NLP-based Semantic model healing for calculating embodied carbon in early building design stages

K. Forth & J. Abualdenien & A. Borrmann
Technical University of Munich, Germany

ABSTRACT: To reach the goals of the Paris agreement of limiting the global warming, the environmental impacts of new buildings need to be quantified and optimized already in the early design stages. Design decisions made during the early stages have a major impact on the success of projects in achieving their performance goals. Semantically rich models, such as Building Information Modeling (BIM), facilitate deriving consistent and automated quantity take-offs of the relevant components for calculating whole building life cycle assessments (LCA). Nevertheless, the early design stages are characterized by high uncertainty due to the lack of information and knowledge, impeding a holistic and consistent LCA for supporting design decisions and optimizing performance. A particular challenge is that early-stage BIM models typically lack stringency in terms of component modeling and material classification. Hence, this paper presents a methodology for enriching knowledge and characteristics from the coarse information available at the early design stages, in a process denoted as semantic model healing. In more detail, the proposed method employs different Natural Language Processing (NLP) strategies to increase the performance of automatically mapping materials defined in a BIM model to a knowledge database with environmental indicators of commonly used components. The knowledge database contains all missing information for LCA and has different granularities for a range of several potential design options of components, elements, and materials, including their dependencies. Accordingly, this paper investigates multiple NLP techniques and evaluates the performance of state-of-the-art deep learning models such as GermaNet, spaCy, or BERT. Finally, the most performant NLP approach is used to provide an automatic workflow for mapping IFC elements to the knowledge database, facilitating a seamless LCA in the early stages of design.

1 INTRODUCTION & MOTIVATION

In order to reach the international goals of the Paris Agreement and improve the ecological impacts of new buildings, life cycle assessments (LCA) are an established method to calculate several environmental indicators along the whole life cycle of buildings. According to the United Nations, manufacturing of materials for building construction cause 11% of the global energy-related carbon emissions (Abergel et al. 2017). Accordingly, a careful LCA of the different design options is required in order to identify the main drivers and optimize the building design accordingly.

Up to now, LCA has been performed mainly manually, which is time-consuming, especially quantifying the building components and mapping them to LCA databases (Llatas et al. 2020). Building Information Models (BIM) combine geometry and semantics and thus facilitate deriving consistent and automated quantity take-offs of the relevant components for calculating whole building life cycle assessments (LCA). Additionally, using and enriching the semantic information of e.g. materials provides a great potential to completely automate the calculation process (Safari & AzariJafari 2021).

However, in early design stages, essential decisions are taken that have a significant impact on the carbon footprint of the final building design. At the same time, the early design stages are characterized by high uncertainty due to the lack of information and knowledge, making a holistic and consistent LCA for supporting design decisions and optimizing performance challenging (Schneider-Marin et al. 2020).

In more detail, in the "rough" BIM models of early design stages, materials are rather defined by material groups than by specific types (e.g. "concrete" rather than "concrete C20/25"), which leads to a range of possibilities for each material group. Furthermore, several materials or component layers might not intentionally be defined yet, which opens the opportunity to explore and compare different design options.

This paper aims to answer the following research question to facilitate a reliable LCA in the early stages: Is automated semantic healing of "rough" BIM models possible that allows assigning correct element types and materials to the respective model elements?

2 STATE OF THE ART

2.1 *LCA in early design stages using 3D models*

For a whole building LCA in early design stages, there are different established approaches. One approach is using benchmarks, which are derived from already finished projects and are then transferred to early design stages. Gantner et al. introduced this approach, based on several design stages, where different input information are needed (Braune et al. 2018). The early design stages are subdivided into occasion and initialization phase, where building types and general systems are decided, and design and approval planning, where element systems are decided. Nevertheless, with this approach, optimization of LCA using different design options is difficult because they don't include all later details but rather benchmarks.

On the other hand, Hollberg suggested a parametric approach, based on LCA profiles for several construction types and using the Visual Programming Language Grasshopper with Rhino (Hollberg 2016), where optimizations can be automatically performed. However, the modell intup of the calculation depends only on geometric and doesn't include semantic information, same as with benchmarks.

2.2 *BIM-LCA integration*

Wastiels & Decuypere classified five different integration workflows for calculating LCA using BIM models (Wastiels & Decuypere 2019). While the first three still require manual work, mainly for mapping IFC element information to LCA profiles, the fourth workflow is based on plugins of LCA software in BIM authoring tools. The fifth workflow by Wastiels & Decuypere includes a BIM object enrichment with LCA profiles.

Rezaei et al. developed a method based on Revit models to calculate LCA in early and detailed design stages (Rezaei et al. 2019). LCA profiles on element levels are detailed into layers and material options, but the mapping to match Revit and LCA database assemblies is carried out manually. Nevertheless, the LCA results are given in ranges, due to uncertainties in early design stages, and not as total result.

