Multilingual semantic enrichment of room-specific load profiles using BIM models for whole building energy simulation

Kasimir Forth[1]
[1]Lehrstuhl für Computergestützte Modellierung und Simulation, Technische Universität München, Arcisstr. 21, 80333 München, Deutschland
E-mail(s): kasimir.forth@tum.de

Abstract: Whole building energy simulations are an established method for predicting the operational energy performance of new constructions. Nevertheless, these tools are mostly expert systems and setting up a Building Energy Model (BEM) still requires a lot of knowledge and effort. Often, current approaches, which transform Building Information Models (BIM) models for energy simulation in early design stages, don't take model-specific information related to energy properties into account or need to be enriched manually in the simulation tools. Sometimes, only generic construction sets or program types are assigned to the model, so the simulation results don't correspond with the specific information of the architectural BIM model. In this paper, we are proposing a novel methodology which creates BEM based on BIM by automatically enriching HVAC (Heating, Ventilation, Air Conditioning) program types for each room using Natural Language Processing (NLP). Based on the architectural room names as input, we propose to use multilingual large language models (LLM) to identify the semantically most similar program type to automatically enrich the model with the relevant information. Finally, we are testing the proposed approach using an office building as a case study.

Keywords: BIM, IFC, NLP, Energy Simulation

Erschienen in Tagungsband 34. Forum Bauinformatik 2023, Bochum, Deutschland, *DOI:* 10.13154/294-10093

1 Introduction

According to latest the United Nations Environment Programme, buildings and the construction industry are responsible for 40% of the global CO2 emissions [1]. 28% of these greenhouse gas (GHG) emissions are related to direct (e.g. fossil energy) and indirect (e.g. electricity) effects of the operation of residential and non-residential building stock, while 11-12% are mainly embodied emissions of construction materials and others [1]. To lower the environmental impacts of building operations, energy simulations during the entire design process help to quantify and optimize the building energy performance. Gao et al. identified the early design stages for having the most significant impact for decision making and optimizing of the energy demand [2]. To automate the simulation workflow, Building Information Models (BIM) support to reduce the modeling effort for the required Building Energy Models (BEM). Nevertheless, there are still a few manual steps required in the current BIM to BEM workflows, which increases the effort of the whole optimization process in early design stages.

2 State of the Art

2.1 Current challenges of BIM-based BEM

Although the research field of BIM to BEM has been investigated for more than 15 years, there are still unsolved issues. Gao et al. stated an automated workflow for transforming detailed space types of all rooms as future research, as this assignment is still done manually today [2]. Elnabawi also highlighted in his review the manual assignment of occupancy operating schedules as one of the main unsolved interoperability issues [3]. Raggi et al. came to a similar conclusion that "additional data (e.g. regarding some HVAC data [...], must be manually added [...] to the models before the energy simulation can run" [4]. The occupancy and operating schedules are mentioned as data related to Heating, Ventilation and Air-Conditioning (HVAC). Therefore, in this publication we will focus on the automated enrichment of these room-specific load profiles.

2.2 Interoperability in BIM to BEM workflows

Eckstädt et al. compared three different workflows of calculating whole building energy simulation, which is often also known as building performance simulation, based on an IFC model as an input file [5]. Nevertheless, the existing tools seem to still have a couple of problems due to correct IFC export settings for each simulation tools, limitations of the IFC import and the simulation tools themselves. Ramaji et al. introduced another IFC-based BIM to BEM transformation approach [6]. They directly transform IFC models to OpenStudio's native IDD format, which is still challenging and was tried to be handled using MVDs. Spielhaupter compared in his master thesis different IFC-based approaches for BIM to BEM transformation [7]). Yang et al. also used IFC files as input in their approach, but did not directly export to IDF (EnergyPlus native file), but first to gbXML format [8]. Nevertheless, their workflow overview shows, that there are often still fixups needed. Therefore, for my approach, I want to use the HBJSON format as transformation schema, as it is open source, easily transformable to other file formats (gbXML, IDF, etc.) and more robust in its geometric export.

