can be determined from site location).

Geometry-Based BIM Retrieval

Although geometry is part of the BIM content, it is categorized separately, because, the visual aspect of geometry sets it apart from text-based search from several aspects including query modality, result visualization, indexing, similarity assessment, implementation technologies, and distinctness.

Table 1: Main subcategories of context-based retrieval

Category	**Items**
Site	• Shape • Orientation • Ground Properties • Latitude (GIS)
Urban	• Fabric • Social • Utilities • Amenities
Climate	• Temperature • Humidity • Wind speed • Precipitation • Sky conditions
Project	• Users • Functions & Activities • Budget • Time • Sustainability
Market	• Material • Construction Technology
Regulations	• Mandatory Standards • Optional Certificates

Geometry is one of the most important design factors in building design projects, especially the envelop geometry, which often times is constrained by the dimensions of the designated construction site. Envelop geometry has critical impacts on the performance of most of the building systems such as lighting, HVAC, structural and architectural systems. In a needs analysis study for BIM component retrieval within a single model, Demian et al. found that users need the ability to search and visualize the results in both textual and graphical mode as well as the relationship between components (a.k.a. topology) [56]. Based on this, they developed two prototype BIM 'object' retrieval system; one based on pure geometry and another based on the combination of geometry and topology. User evaluations showed that the former outperformed the latter in all the ten questions answered by the users mostly due to the inconvenience caused by the added complexity of topology-based retrieval. Langenhan et al. proposed a new concept, namely, semantic fingerprint to capture the topological relationships between spaces in terms of adjacency and accessibility [57]. Later, they built on this concept by modeling this topological fingerprint as a graph and developed a graph-based BIM retrieval system [7], [42].

Although, geometry-based retrieval has challenging aspects due to its non-textual modality, it has positive sides as well. Firstly, due to its importance, geometrical data in BIM models are more reliable than semantic contents [53]. Secondly, sufficient research and tools for geometry-based retrieval are available for use and this can speed up the development of a BIM retrieval engine [53], [58]. Table 2 shows various subcategories of geometry-based retrieval along with references to notable research. It is worth mentioning that well-developed commercial pure 3D (i.e. excluding topology) search engines already exist in the market (e.g. yobi3d.com).

Table 2: Main subcategories of geometry-based retrieval

Categories	**Items**
Graphical search (shape, dimensions & orientation)	• 2D [59] • 3D [58]
Topological search (space composition) [7], [42], [53]	• Space set • Space adjacency • Space accessibility
Combined graphical and topological [56]	• Taking into account all the above items

Content-Based BIM Retrieval

Content-based BIM retrieval has been researched and commercialized for other types of media including both non-structured media such as text, images, and music, and for structured media such as product data that is stored in databases. BIM 'objects' can also be searched using product/database search technologies because they usually have simple data structures. Bimobject.com is an example of a commercial BIM object search engine, which utilizes contents of BIM objects to some extent. However, the same technologies cannot be easily adapted for retrieval of BIM models because BIM models are usually made up of systems of complex compositions of simpler BIM objects that are associated with high-levels (design) concepts. We speculate that designers would mostly be interested in querying BIM models based on these high-level concepts rather than low-level raw data stored in BIM models. This requires interpreting and understanding complex compositions of BIM objects by the search engine. For example, window-to-wall ratio, a rather high-level concept, is kept low for designing passive houses in cold climates. To index models based on this characteristic, one needs to 1) extract the windows and walls of the building model envelop; 2) extract the area of these windows and walls; 3) calculate the ratio; and 4) index the model according to this ratio.

Content-based retrieval can be seen at three levels of granularity as objects, systems, and building levels. At object and system levels, users can search for building models that 'contain' objects or systems with specific properties, while at building level, it is the building itself as a whole that should meet search criteria defined by the user. Table 3 shows subcategories of content-based retrieval at systems and building level, which are partially based on the PhD work of Ajla Aksamija (refer to Appendix A of her PhD dissertation for further details [55]).