Eleftheriadis et al. proposed an BIM-embedded LCA approach focusing on structural design alternatives in early design stages (Eleftheriadis et al. 2018). However, they do not consider all life-cycle modules (only A1-A3) and is also based on Autodesk Revit.

Horn et al. proposed an integration approach based on open BIM using Industry Foundation Classes (IFC) as data format (Horn et al. 2020). With the help of Information Delivery Manuals (IDM) and Model View Definitions (MVD), LCA for several level of development of building design are realized, also for early design stages.

2.3 *BIM data extraction & mapping methods*

Extracting data from IFC models and map those to different data structures or ontologies of the chosen use case is a complex task. Several approaches for different use cases have been developed recently.

Koo et al. explored the use of 3D geometric deep neural networks to distinguish BIM element subtypes and enrich semantics of IFC model entities (Koo et al. 2021). Costa & Sicilia propose a methodology of transforming and mapping building data from BIM models using Semantic Web technologies for an automated and flexible exchange with other software, e.g. whole building energy simulation (Costa & Sicilia 2020).

Wu et al. proposed a natural-language-based retrieval engine for BIM object database (Wu et al. 2019). Their use case is to mapping building components in BIM object databases with a higher accuracy than with keyword-based methods.

Reitschmid proposed a mapping algorithm of IFC materials to the LCA database Ökobaudat based on tokenization of material names and a distinct mapping or via Levenshtein distance (Reitschmidt 2015). Nevertheless, no Natural Language Processing (NLP) model was used and also no integration to element-specific mapping was proposed.

Locatelli et al. investigated in their scientometric analysis the synergies between NLP and BIM (Locatelli et al. 2021). Beside the field of Automatic Compliance Checking, they identified also Information Retrieval from BIM models and Information Enrichment of BIM objects as a further field of relevant application.

Nevertheless, an automated mapping of LCA and IFC data on element level has not been developed yet (Safari & AzariJafari 2021).

3 SEMANTIC MODEL HEALING

The semantic model healing process is part of a bigger framework, which we previously proposed (Forth et al. 2021). In the paper at hand, the focus is on how NLP techniques help to heal the BIM model semantically for the use case of LCA in early design stages. Typically, design decisions are finally decided by the client and not the architect, hence, the proposed methodology is leveraging open BIM data models.

The proposed healing process is based on NLP, using different strategies to increase the performance of mapping materials from a rough BIM model to a knowledge database with environmental indicators of commonly used components. The knowledge database contains all missing information for LCA and