2.3 Semantic enrichment of BIM

There are different State-of-the-Art methods for automated semantic enrichment for different purposes. Tanya Block showed in her review different approaches, methods and application field for semantic enrichment of BIM [9]. She identified two main approaches, such as IFC representation of building information including inference based enrichment and integrating IFC with external data source, as well as Semantic Web technologies for building information. One of the three main application fields includes building design and performance evaluation, focusing also on energy simulations [10]. Costa and Sicilia use semantic query languages for facilitating automatic transformation of BIM data, focusing on interoperable data models for building-scale energy simulation using EnergyPlus [11]. Baumgärtel et al. used in their approach ontologies to automatically vary and evaluate thermal energy building performances [12]. In general, most semantic approaches use ontologies, semantic web and linked-data-related technologies for automatic semantic enrichment of BIM. A few use NLP, but not for enriching detailed information from BIM models for energy simulations yet. In this paper, we will therefore use NLP-based automated semantic enrichment for energy models.

2.4 Natural Language Processing for the built environment

Natural Language Processing (NLP) showed significant improvement in research considering performance and usability, which also leads to a more relevant in other industries, such as the construction sector. As main challenges in this field, Wu et al. list three main potentials, such as excessive manual work, poor relation extraction and poor mapping between functions and NLP techniques [13]. Zheng et al. introduced domain-specific language models for the AEC domain, mainly focusing on text classification and named entity recognition [14]. In this publication, current State-of-the-Art large language models (LLM) are used to reduce the manual matching effort for exporting semantic rich building energy models for whole building energy simulation.

3 Methodology

3.1 General workflow

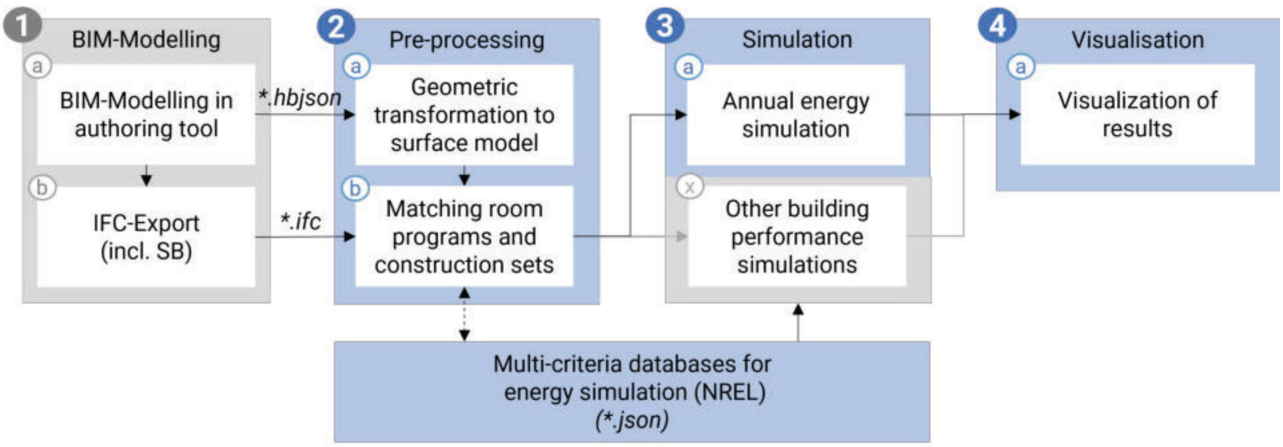

Figure 1: Overall methodology for semantic enrichment using BIM models for building performance simulation

The proposed methodology is divided in four main steps including the BIM modeling part, a pre-processing step, the energy simulation part and finally the visualization of the simulation results part, as shown in Figure 1. After the modeling in the authoring software (1.a), the model is exported as an IFC file (Industry Foundation Classes) including first level of space boundaries(1.b). Afterwards, the BIM model is used to transform the model to a room-specific surface model as a HBJSON file (2.a), needed as an input for most building performance simulation software.

Nevertheless, the main contribution to the existing state of the art is step 2.b, which contains the automated semantic enrichment using architectural BIM models in early design stages. We are focusing on matching program types to architectural rooms, while for the construction sets generic data are enriched. When all relevant semantic data are enriched, the annual energy demand is simulated based on the whole building using EnergyPlus (3.a). Finally, the simulation results are visualized and exported by the existing export settings of OpenStudio (4.a).

