Linking ontology metrics with BIM modeling stages

Silke Dierssen[1], Mario Cichonczyk[1], Uwe Weitkemper[1] and Dominic Becking[1]
[1]Bielefeld University of Applied Sciences, Campus Minden
E-mail(s): sdierssen,mcichonczyk,uweitkemper,dbecking@fh-bielefeld.de

Abstract: Before starting a project, stakeholders of the Architecture, Engineering and Construction (AEC) industry aim to agree on common requirements, such as which information and geometry should be provided at any given stage of the engineering process. In particular, this consensus includes the geometric depth of the processed building information model and its enhancement with information artifacts. A suitable determination of a categorical standard can be achieved on the basis of Level of Geometry (LOG) and Level of Information (LOI). Since no international standard exists for these qualifiers, individual catalogs and checklists are currently used, which often do not allow an unambiguous classification. Given the subjective interpretability of existing methods, a more explicit and objective way of categorization would be beneficial.
The purpose of this paper is to investigate whether this process can be supported by a mathematically rigorous methodology. If LOG and LOI could be determined by a formal computational process, the influence of subjectivity would be reduced to introduce a more objective categorization. With the proposed method, stakeholders can achieve a more reliable assignment of criteria and a common understanding.
We examine whether the Resource Description Framework (RDF) representation of a building information model in Industry Formation Classes (IFC) format allows conclusions about categorization levels. Existing metrics for the analysis of ontologies are applied to RDF graphs. We then assess their correlation with LOG and LOI levels. Finally, we evaluate these mathematically objective qualifiers in terms of their relevance to the AEC domain. Our results show a promising framework for further exploration of a new class of BIM tools that facilitate collaboration and standardization.

Keywords: BIM, IFC, Ontology Metrics, Level of Development (LOD), Level of Information (LOI)

1 Introduction

For any construction project that uses a Building Information Model (BIM), project stakeholders must agree on the level of elaboration of the included model objects. This is usually done by specifying the elaboration via the classification into a Level of Development (LOD). The specified LOD usually

includes requirements for geometric detailing (LOG) and alphanumeric information content (LOI). The classification into the individual levels is mostly done for LOG via textual description in combination with a visual illustration of the individual elements. For the LOI, predefined sample checklists are common. For this purpose, numerous concepts, catalogs and publications exist to facilitate the application process and to achieve a standard. For example, a well-known and often used guideline is the *Level of Development Specification* from BIMForum [1]. These publications provide a suitable guide but still leave much room for interpretation. Since the given definitions are ambiguous, users have to make subjective decisions about what a particular LOD exemplifies and what graphical representation and information to include in the model. Abualdenien and Borrmann [2] highlight discrepancies, misunderstandings and misinterpretations that practitioners face when following the definitions of the various LOD specifications and the need to standardize the different concepts internationally. The concept of LOD itself has been established in the AEC industry, but lacks a formal method. An approach that provides a more objective and reproducible categorization would be beneficial.
Our hypothesis is that the established approach to represent IFC data as an RDF graph lends itself for further analysis beyond model enrichment and querying. Based on the fact that ontological representations have been studied thoroughly in the semantic web domain [3], we assume that the interpretation of a given BIM model only as an abstract ontological construct in a superordinate sense holds information centric value. Viewed from an ontology theoretic perspective, a BIM model can be regarded as a knowledgebase describing a graph realization of an abstract information schema. This analogy arises from the observation that IFC files constitute an object-based instance collection of the IFC entity relationship model. As such, generic semantic analyses produce valid results if applied to BIM applications. We demonstrate that inherent properties of the internal graph topology spanning an IFC instance can be linked to abstract concepts like LOD and LOI. The very same structure originates the visual renderings that would otherwise be interpreted subjectively in e.g. quality control. Therefore, we test whether existing ontology metrics can be applied for the same assessment.