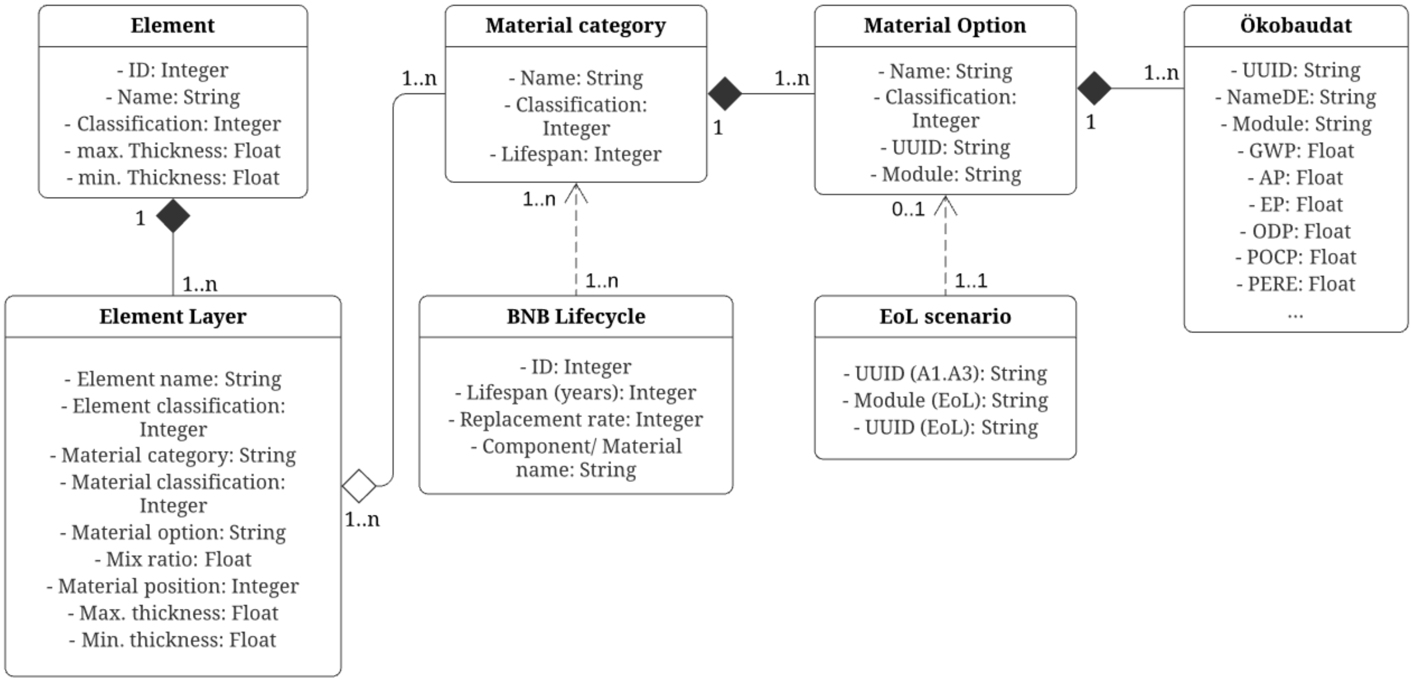

Figure 1: Structure of the LCA Knowledge Database as UML schema

has different levels of detail for a range of several potential design options of components, elements, and materials, including their dependencies. The semantic model healing process happens, when the incomplete IFC element data are matched with the detailed LCA knowledge database (LKdb). First in chapter 3.1, the structure of the LKdb is introduced, before in chapter 3.2 the method for matching is described.

3.1 *LCA Knowledge Database*

The aim of the LCA Knowledge Database is to store all detailed information of typical building elements including all relevant information for calculating a holistic LCA. After the matching of IFC materials to material options in the LKdb and selecting the most similar element, all relevant data are queried for calculating the LCA.

As shown in Figure 1, the general structure of the proposed LKdb consists of three different levels: element, material category and material option. As the LCA database, Ökobaudat was chosen (BBSR 2021), because it consists of more than 1400 datasets specifically of building products and is the most used LCI database in Germany.

The Ökobaudat datasets consist of a Universally Unique Identifier (UUID) and the relevant life cycle modules. All datasets from Ökobaudat consist of several environmental impact categories, such as Global Warming Potential (GWP), Acidification Potential (AP), Eutrophication Potential (EP), Ozone Depletion Potential (ODP), Photochemical Creation Potential (POCP), Primary Energy Renewable (PERE) and many more.

As the quality of some datasets in Ökobaudat is lacking such as missing data of End-of-Life (EoL) module, generic EoL scenarios have to be mapped manually from Ökobaudat. In this case, for each material option two UUID are mapped, one for the LCA Modules A1-A3 and another for the missing End-of-Life scenario (C3/ C4/ D). Stenzel conducted this manual mapping as well as a classification of all UUIDs according to German cost groups using DIN 276 (Stenzel 2020).

All material options have a name and classification according to DIN 276, which is derived by the German name in Ökobaudat. They are called options, as they are the most detailed level of a design option for LCA. Further entries are the UUID, included Modules and the encoded NLP vectors of the name (spans and tokens).

Each material option is related to a material category, which is also stored in Ökobaudat. There are three different levels of categories, but for LKdb only the last level of categories is used, as it groups the datasets of the relevant material options. The category level is extended with another external input describing the service life of building components (BBSR 2017). For these, the IDs are mapped once to the corresponding material category for each classification. Also, for the material categories, the name and the classification are the keys and the encoded NLP vectors of the name (spans and tokens) are also stored.

In the next level, material categories and options are used for setting up element layers. Different elements can consist of the same material category or option. The element layer and the element have default maximum and minimum thicknesses and are also classified to the third level of the German cost group classification according to DIN 276. The element layers can have different mixtures ratios, as e.g. reinforced concrete consists of two material inputs: concrete and reinforcement steel. Each element layer