Table 3: Main subcategories of content-based retrieval at systems and building levels

Categories	Items
Envelope energy efficiency (e.g. dimensions & form, material, thermal properties)	• Window-to-wall ratio • Window (glazing & frame) • Exterior wall • Roof • Bottom floor/slab • Facade
HVAC systems	• Heating & cooling source (district, heat pump, fuel, el., etc.) • HVAC equipment efficiency • Air leak • Occupancy control (IoT) • Operating schedule (typical hours of use by occupants) • Lighting efficiency (internal heat gains and power consumption) • Heat from equipment • Overall heating and cooling quality and performance
Lighting systems	• Daylighting Efficiency • Occupancy control (IoT) • Operating schedule (typical hours of use by occupants) • Shading System • Overall performance and quality
Structural systems	• Material (steel, concrete, wood, mixed, etc.) • Structural systems (frame, truss, etc.) • (Min, max, mean) span • Foundation properties (material, type, etc.)
Building	• Performance (standards and certificates) • Aesthetics and architectural style • Vertical Transportation • Accessibility for special needs • Fire safety standard • Electricity source (renewable, grid, mixed, etc)

VI Discussions

The current state of BIM search in the industry is as primitive as the first days of image search, which was based on surrounding metadata such as file title and textual context around the image [60]. From a data model viewpoint, BIM data has a clear advantage over image data; BIM data is machine-readable, structured data while image data is unstructured data that can be understood by machines only after processing with advanced machine learning algorithms. Yet, the image search community has overcome this challenge to a great degree and today one can enjoy Content-Based Image Retrieval (CBIR) as seen in the likes of Google and Bing. BIM search has not received enough attention from researchers let alone from software developers. The reasons might be that 1) BIM is still not in widespread use and 2) BIM is a relatively niche market; the number of BIM models and users is just a fraction of the number of images and their users. With the increase in the adoption of BIM and the awareness regarding the benefits and applications of BIM search, these barriers will be reduced to some extent. Similar to the stock photography industry, which is becoming more effective with advancements in CBIR, it is possible to build a stock BIM ecosystem. An effective BIM retrieval system (especially one equipped with context-based BIM retrieval) would play a critical role in realizing such an enterprise.

This paper categorized approaches to BIM search in terms of context, geometry, and content, and in turn subdivided each into subcategories. These subcategories are not comprehensive in either breadth or in depth. Breadth-wise, plumbing, electrical, security, acoustics, etc are not taken into account. Depth-wise, it is possible to further subdivide some of these subcategories. For example, a 'roof system' can be subdivided into: material, structural system (truss, frame, arch, etc.), load bearing capacity, etc. Such details will be elaborated upon during the actual implementation of a BIM retrieval engine in a later stage of the research.

BIM retrieval based on geometry and content has recognizable parallels in other domains such as content-based image retrieval. However, we believe that the nature of architectural design (high volume of projects with each to achieve certain goals while satisfying a set of constraints) presents a unique opportunity for retrieval of BIM models based on a comparison with the context for which the building models were originally designed. Context-based retrieval could also be utilized together with clients in the project planning phase, in order to explore ideas and possibilities.

VII Conclusion

The aim of this research was to review the current state of BIM search and to investigate various approaches therein. We first highlighted the need for BIM search for design recycling and continuous improvement of architectural design practices, as well as its key role for a successful BIM-based knowledge management system and stock BIM ecosystem. The concept of 'indexing depth' was also introduced in four levels to clarify 'what' is being indexed by the information retrieval engine. These levels included metadata, data, extracted information, and incorporated domain information and knowledge. Finally, BIM search was categorized in terms of context, geometry, and content, and their different subcategories were elaborated. To have an intelligent and comprehensive BIM retrieval engine, all of these approaches need to be covered. Geometry-based retrieval (especially topological aspects) is researched more than the other two approaches. Research about context-based BIM retrieval is almost non-existent, perhaps because such an approach is not well-researched in other domains. Context-based retrieval is most relevant in the domain of Architecture because of its unique project-based and goal-oriented nature. Therefore, it is the duty of the researchers in AECO community to advance this approach.