3.2 Implementation

As the focus of this paper is the semantic enrichment process, the geometric transformation to surface models is conducted with conventional tools using Pollination's Revit plugin [15]. For the whole building energy simulation, the open source simulation engine of the EnergyPlus and Ladybug Tools interfaces are used. Therefore, the Building Energy Model is transformed to Ladybug Tools' internal data format HBJSON, as this format can also be used for further criteria such as thermal comfort and daylight simulation (potential step 3.x from Section 3.1). For enriching the room-specific End-Use load profiles we use the standardized database by the National Renewable Energy Laboratory (NREL) of the U.S. Department of Energy [16]. It contains 23 different building types and 224 different program types and its ontology is shown in Figure 2, including internal loads of lighting, infiltration, ventilation, electric equipment, and different schedules.

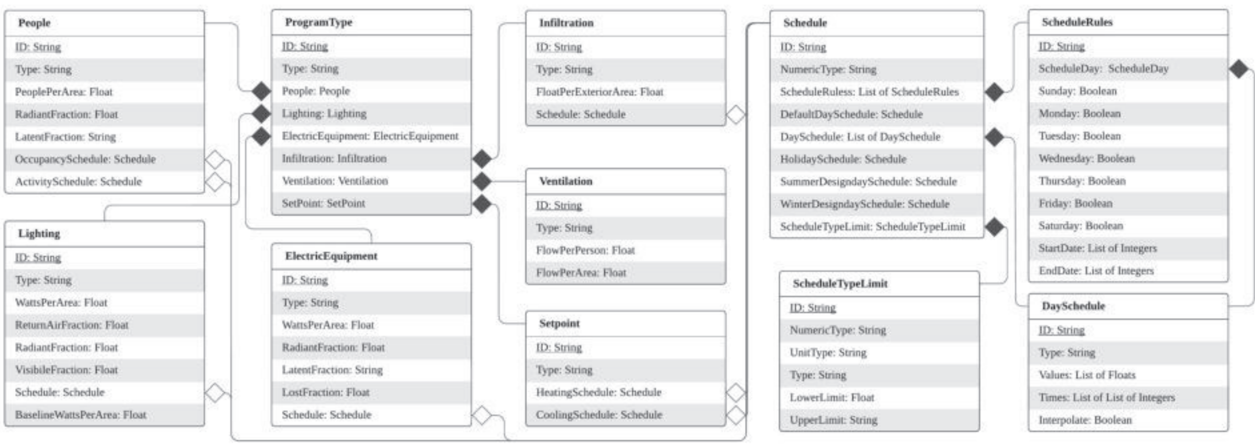

Figure 2: Ontology of End-Use Load Profiles by the NREL [16]

As the case study in Section 4.1 includes German language and the NREL database in English language, the NLP-based matching needs to use multi-lingual Large Language Models (LLM). Different multi-lingual language models are selected for testing the matching performance using the program types of case study, which are the following: "text-embedding-ada-002" (best performing embedding

LLM by openAI) [17], “xlm-roberta-large” (based on Google’s BERT) [18], “distiluse-base-multilingual-cased-v1” [19], “distiluse-base-multilingual-cased-v2” [19], and “LaBSE” [20]. After the matching of the semantically most similar program types to the architectural rooms, the HBJSON format is enriched with the information of the NREL program type database. Semantic similarity is derived by the cosine similarity, as previously described in more detail in [21]. The enriched BEM model is uploaded to the Pollination simulation server adding the missing information of weather data as *.epw and *.ddy files and run by using the "annual energy demand" recipe.