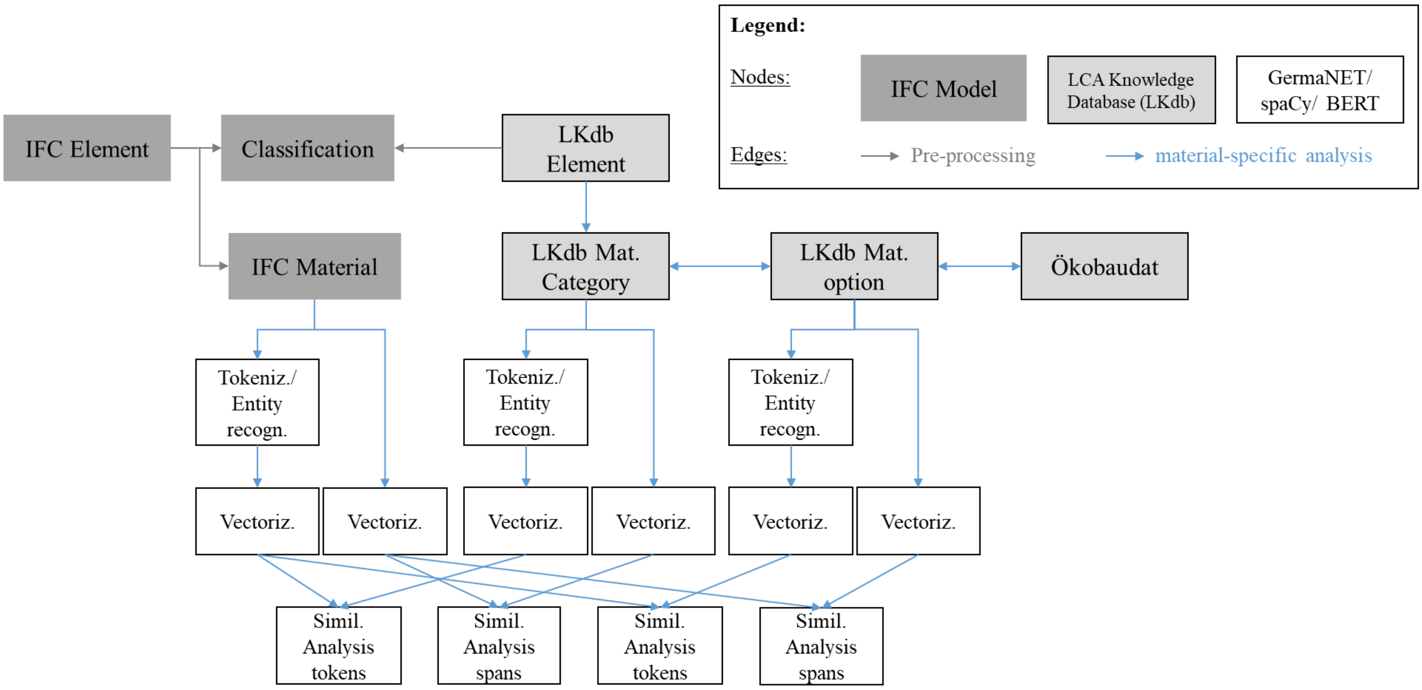

Figure 2: Material-specific similarity analysis of IFC elements and LCA Knowledge Database using GermaNET, spaCy and BERT

has a unique material position, so that elements consist of one or more layers with different material position orders.

3.2 *Method for matching materials*

For matching the elements and materials of an IFC model to the LCA Knowledge Database, we propose employing NLP techniques to measure "semantic similarity" (Forth et al. 2021). Measuring the semantic similarity between the IFC element's material information and the material names of the database involves converting the text of every material type to a vector representation. A vector is a list of numerical values, where the combination of them represents the overall meaning. When comparing two material names, the similarity between vectors A and B can be measured using the cosine similarity, while n is the dimension of the vector:

$$\cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

To compute the vectors for similarity analysis, this paper investigates multiple NLP techniques and evaluates the performance of state-of-the-art deep learning models such as GermaNet (Hamp & Feldweg 1997; Henrich & Hinirchs 2010), SpaCy (Honnibal & Montani 2017), or BERT (Devlin et al. 2018), which will be introduced in the following sections.

Figure 2 shows the general workflow for matching IFC elements to the previously introduced LCA Knowledge Database. Generally, all IFC elements consist of an element name, a required classification and IFC materials, including their names. As the LCA knowledge database is based on the German ÖKOBAUDAT and the German cost group classification system according to DIN 276, it is also required as classification of the IFC elements. The matching method is developed with the aim to perform as robust as possible. For this reason, if there is no IFC material available for specific elements, the element name itself will be evaluated for NLP similarity analysis. Furthermore, only the elements corresponding to the classification are filtered and compared. The material names from the IFC elements as well as from the LCA Knowledge Database can be encoded either as whole expressions/ spans or be tokenized beforehand.

3.3 *NLP techniques*

This section introduces the three NLP techniques GermaNET, spaCy and BERT. In the next section, the performance of these techniques is evaluated and compared for measuring the similarity between the different material types.

3.3.1 *GermaNET*

GermaNET is a Lexical-Semantic Net specialized for the German language, also known as the German version of the Princeton WordNet (Hamp & Feldweg 1997; Henrich & Hinirchs 2010). GermaNET relates German nouns, verbs and adjectives semantically by grouping lexical units that express the same concept into synsets (set of synonyms) and by defining semantic relations between these synsets. It can be represented as a graph, whose nodes are synsets and edges represent the semantic relations (Navigli & Martelli 2019). Therefore, the similarity is not measured using cosine similarity, but graph-related shortest path similarity, which is equal to the inverse of the shortest path length between two synsets. There are other path-related similarity analyses such as Wu-

Palmer similarity or Leacock-Chodorow similarity, which were not considered in this paper.