REFERENCES

[1] G. Gao, Y.-S. Liu, M. Wang, M. Gu, and J.-H. Yong, "A query expansion method for retrieving online BIM resources based on Industry Foundation Classes," *Autom. Constr.*, vol. 56, pp. 14–25, 2015.

[2] J. Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Nicholas Brealey Publishing, 2011.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[4] G. Gao, Y.-S. Liu, P. Lin, M. Wang, M. Gu, and J.-H. Yong, "BIMTag: Concept-based automatic semantic annotation of online BIM product resources," *Adv. Eng. Informatics*, 2015.

[5] V. Ayzenshtadt, C. Langenhan, S. S. Bukhari, K.-D. Althoff, F. Petzold, and A. Dengel, "Thinking With Containers: A Multi-Agent Retrieval Approach for the Case-Based Semantic Search of Architectural Designs.," in *ICAART (1)*, 2016, pp. 149–156.

[6] A. Aksamija and I. Iordanova, "Computational Environments with Multimodal Representations of Architectural Design Knowledge," *Int. J. Archit. Comput.*, vol. 8, no. 4, pp. 439–460, 2010.

[7] C. Langenhan, M. Weber, M. Liwicki, F. Petzold, and A. Dengel, "Graph-based retrieval of building information models for supporting the early design stages," *Adv. Eng. Informatics*, vol. 27, no. 4, pp. 413–426, 2013.

[8] C. J. Anumba, C. Egbu, and P. Carrillo, *Knowledge management in construction*. John Wiley & Sons, 2008.

[9] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994.

[10] I. Watson and S. Perera, "Case-based design: A review and analysis of building design applications," *AI EDAM*, vol. 11, no. 1, pp. 59–87, 1997.

[11] D. Miller and J. Shamsie, "The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965," *Acad. Manag. J.*, vol. 39, no. 3, pp. 519–543, 1996.

[12] A. Edmunds and A. Morris, "The problem of information overload in business organisations: a review of the literature," *Int. J. Inf. Manage.*, vol. 20, no. 1, pp. 17–28, 2000.

[13] G. K. Kululanga and R. McCaffer,

"Measuring knowledge management for construction organizations," *Eng. Constr. Archit. Manag.*, vol. 8, no. 5/6, pp. 346–354, 2001.

[14] T. Groff and T. Jones, *Introduction to knowledge management*. Routledge, 2012.

[15] R. Amor, D. Bloomfield, M. Langham, J. Fortmann, N. Jarrett, and P. Goodwin, "The UK industry knowledge base feasibility study," *CIB Rep.*, pp. 35–44, 1996.

[16] G. De Paoli, I. Iordanova, and T. Tidafi, "Is A Digital Model Worth A Thousand Pictures?," in *Computer-Aided Architectural Design Futures (CAADFutures) 2007: Proceedings of the 12th International CAAD Futures Conference*, 2007, p. 85.

[17] S. E. O'Keeffe, "Synergy of the developed 6D BIM framework and conception of the nD BIM framework and nD BIM process ontology." The University of Southern Mississippi, 2013.

[18] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha, "Big Data in the construction industry: A review of present status, opportunities, and future trends," *Adv. Eng. Informatics*, vol. 30, no. 3, pp. 500–521, 2016.

[19] I. Motawa, S. Janarthanam, and A. Almarshad, "Live Capture of Energy-related Knowledge into BIM Systems," in *Construction Research Congress 2014: Construction in a Global Network*, 2014, pp. 249–258.

[20] A. W. Court, D. G. Ullman, and S. J. Culley, "A Comparison Between the Provision of Information to Engineering Designers in the UK and the USA," *Int. J. Inf. Manage.*, vol. 18, no. 6, pp. 409–425, 1998.

[21] G. J. Hahm, M. Y. Yi, J. H. Lee, and H. W. Suh, "A personalized query expansion approach for engineering document retrieval," *Adv. Eng. Informatics*, vol. 28, no. 4, pp. 344–359, 2014.