4 Case study and results

4.1 Case study

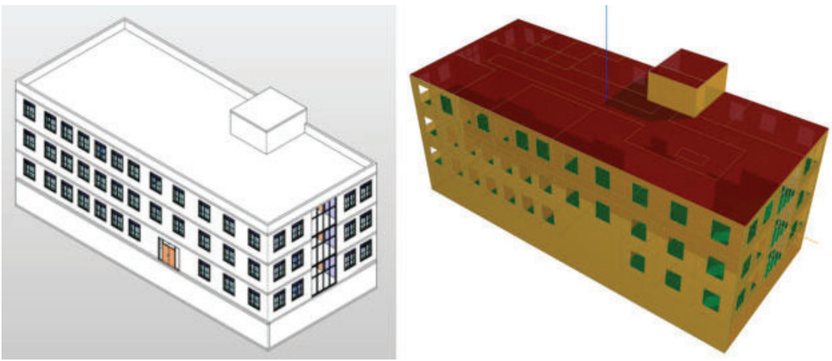

Figure 3: Case study building as Revit model (left) and energy model (right, using gbXML viewer)

For testing the proposed methodology, different LLM are evaluated using a real-world case study and compared to a manual matched ground truth. The chosen case study is an office building in Munich, consisting of in total five storeys and 80 different architectural rooms and 22 different room descriptions or usages, as shown in Figure 3.

4.2 Results of room-specific matching of load profiles using different LLM

In a first test, the NLP-based program type matching was conducted only using the inital LLM and compared to the manually matched ground truth, as shown on the left side in Figure 4. Nevertheless, the result of most LLM show only a success rate between 60%-70%, while the LLM RoBERTa even performs for only 10% of the rooms correct matches, which are rather insufficient results for a simple case study.

To improve the matching results, domain knowledge needs to be added to the LLM, e.g. by fine-tuning abbreviation which are often used for architectural rooms, for example "WC" to "toilet" or "ELT" as "Elektrotechnik". The matching results of the domain knowledge extension show significant better results accross all LLM, with LaBSE, distiluse-v2 and ada-002 reaching even 95%. To identify the best performing LLM also considering the final results, we simulate the annual energy demand in the next step.

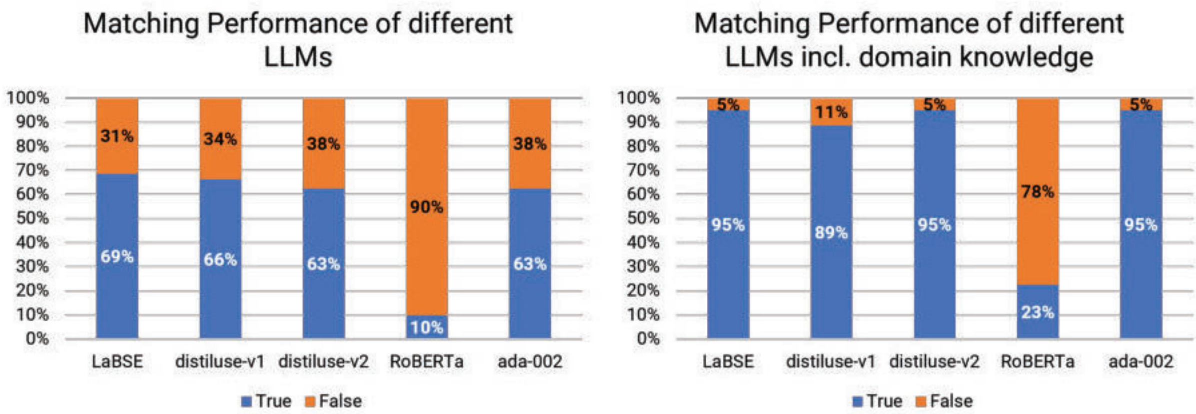

Figure 4: Results of the initial (left) and refined including domain knowledge (right) NLP-based matching performance using different LLMs

4.3 Energy simulation results of matched load profiles

In order to identify the performance effects of wrong matching, the matched program types considering domain knowledge including typical abbreviations are used to enrich the HBJSON data format using the information of the NREL database, as described in Section 3.2. In Figure 5, the results of the whole building simulation with Ladybug Tools and EnergyPlus based on the different improved LLM are shown.