3.3.2 *spaCy*

SpaCy is a pre-trained neural network model which offers state of-the-art accuracy in multiple languages (Honnibal & Montani 2017). Its large German model (de_core_news_lg) includes 500k unique vectors in its corpus and represents every word or expression with a vector of 300 dimensions. As sources for training data, existing corpi were used such as e.g. TiGer Corpus (Brants et al. 2004).

3.3.3 *BERT*

BERT stands for Bidirectional Encoder Representations from Transformers and was released by Google in 2018 (Devlin et al. 2018). Transformers-based pre-trained models are currently state-of-the-art and are capable of solving a different set of tasks as they "can represent the characteristics of word usage such as syntax and how words are used in various contexts" (Locatelli et al. 2021). Nevertheless, BERT represents each word or expression with a vector of 768 dimensions, which is significantly higher compared to spaCy, making the similarity calculation more time-consuming.

4 EXPERIMENTS & RESULTS

In the following sections, first the case study is shortly introduced. Afterwards, the performance results of three different NLP techniques based on a manual matching is compared. Last, one IFC element is chosen to be prototypically matched and the LCA calculation is conducted using the LKdb and compared to conventional workflow results.

4.1 *Case study*

For comparing the three different NLP techniques and their performance of their workflows, a real-world office building was chosen as a case study. This real-world project guarantees that the material naming is not optimized but according to current industry standards, so that the matching performances are tested under realistic conditions.

In total, the case study office model consists of 2110 individual elements, which are summed up to 133 unique elements when grouped by element type. Those consists of 59 unique IFC materials, which were manually matched to LCA material options and categories from the LKdb, as a ground truth.

4.2 *Results*

The following results of each NLP technique performance are based on the 59 pairs of IFC materials and the matched LCA material options and categories from the LKdb based on Ökobaudat.

4.2.1 *GermaNET*

As the workflow of the GermaNET differs from the other two NLP techniques, the identification rate of the synsets need to be analyzed before analyzing the shortest path similarity.

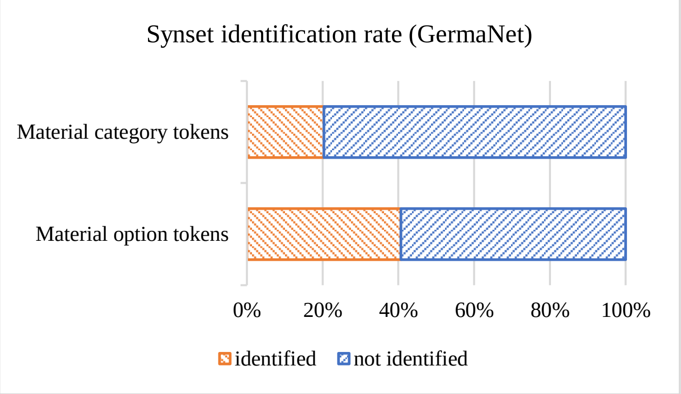

Figure 3: Synset identification rate of material pairs with GermaNET

After the tokenization of the IFC material names, material options and their related material categories, synsets were identified to calculate the shortest path similarity. Nevertheless, synsets could not be identified for every token set, so that not for all 59 pairs synsets could be identified. As shown in Figure 3, only for 20,3% of the material category tokens and 40,7% of the material option tokens, a pair of synsets could be identified.

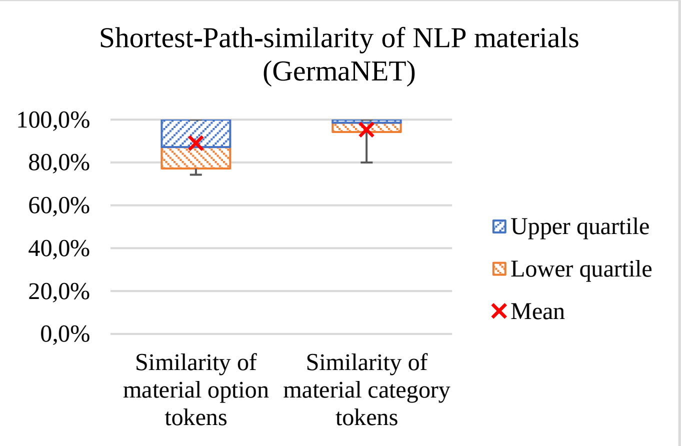

Figure 4: Shortest path similarity of NLP material using GermaNET

Nevertheless, the shortest path similarity of the identified pairs of synsets show promising results (Figure 4). The mean of the similarity of material option tokens is 88,9% and of the material category tokens even 95,2%, both with little deviation. However, including the little synset identification rate of both, material options and material categories from the LKdb, the total similarity is very low and not sufficient for being used in the proposed matching methodology.

4.2.2 *spaCy*

For the results of spaCy and BERT, the similarities of tokens and whole spans of the material options and material categories are compared according Figure 2.