[22] Z. Li, V. Raskin, and K. Ramani, "Developing engineering ontology for information retrieval," *J. Comput. Inf. Sci. Eng.*, vol. 8, no. 1, p. 11003, 2008.

[23] R. Fidel and M. Green, "The many faces of accessibility: engineers' perception of information sources," *Inf. Process. Manag.*, vol. 40, no. 3, pp. 563–581, 2004.

[24] A. K. Chakrabarti, S. Feineman, and W. Fuentevilla, "Characteristics of sources, channels, and contents for scientific and technical information systems in industrial R and D," *IEEE Trans. Eng. Manag.*, no. 2, pp. 83–88, 1983.

[25] T. E. Pinelli, A. P. Bishop, R. O. Barclay, and J. M. Kennedy, "The information-seeking behavior of engineers," *Encycl. Libr. Inf. Sci.*, vol. 52, no. 15, pp. 167–201, 1993.

[26] G. J. Leckie, K. E. Pettigrew, and C. Sylvain, "Modeling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers," *Libr. Q.*, vol. 66, no. 2, pp. 161–193, 1996.

[27] D. Ellis and M. Haugan, "Modelling the information seeking patterns of engineers and research scientists in an industrial environment," *J. Doc.*, vol. 53, no. 4, pp. 384–403, 1997.

[28] D. E. O'Leary, "Enterprise knowledge management," *Computer (Long. Beach. Calif).*, vol. 31, no. 3, pp. 54–61, 1998.

[29] R. Sommer, *Design awareness*. Rinehart Press, 1971.

[30] R. Gifford, D. W. Hine, and J. A. Veitch, "Meta-analysis for environment-behavior and design research, illuminated with a study of lighting level effects on office task performance," in *Toward the integration of theory, methods, research, and utilization*, Springer, 1997, pp. 223–253.

[31] J. Alam and J. J. Ham, "Towards a BIM-based energy rating system," in *Proceedings of the 19th International Conference on Computer-Aided Architectural Design Research in Asia: Rethinking Comprehensive Design: Speculative Counterculture; 2014*, 2014, pp. 285–294.

[32] D. Davis, "Evaluating Buildings with Computation and Machine Learning," *ACADIA*, 2016.

[33] J.-R. Lin, Z.-Z. Hu, J.-P. Zhang, and F.-Q. Yu, "A Natural-Language-Based Approach to Intelligent Data Retrieval and Representation for Cloud BIM," *Comput. Civ. Infrastruct. Eng.*, vol. 31, no. 1, pp. 18–33, 2016.

[34] K.-Y. Lin and L. Soibelman, “Incorporating domain knowledge and information retrieval techniques to develop an architectural/engineering/construction online product search engine,” *J. Comput. Civ. Eng.*, vol. 23, no. 4, pp. 201–210, 2009.

[35] P. Demian and R. Fruchter, “Measuring relevance in support of design reuse from archives of building product models,” *J. Comput. Civ. Eng.*, vol. 19, no. 2, pp. 119–136, 2005.

[36] P. Demian and R. Fruchter, “An ethnographic study of design knowledge reuse in the architecture, engineering, and construction industry,” *Res. Eng. Des.*, vol. 16, no. 4, pp. 184–195, 2006.

[37] P. Demian and R. Fruchter, “Methodology for usability evaluation of corporate memory design reuse systems,” *J. Comput. Civ. Eng.*, vol. 20, no. 6, pp. 377–389, 2006.

[38] D. A. Schon, *The reflective practitioner: How professionals think in action*, vol. 5126. Basic books, 1984.

[39] R. Fruchter, “Conceptual, collaborative building design through shared graphics,” *IEEE Expert Intell. Syst. Their Appl.*, vol. 11, no. 3, pp. 33–41, 1996.

[40] K. Reiner and R. Fruchter, “Project memory capture in globally distributed facility design,” in *Computing in Civil and Building Engineering (2000)*, 2000, pp. 820–827.