2 Background

Opposed to conventional CAD drawings, digital building models do not have a measurement scale which has a significant influence on the degree of elaboration. Other regulations must be made regarding the specificity of the model for the individual project phases. This can be determined with the Level of Development, a combination of specifications for the geometric representation (LOG) and the alphanumeric information to be provided (LOI) which comprises five levels ranging from LOD100 and reaching up to LOD500. Approaches have been developed to automate the classification of model objects into LOD categories. Abualdenien and Borrmann [2][4] have shown that features which exploit the geometric surface complexity of primitives composing an assembly are useful for ensemble learning methods. For a quality check and model analysis, there are a number of software tools that can automatically validate the model according to predefined rules. These are suitable for verifying information contained in the model, but checks have to be defined manually.
It has become a common approach in the AEC domain to make use of ontologies, the Resource

Description Framework (RDF) and other semantic web technologies to represent knowledge and information in a machine-readable form for reuse and processing. Information is formulated as directed graphs (e.g. RDF graphs). By applying ontological linkages to this information, semantic relations and interconnections can be encoded. Content can be queried using a query language, e.g. SPARQL. The application of these technologies is often used as a complement to existing BIM software. This is mainly done to overcome the problem of interoperability of different software tools within individual disciplines, to establish connections between different application areas in order to integrate unused, valuable resources and to make use of the intrinsic logical features of semantic technologies [3]. For these purposes, a universal entity relationship model is described by the Industry Foundation Classes schema, itself representing a highly descriptive ontology.

Deeper formal analysis using tailored metrics have been accepted as necessary means for ontological quality control and evolution tracking during development and application [5]. Based on the propositions of Franco, Vivo, Quesada-Martínez, *et al.* [6], within the scope of this work, metrics for ontology engineering are understood as objective and reproducible measuring instruments which asses quantitative and qualitative criteria of a structured information corpus adhering to the definition described in [7]. As Gangemi, Catenacci, Ciaramita, *et al.* [8] point out, different dimensions for ontology measurement can be quantified. We loosely summarize potential dimensions as follows [9]:

- **Schema Metrics:** Aspects of ontology design are measured. The correct modeling of intended knowledge cannot be ascertained given an ontology, but its semantic potential to do so is measurable [10].
- **Graph Metrics:** The ontology is viewed as an information *object*. The metric space is defined as a function of its topology and logical syntax [8].
- **Knowledgebase Metrics:** Describe the way in which knowledge is placed within an ontology as a whole and its effectiveness to encapsulate information [10].
- **Class Metrics:** Relate defined classes provided by the schema with their utilization in the knowledgebase [10].

Chosen mathematical quantifiers from these dimensions are being tested for their explanatory capabilities in the context of semantic building representations.

3 Method

To test the explanatory capability of standard criteria from the semantic web domain for model quality assessment, we applied different ontology and graph metrics to a selected set of IFC files. These sample files have been derived from a pre-existing building information model by successively reducing first the geometric complexity and thereafter the information content to obtain 7 different models of the same building in total (LOG100-LOG400, LOI100-LOI300, LOI400 being equivalent to LOG400). The model and all related data artifacts will be released as a reference data set in a separate publication. Figure 1 shows the incremental LOG stages. Table 1 gives a short excerpt over the specified information artifacts for each LOI level. These levels have been declared for the entity set {*IfcWall, IfcDoor, IfcColumn, IfcRoof, IfcWindow, IfcSlab, IfcStair, IfcBeam, IfcCurtainWall, IfcBuildingElementProxy, IfcRailing*}.

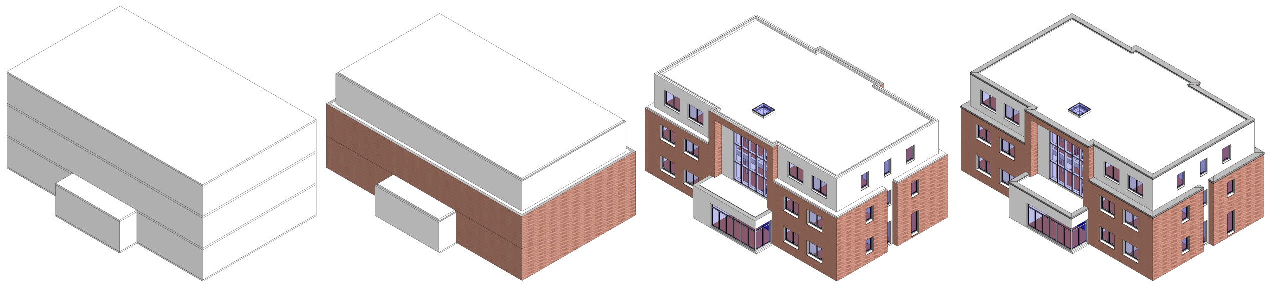

Figure 1: Sample IFC models.