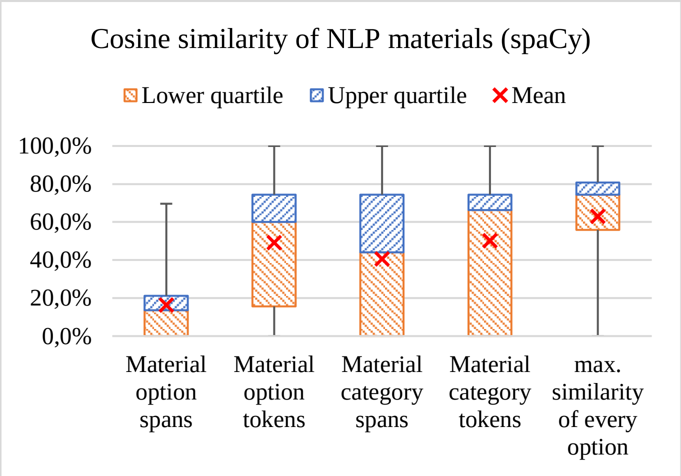

Figure 5: Cosine similarity of NLP materials using spaCy

As shown in Figure 5, the ranges of the cosine similarity of all different comparisons differ a lot. Generally, the similarity of IFC materials to the material option spans have the worst performance with the mean being at 16,3%. This means, that the similarity for most matched pairs using material option spans and spaCy is only very little. The tokenization improves the performance of matching the material performances up to a mean of 49,2%. Also, the spans of the material categories are much better (mean at 40,5%). The tokenization of the material categories improves the performance results up to 50,2%. As an additional, performance result, the maximum similarity of all options (material option spans and tokens, as well as material category spans and tokens) is calculated. Its mean is 63,0%, but also the quartile ranges improved, compared to all other ranges. In general, the results are not sufficient, but show a promising strategy of getting the maximum similarity of every option.

4.2.3 *BERT*

When evaluating BERT, the same similarity results are calculated as previously shown with spaCy also using cosine similarity. Figure 6 is showing the results as ranges of the material option spans and tokens and material category spans and tokens. Generally, all result ranges differ much less compared to the results using spaCy, which means that for all pairs more satisfying performances can be reached using BERT.

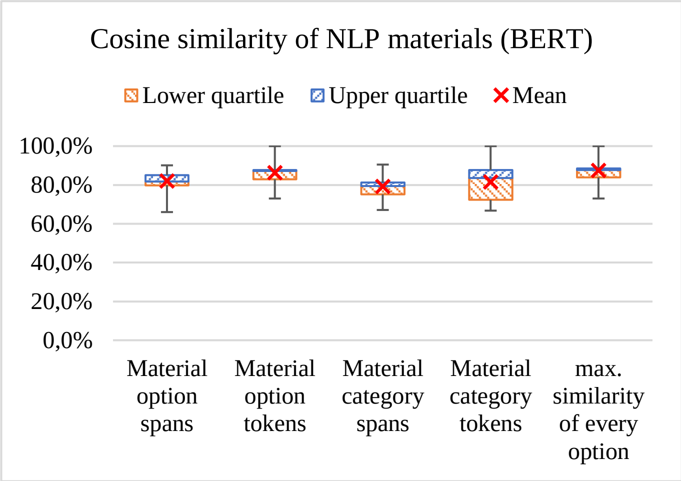

Figure 6: Cosine similarity of NLP materials using BERT

Additionally, all means are between 79,3% (material category spans) and 86,4% (material option tokens). Also, the strategy of getting the maximum similarity of every option is improving the general promising results (mean 87,6%).

Also, the minimum values of each result ranges show that BERT, generally performs much better than spaCy.

4.3 *Summary*

Generally, all three NLP techniques could be applied to the case study. Although GermaNET showed promising results in the ranges of shortest path similarity, the identification rate of synsets was too low. Therefore, a further implementation to the proposed matching methodology was not pursued further in this paper. The second tested NLP technique, spaCy, showed that different strategies of calculating the cosine similarity of material option spans and material category spans are improving the results. Furthermore, the tokenization of both material options and material categories, as well as choosing the maximum similarity of every calculated option improved the result ranges significantly. However, the deviations in ranges were substantial and are generally too low, so that a further consideration for implementation was not investigated. Finally, BERT showed the most promising results. Low deviations of the result ranges and high cosine similarity of all strategies lead to a further implementation of the matching approach. Nevertheless, due to its large vectors with 786 dimensions, the calculation time is significantly higher than with spaCy and needs to be considered for further optimization.

4.4 *Prototypical element matching and calculation of LCA results*

Next, the proposed matching methodology was prototypically tested using LKdb and BERT. The LKdb was filled with example elements and element layers, based on domain knowledge and the structured Ökobaudat. As a test element, the exterior wall "Basiswand: STB 250_außen" from the case study was chosen (cost group 331, single material "Ortbeton", total area 415.32 m², layer thickness 20 cm).