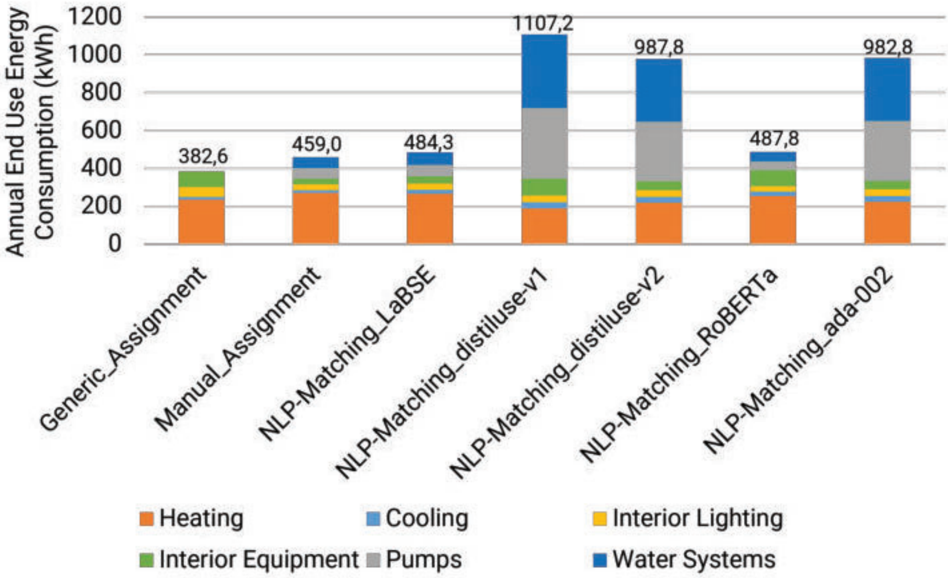

Figure 5: Simulation results of annual energy demand based on the refined matching algorithm using different LLM

Although the matching performance of "distiluse-v2" and "ada-002" showed promising results in the matching performance, the difference to the ground truth considering the annual energy results are significantly higher. Although "RoBERTa" had by far the worst matching results, the difference to the ground truth of the annual energy demand seems very small. Nevertheless, "LaBSE" is showing most sufficient results in both, matching performance and annual energy demand, which will be also considered in future work. The deviation in the matching performance is only 5% and the annual energy demand 5.5% compared to results of manual assignment, and thus the most suitable LLM.

5 Outlook and next steps

The proposed methodology aims to automatically enrich room-specific load-profiles for whole building energy simulation using BIM models. For the semantic enrichment, Natural language Processing (NLP) is used and test with different multilingual Large Language Models (LLM). The results of the the room-specific matching of load profiles using different LLM and the resulting energy demand simulation showed that the LLM "LaBSE" leads to sufficient accuracy for the given case study.

As a next step, more case studies need to be applied and evaluated to further prove the initial findings. Furthermore, the LLMs need to be more domain-specifically fine-tuned. Additionally, the method shall be also extended to automatically identify the semantically most similar construction sets based on the element and material information in the BIM models according to [21]. Finally, the scope of whole building energy simulation will be extended also to thermal comfort and daylight simulation to perform a multi-criteria analysis, as shown in step 3.x of Section 3.1.

References

[1] United Nations Environment Programme, *2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector*, Nairobi, 2022.

[2] H. Gao, C. Koch, and Y. Wu, “Building information modelling based building energy modelling: A review”, *Applied Energy*, vol. 238, pp. 320–343, 2019, ISSN: 03062619. DOI: 10.1016/j.apenergy.2019.01.032.

[3] M. H. Elnabawi, “Building Information Modeling-Based Building Energy Modeling: Investigation of Interoperability and Simulation Results”, *Frontiers in Built Environment*, vol. 6, 2020. DOI: 10.3389/fbuil.2020.573971.

[4] E. Raggi, F. Tavaroli, A. Zinno, M. Regidor Sànchez, and A. Gonzales Viegas, *Analysis of BIM-BEM critical parameters and recommendations to solve the current bottlenecks*, 2021.

[5] E. Eckstädt, C. Liersch, R. Hoch, and A. Hentschel, “Comparing BIM2SIM Workflows for coupled building and HVAC simulation”, *Proceedings of BauSIM 2022, Weimar*, 2022.