[41] P. Demian and P. Balatsoukas, “Information retrieval from civil engineering repositories: importance of context and granularity,” *J. Comput. Civ. Eng.*, vol. 26, no. 6, pp. 727–740, 2012.

[42] J. Bayer, S. S. Bukhari, C. Langenhan, M. Liwicki, K.-D. Althoff, F. Petzold, and A. Dengel, “Migrating the classical pen-and-paper based conceptual sketching of architecture plans towards computer tools-prototype design and evaluation,” in *International Workshop on Graphics Recognition*, 2015, pp. 47–59.

[43] Merriam-Webster, “Definition of Metadata,” *Merriam-Webster*. [Online]. Available: https://www.merriam-webster.com/dictionary/metadata. [Accessed: 07-Jul-2017].

[44] S. Daum and A. Borrmann, “Processing of topological BIM queries using boundary representation based methods,” *Adv. Eng. Informatics*, vol. 28, no. 4, pp. 272–286, 2014.

[45] S. Daum and A. Borrmann, “Checking spatio-semantic consistency of building information models by means of a query language,” in *Proc. of the Intl Conference on Construction Applications of Virtual Reality*, 2013.

[46] K. Park and R. Krishnamurti, “DIGITAL DIARY OF A BUILDING: A System for Retrieval and Update of Information Over a Building Life Cycle,” *CAADRIA, New Delhi-India*, 2005.

[47] J. D. Goedert and P. Meadati, “Integrating construction process documentation into building information modeling,” *J. Constr. Eng. Manag.*, vol. 134, no. 7, pp. 509–516, 2008.

[48] P. Pauwels, R. De Meyer, and J. Van Campenhout, “Interoperability for the design and construction industry through semantic web technology,” in *International Conference on Semantic and Digital Media Technologies*, 2010, pp. 143–158.

[49] S. Jiang, H. Zhang, and J. Zhang, “Research on BIM-based Construction Domain Text Information Management.,” *JNW*, vol. 8, no. 6, pp. 1455–1464, 2013.

[50] J. McKechnie, S. Shaaban, and S. Lockley, “Computer assisted processing of large unstructured document sets: a case study in the construction industry,” in *Proceedings of the 2001 ACM Symposium on Document engineering*, 2001, pp. 11–17.

[51] C. H. Caldas, L. Soibelman, and J. Han, “Automated classification of construction project documents,” *J. Comput. Civ. Eng.*, vol. 16, no. 4, pp. 234–243, 2002.

[52] A. Aksamija and F. Grobler, “Architectural ontology: development of machine-readable representations for building design drivers,” in *Computing in Civil Engineering (2007)*, 2007, pp. 168–175.

[53] S. Daum, A. Borrmann, C. Langenhan, and F. Petzold, “Automated generation of building fingerprints using a spatio-semantic query language for building information models,” *eWork Ebus. Archit. Eng. Constr. ECPPM 2014*, p. 87, 2014.

[54] M. Strohmaier, M. Lux, M. Granitzer, P. Scheir, S. Liaskos, and E. Yu, “How do

users express goals on the web?-an exploration of intentional structures in web search," in *Web Information Systems Engineering–WISE 2007 Workshops*, 2007, pp. 67–78.

[55] A. Aksamija, *Knowledge-based model for integrated tall building design factors*. University of Illinois at Urbana-Champaign, 2008.

[56] P. Demian, K. Ruikar, T. Sahu, and A. Morris, "3DIR: exploiting topological relationships in three-dimensional information retrieval from BIM environments," 2016.

[57] C. Langenhan and F. Petzold, "The fingerprint of architecture-sketch-based design methods for researching building layouts through the semantic fingerprinting of floor plans," *Int. Electron. Sci. J. Archit. Mod. Inf. Technol.*, vol. 4, p. 13, 2010.

[58] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3D models," *ACM Trans. Graph.*, vol. 22, no. 1, pp. 83–105, 2003.

[59] S. Inanc, "Casebook. an information retrieval system for housing floor plans," 2000.

[60] R. Datta, D. Joshi, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Comput. Surv*, vol. 40, no. 5, 2008.