The final matching shows if the highest cosine similarity was derived from a material category or the material option. In this test case, it is the material category with a cosine similarity of 85,4%. Therefore, the matched element within the cost group 331 is "Stahlbeton", so also the reinforcement steel is included beside the range of different concrete options.

As a comparison for manual matching and manual calculation, the software eLCA is used (BBSR). Only specific datasets of the Ökobaudat can be used. Therefore, the assumed LCA dataset is "Transportbeton C20/25".

For simplicity, GWP [kg CO_2-eq./a] is chosen as the comparing indicator with a lifespan of 50 years. The results are shown in Figure 7:

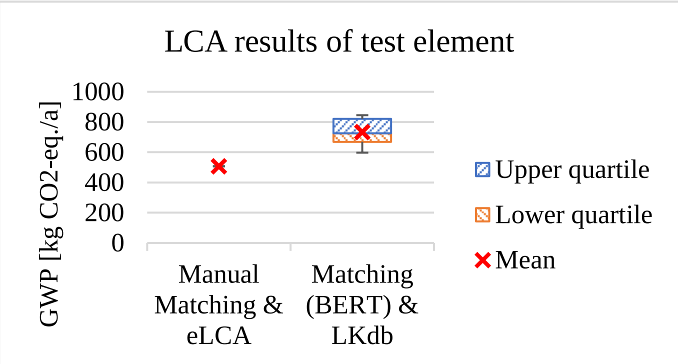

Figure 7: LCA result (GWP) of test element comparing manual matching with BERT matching and LKdb

Besides the necessary effort and knowledge of the manual matching, the accuracy of the results is different. While the result with manual matching and eLCA is a single value, the proposed methodology returns a material category, considering the uncertainty of material choice in the early design stages. Therefore, the LKdb returns a range of material options for LCA calculation. Because the matching does not take place on a material level, but on an element level, the reinforcement steel is getting included in the LKdb element of reinforced concrete, which gives more realistic results. This is the reason, why the range of results is more than 200 kg CO_2-eq./a higher than the manually matched eLCA result and is therefor more correct.

Accordingly, it can be stated that by using the LKdb and proposed matching methodology the inaccurate BIM model can be semantically healed for a more accurate LCA in early design stages.

5 DISCUSSION & OUTLOOK

By semantically healing BIM models for LCA, the analysis of embodied carbon becomes holistically more consistent and more comparable for early design stages. Furthermore, the LCA knowledge database provides design options for optimizing the building performance according to LCA results. The limitations of this research are the chosen LCA database (Ökobaudat) and the correlating German language. Other NLP models of different languages might perform differently, as well other LCA databases might have less datasets.

In a next step, the matching should be carried out on multiple case studies and verified with manually calculated LCA results. Furthermore, the performance shall be increased by checking domain specific abbreviations, as for example "STB" stands for "Stahlbeton" (reinforced concrete), but could not be identified by existing NLP models. As in this paper, the focus was set to material matching for comparing the performance of several NLP models, in a next step the element-specific matching shall be included in the performance analysis. Finally, to make the method more robust, commonly used elements for each classification with default values shall be defined in the LCA Knowledge database.

6 REFERENCES

Abergel, Thibaut; Dean, Brian; Dulac, John (2017): Global Status Report 2017. Hg. v. United Nations Environment Programme. International Energy Agency (IEA) for the Global Alliance for Buildings and Construction (GABC).

BBSR: eLCA. Online verfügbar unter https://www.bauteileditor.de/, zuletzt geprüft am 29.12.2021.

BBSR (2017): Nutzungsdauern von Bauteilen - Informationsportal Nachhaltiges Bauen. Hg. v. Bundesinstitut für Bau-, Stadt- und Raumforschung. Online verfügbar unter http://www.relaunch-nb.online-now.de/index.php?id=91&L=0, zuletzt geprüft am 28.12.2021.

BBSR (2021): ÖKOBAUDAT. Online verfügbar unter https://www.oekobaudat.de/datenbank/browser-oekobaudat.html, zuletzt aktualisiert am 06.04.2021, zuletzt geprüft am 06.04.2021.

Brants, Sabine; Dipper, Stefanie; Eisenberg, Peter; Hansen-Schirra, Silvia; König, Esther; Lezius, Wolfgang et al. (2004): TIGER: Linguistic Interpretation of a German Corpus. In: *Res Lang Comput* 2 (4), S. 597–620. DOI: 10.1007/s11168-004-7431-3.

Braune, Anna; Ruiz Durán, Christine; Gantner, Johannes (2018): Leitfaden zum Einsatz der Ökobilanzierung.

Costa, G.; Sicilia, A. (2020): Alternatives for facilitating automatic transformation of BIM data using semantic query languages. In: *Automation in Construction* 120, S. 103384. DOI: 10.1016/j.autcon.2020.103384.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Online verfügbar unter https://arxiv.org/pdf/1810.04805.