Table 1: Example of LOI specification

Element type	LOI 100	LOI 200	LOI 300	LOI 400
...				
Walls	Name	+ is external	+ load-bearing	+ thermal transmittance
Window	Name	+ is external		+ thermal transmittance
...				

Our test sample is a 3-story office building in solid construction with a flat roof, built in the year 2018. All sample IFC files were exported from Autodesk Revit 2021 in IFC4 Design Transfer View and afterwards converted to RDF graphs in Turtle format using the IFCtoRDF[1] converter. This approach was chosen to transform the IFC representation into a more flexible, fully interoperable ontology that supports semantic analysis. The underlying EXPRESS schema has limited functionality beyond descriptive capabilities [11][12] and our analysis requires the custom implementation of metrics as well as graph query methods [13].
We specifically focused on quantifiers which give weight to topological and taxonomical dimensions to prevent indirect benchmarking of the IFC schema itself. As our interest lies in the metric behaviour of the test sample as a specific realization of the schema, we have selected a set of promising quantifiers from graph and knowledge base analytics to illustrate the process (see Table 2). These metrics were implemented as either graph operations or SPARQL queries and then applied to our 7 sample IFC files.

4 Results

After conversion to RDF, the different modeling stages of our IFC sample set exhibit topological graph properties as listed in Table 3.

[1] Pieter Pauwels: https://github.com/pipauwel/IFCtoRDF

Table 2: Tested metrics.

Metric	Description	Formula
Density	Measures how connected a graph is. Defined as the relationship between all theoretically possible connections in V and the observed edges E.	$D = \frac{2\|E\|}{\|V\|*(\|V\|-1)}$
Degree Centrality	The degree d(v) of a node v is equal to the sum of all edges incident upon that node. Degree centrality is the mean degree contained in V.	$DC = \frac{\sum_{v_i}^{V} d(v_i)}{\|V\|}$
Degree Centrality stdev.	Introduced to capture the second statistical moment of the dispersion properties of DC. Commonly calculated as the root of the variance.	$\sigma = \sqrt{\frac{\sum_{v_i}^{V} \|d(v_i)-DC\|^2}{\|V\|}}$
Average Population	Indicates how extensively populated a knowledgebase is. Calculated through the ratio between the number of instances I provided by the knowledgebase divided by the number of classes C defined in the ontology schema [10].	$AP = \frac{\|I\|}{\|C\|}$
Class Richness	Represents a percentage indicating how well a knowledgebase exemplifies the potential knowledge in the schema. Calculated by dividing the cardinality of the set of all classes C′ within the knowledgebase by the total number of classes C defined in the ontology [10].	$CR = \frac{\|C'\|}{\|C\|}$

Table 3: Modeling stages graph topology.

	Nodes	Edges	Pendants		Nodes	Edges	Pendants
LOG 100	2888	6525	732	**LOI 100**	275818	778128	20206
LOG 200	8857	23295	1457	**LOI 200**	298321	833729	27510
LOG 300	221009	627970	20686	**LOI 300**	298325	834046	27511
LOG 400	298412	834687	27537	**LOI 400**	298412	834687	27537

On average, graph density decreased by –61.6% per LOG stage. Degree centrality behaves accordingly but exerts an average decrease in standard deviation by –40.8%. Average population increased 9.8-fold with the most significant change between LOG200 and LOG300, demonstrating a median increment of 254.4%. Class richness expanded by 20.7%, mostly saturating at the LOG300-LOG400 level.

For LOI, density decreased by –2.9% on average. The main contribution can be observed at the transition from LOI100 to LOI200 at –8.4% and stabilizing thereafter. Again, degree centrality mirrors this behavior. Standard deviation is reduced by –3.4% presenting a median rate of change with –0.04%. Average population saturates after LOI200 and gains 2.0% on average. Likewise, class richness does not increase significantly after LOI200 with a pronounced difference of 6.6% before saturation.