[6] I. J. Ramaji, J. I. Messner, and E. Mostavi, “IFC-Based BIM-to-BEM Model Transformation”, *Journal of Computing in Civil Engineering*, vol. 34, no. 3, 2020, ISSN: 0887-3801. DOI: 10.1061/(ASCE)CP.1943-5487.0000880.

[7] O. Spielhaupter, *BIM to BEM Transformation Workflows: A Case Study Comparing Different IFC-Based Approaches*, Wien, 2021.

[8] Y. Yang, Y. Pan, F. Zeng, Z. Lin, and C. Li, “A gbXML Reconstruction Workflow and Tool Development to Improve the Geometric Interoperability between BIM and BEM”, *Buildings*, vol. 12, no. 2, p. 221, 2022. DOI: 10.3390/buildings12020221.

[9] T. Bloch, “Connecting research on semantic enrichment of BIM - review of approaches, methods and possible applications”, *Journal of Information Technology in Construction*, vol. 27, pp. 416–440, 2022. DOI: 10.36680/j.itcon.2022.020.

[10] R. J. Scherer and S.-E. Schapke, “A distributed multi-model-based Management Information System for simulation and decision-making on construction projects”, *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 582–599, 2011, ISSN: 14740346. DOI: 10.1016/j.aei.2011.08.007.

[11] G. Costa and A. Sicilia, “Alternatives for facilitating automatic transformation of BIM data using semantic query languages”, *Automation in Construction*, vol. 120, p. 103 384, 2020, ISSN: 09265805. DOI: 10.1016/j.autcon.2020.103384.

[12] K. Baumgärtel and R. Scherer, “Automatic ontology-based Green Building Design Parameter Variation and Evaluation in Thermal Energy Building Performance Analyses”, *Proc. eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2016*, 2016.

[13] C. Wu, X. Li, Y. Guo, J. Wang, Z. Ren, M. Wang, and Z. Yang, “Natural language processing for smart construction: Current status and future directions”, *Automation in Construction*, vol. 134, p. 104 059, 2022, ISSN: 09265805. DOI: 10.1016/j.autcon.2021.104059.

[14] Z. Zheng, X.-Z. Lu, K.-Y. Chen, Y.-C. Zhou, and J.-R. Lin, “Pretrained domain-specific language model for natural language processing tasks in the AEC domain”, *Computers in Industry*, vol. 142, p. 103 733, 2022, ISSN: 01663615. DOI: 10.1016/j.compind.2022.103733.

[15] Pollination, *Pollination Revit Plugin*, 2023. [Online]. Available: https://www.pollination.cloud/revit-plugin (visited on 05/06/2023).

[16] E. Wilson, A. Parker, A. Fontanini, E. Present, J. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, A. LeBar, L. Liu, M. Praprost, L. Zhang, P. DeWitt, N. Merket, A. Speake, T. Hong, H. Li, N. Mims Frick, Z. Wang, A. Blair, H. Horsey, D. Roberts, K. Trenbath, O. Adekanye, E. Bonnema, R. El Kontar, J. Gonzalez, S. Horowitz, D. Jones, R. Muehleisen, S. Platthotam, M. Reynolds, J. Robertson, K. Sayers, and Q. Li, *End-Use Load Profiles for the U.S. Building Stock*, 2021. DOI: 10.25984/1876417.

[17] R. Greene, T. Sanders, L. Weng, and A. Neelakantan, *New and improved embedding model*, 2.03.2023. [Online]. Available: https://openai.com/blog/new-and-improved-embedding-model (visited on 03/02/2023).

[18] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov, “Unsupervised Cross-lingual Representation Learning at Scale”, *CoRR*, vol. abs/1911.02116, 2019.

[19] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Nov. 2019. [Online]. Available: http://arxiv.org/abs/1908.10084.

[20] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, *Language-agnostic BERT Sentence Embedding*, 2020. DOI: 10.48550/arXiv.2007.01852.

[21] K. Forth, J. Abualdenien, and A. Borrmann, “Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing”, *Energy and Buildings*, vol. 284, p. 112 837, 2023, ISSN: 03787788. DOI: 10.1016/j.enbuild.2023.112837.