Eleftheriadis, S.; Duffour, P.; Mumovic, D. (2018): BIM-embedded life cycle carbon assessment of RC buildings using optimised structural design alternatives. In: *Energy and Buildings* 173, S. 587–600. DOI: 10.1016/j.enbuild.2018.05.042.

Forth, Kasimir; Abualdenien, Jimmy; Borrmann, André; Fellermann, Sabrina; Schunicht, Christian (2021): Design optimization approach comparing multicriterial variants using BIM in early design stages. In: *Proceedings of 38th International Symposium on Automation and Robotics in Construction (ISARC 2021)*, S. 235–242. DOI: 10.22260/ISARC2021/0034.

Hamp, Birgit; Feldweg, Helmut (1997): GermaNet - a Lexical-Semantic Net for German. In: In Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, S. 9–15.

Henrich, Verena; Hinirchs, Erhard (2010): GernEdiT – The GermaNet Editing Tool. In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010). Valletta, Malta*, pp. 2228-2235.

Hollberg, Alexander (2016): A parametric method for building design optimization based on Life Cycle Assessment. PhD thesis. Bauhaus-Universität Weimar.

Honnibal, Matthew; Montani, Ines (2017): spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. In: *To appear* 7 (1), S. 411–420.

Horn, Rafael; Ebertshäuser, Sebastian; Di Bari, Roberta; Jorgji, Olivia; Traunspurger, René; Both, Petra von (2020): The BIM2LCA Approach: An Industry Foundation Classes (IFC)-Based Interface to Integrate Life Cycle Assessment in Integral Planning. In: *Sustainability* 12 (16), S. 6558.

Koo, Bonsang; Jung, Raekyu; Yu, Youngsu (2021): Automatic classification of wall and door BIM element subtypes using 3D geometric deep neural networks. In: *Advanced Engineering Informatics* 47, S. 101200. DOI: 10.1016/j.aei.2020.101200.

Llatas, Carmen; Soust-Verdaguer, Bernardette; Passer, Alexander (2020): Implementing Life Cycle Sustainability Assessment during design stages in Building Information Modelling: From systematic literature review to a methodological approach. In: *Building and Environment* 182, S. 107164. DOI: 10.1016/j.buildenv.2020.107164.

Locatelli, Mirko; Seghezzi, Elena; Pellegrini, Laura; Tagliabue, Lavinia Chiara; Di Giuda, Giuseppe Martino (2021): Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis. In: *Buildings* 11 (12), S. 583. DOI: 10.3390/buildings11120583.

Navigli, Roberto; Martelli, Federico (2019): An overview of word and sense similarity. In: *Nat. Lang. Eng.* 25 (06), S. 693–714. DOI: 10.1017/S1351324919000305.

Reitschmidt, Georg (2015): Ökobilanzierung auf Basis von Building Information Modeling. Entwicklung eines Instruments zur automatisierten Ökobilanzierung der Herstellungsphase von Bauwerken unter Nutzung der Ökobau.dat und Building Information Modeling. Masterthesis. Technische Hochschule Mittelhessen, Gießen.

Rezaei, Farzaneh; Bulle, Cécile; Lesage, Pascal (2019): Integrating building information modeling and life cycle assessment in the early and detailed building design stages. In: *Building and Environment* 153, S. 158–167. DOI: 10.1016/j.buildenv.2019.01.034.

Safari, Kaveh; AzariJafari, Hessam (2021): Challenges and opportunities for integrating BIM and LCA: Methodological choices and framework development. In: *Sustainable Cities and Society* 67, S. 102728. DOI: 10.1016/j.scs.2021.102728.

Schneider-Marin, Patricia; Harter, Hannes; Tkachuk, Konstantin; Lang, Werner (2020): Uncertainty Analysis of Embedded Energy and Greenhouse Gas Emissions Using BIM in Early Design Stages. In: *Sustainability* 12 (7), S. 2633. DOI: 10.3390/su12072633.

Stenzel, Valérie (2020): Wissensdatenbank für Graue Energie und Treibhauspotenzial von Baustoffen. Masterthesis. München, Technische Universität München. Lehrstuhl für energieeffizientes und nachhaltiges Planen und Bauen.

Wastiels, L.; Decuypere, R. (2019): Identification and comparison of LCA-BIM integration strategies. In: *IOP Conf. Ser.: Earth Environ. Sci.* 323, S. 12101. DOI: 10.1088/1755-1315/323/1/012101.

Wu, Songfei; Shen, Qiyu; Deng, Yichuan; Cheng, Jack (2019): Natural-language-based intelligent retrieval engine for BIM object database. In: *Computers in Industry* 108, S. 73–88. DOI: 10.1016/j.compind.2019.02.016.