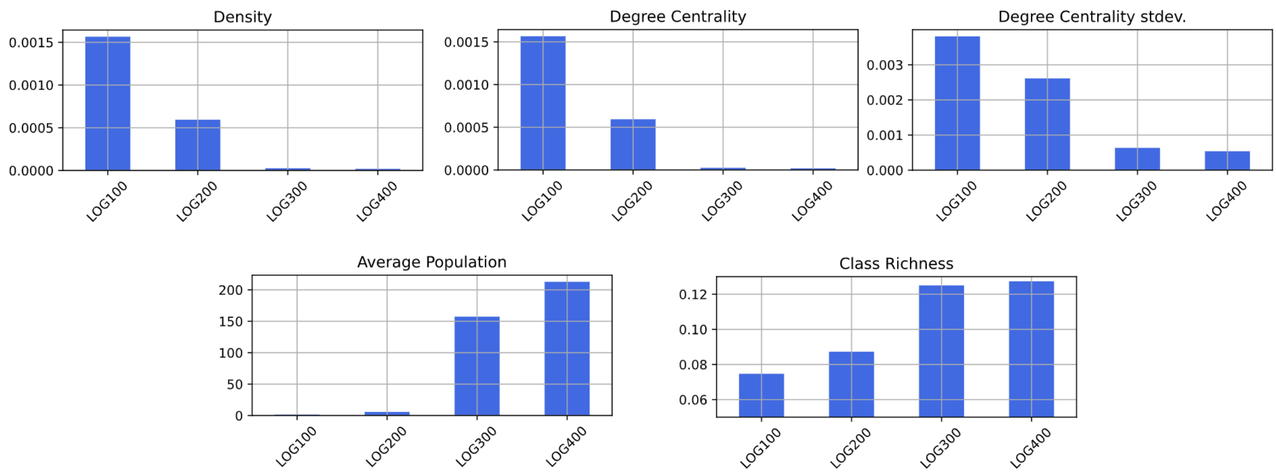

Figure 2: Metrics calculation for LOG stages.

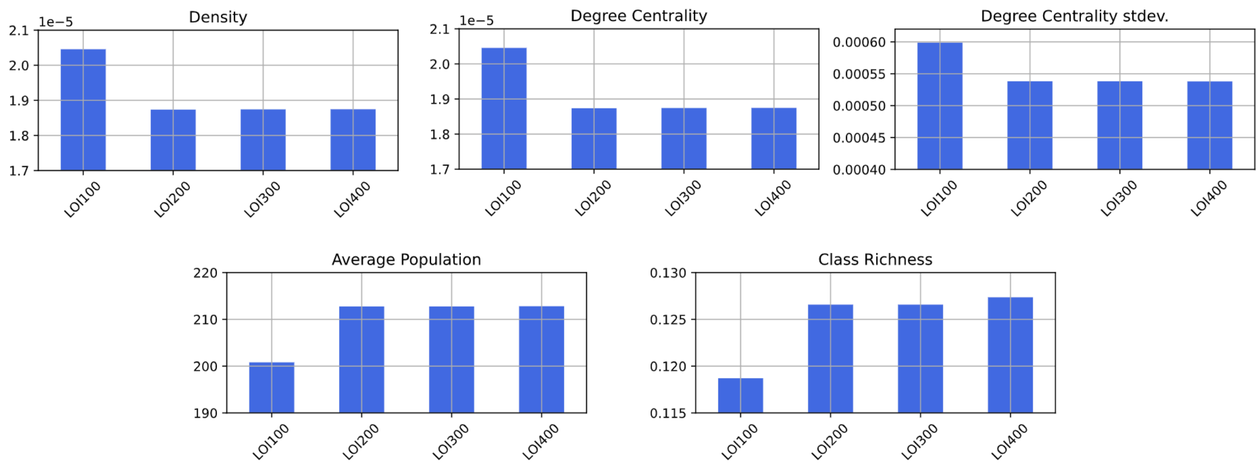

Figure 3: Metrics calculation for LOI stages.

5 Discussion and Future Work

We conclude that generically defined ontology metrics can correlate with BIM modeling stages. The quantifiers we have chosen respond more noticeably to Level of Geometry than Level of Information. We assume this to be an effect of LOI being a more semantically complex categorization, which is predominantly defined by specific IFC classes and their meaning, e.g. *IFCPropertySet*. Once these instances exist within the knowledgebase, the given metrics tend to react less to them as the general ontological topology remains largely unchanged when adding more information afterwards.

We emphasize that ontology metrics are at the center of an active field of research. Existing quantifiers are improved upon and new ones with more specialized purposes are constantly being added. Therefore, many more metrics can be investigated using the presented method.

Our test case has shown the potential value which an ontology theoretic view of an IFC file can add, but measured properties of the underlying RDF graph are bound to the specific model. Future work should aim to find generalizing patterns by testing a wide variety of IFC samples beyond the presented

case. If a large enough sample base is evaluated using our method, converging components of different metrics can be identified. Such patterns would normalize the metric space across all domains and could be universally applied in BIM quality control and replace subjective interpretation providing a vital step towards standardization between BIM stakeholders.
Currently, a BIM quality check can only verify whether previously defined rules have been adhered to. If the requirements for the model change, new agreements and constraints must be defined. Without a requirements catalog, an automated model check is therefore difficult using available tools. However, an initial automated assessment of Level of Development in the event of renovation or redevelopment could be helpful and a significant economic advantage.
We are currently investigating which properties a metric needs for such a purpose and test different custom metric designs against the needs of BIM. We observe that transforming IFC models into RDF representations is a beneficial step in this line of research because it enables querying and semantic enrichment for an increase in metric precision.
Lastly, if the behavior of ontology metrics can be understood in such depth that their measurement can be linked to actual meaning in the context of the AEC industry beyond LOG and LOI, machine-aided control and optimization of complex goal-functions in the planning and design space become viable. The same advantage applies to Scan2BIM for validation of automatically reconstructed semantic models.

Acknowledgements

The project on which this report is based was funded by the German Federal Ministry of Education and Research under the funding code 13FH554IX6 and actively and generously supported by our partners HOCHTIEF ViCon GmbH and Pape Architekten AG. The responsibility for the content of this publication lies with the authors.

References

[1] BIMForum, *Level of development specification*, https://bimforum.org/lod/, Stand: 2022-05-18, 2021.

[2] J. Abualdenien and A. Borrmann, "Levels of detail, development, definition, and information need: A critical literature review", *Journal of Information Technology in Construction*, vol. 27, pp. 363–392, Apr. 2022. DOI: 10.36680/j.itcon.2022.018.

[3] P. Pauwels, S. Zhang, and Y.-C. Lee, "Semantic web technologies in aec industry: A literature overview", *Automation in Construction*, vol. 73, pp. 145–165, 2017, ISSN: 0926-5805.

[4] J. Abualdenien and A. Borrmann, "Formal analysis and validation of levels of geometry (log) in building information models", in *EG-ICE 2020 Workshop on Intelligent Computing in Engineering*, Universitätsverlag der TU Berlin, 2020, p. 33.

[5] D. Vrandečić and Y. Sure, "How to design better ontology metrics", in *The Semantic Web: Research and Applications*, E. Franconi, M. Kifer, and W. May, editors, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 311–325, ISBN: 978-3-540-72667-8.

[6] M. Franco, J. M. Vivo, M. Quesada-Martínez, A. Duque-Ramos, and J. T. Fernández-Breis, "Evaluation of ontology structural metrics based on public repository data", *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 473–485, Feb. 2019, ISSN: 1477-4054.

[7] R. Hoehndorf, P. N. Schofield, and G. V. Gkoutos, "The role of ontologies in biological and biomedical research: A functional perspective", *Briefings in bioinformatics*, vol. 16, no. 6, pp. 1069–1080, 2015.

[8] A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann, "A theoretical framework for ontology evaluation and validation.", in *SWAP*, vol. 166, 2005, p. 16.

[9] B. Lantow, "Ontometrics: Application of on-line ontology metric calculation.", in *BIR Workshops*, vol. 1684, 2016, pp. 1–12.

[10] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth, and B. Aleman-Meza, "Ontoqa: Metric-based ontology quality analysis", 2005.

[11] E. van den Bersselaar, J. Heinen, M. Chaudron, and P. Pauwels, "Automatic validation of technical requirements for a bim model using semantic web technologies", English, 1st 4TU/14USA research day on Digitalization in the Built Environment ; Conference date: 01-04-2022, 2022.

[12] J. Beetz, J. van Leeuwen, and d. Vries, "Ifcowl: A case of transforming express schemas into ontologies", *AI EDAM*, vol. 23, pp. 89–101, Feb. 2009. DOI: 10.1017/S0890060409000122.

[13] E. Tauscher, H.-J. Bargstädt, and K. Smarsly, "Generic bim queries based on the ifc object model using graph theory", in *Proceedings of the 16th International Conference on Computing in Civil and Building Engineering, Osaka, Japan*, 2016, pp. 6–8.