

Master of Science FHNW in Virtual Design and Construction (MSc FHNW VDC)



Smart Semantics for Sustainable Infrastructure

Ein NLP-basierter Ansatz zur kontextsensitiven Materialklassifikation für die
BIM-basierte Ökobilanzierung im Schweizer Infrastrukturbau

Colin Macdonald

22.05.2026

Thesis-Begleiter: Prof. Lukas Schildknecht

Thesis-Experte: Louis Trümpler

Praxispartner*in: AFRY Schweiz AG



Inhaltsverzeichnis

Eigenständigkeitserklärung	vi
Abstract	vii
Vorwort	viii
Abkürzungsverzeichnis	ix
Glossar	x
1 Einleitung	1
1.1 Motivation und Problemstellung	1
1.2 Zielsetzung und Forschungsfragen	1
1.3 Aufbau der Arbeit	3
2 Grundlagen	4
2.1 Normen, Standards und Daten	4
2.1.1 Normen	4
2.1.2 SNBS und NNBS	6
2.1.3 Ökobilanzdaten des KBOB	7
2.2 Ökobilanzierung in der Schweiz	8
2.2.1 Methode der ökologischen Knappheit	8
2.2.2 Ökobilanzierung im Infrastrukturbau	9
2.3 IFC-Standard	10
2.4 Künstliche Intelligenz und Natural Language Processing	11
2.4.1 Künstliche Intelligenz	11
2.4.2 Semantic Textual Similarity und Embedding-Modelle	13
3 Stand der Technik	14
3.1 BIM-LCA-Integration als semantisches Matching-Problem	14
3.2 Semantische Anreicherung von IFC-Modellen	14
3.3 NLP- und LLM-basierte Matching-Verfahren	15
3.3.1 Semantic Model Healing und Embedding-basierte Verfahren	15
3.3.2 Fine-Tuning, LLMs und RAG	15
3.3.3 Hybride Ansätze zur Absicherung semantischer Matches	16
3.4 LCA-Tools	16
3.5 Forschungslücken und Herausforderungen	16
4 Methodik	18
4.1 Literatur- und Dokumentenanalyse	18

4.2	Anforderungsanalyse	19
4.3	Konzeption des Prototyps	20
4.4	Prototypische Umsetzung	20
4.5	Evaluation	20
4.6	Gütekriterien und Abgrenzung	21
5	Konzept	22
5.1	Zielbild des Prototyps	22
5.2	Konzeptioneller Systemablauf	23
5.3	Datenmodell für IFC- und KBOB-Informationen	24
5.3.1	IFC-Informationen	24
5.3.2	Kontextstring für das semantische Matching	25
5.3.3	KBOB-Informationen	26
5.3.4	Mapping-Datensatz	27
5.4	Matching-Strategie	27
5.4.1	Regelbasierte Prüfung	28
5.4.2	Embedding-basiertes Matching	28
5.4.3	Bewertung und Sortierung der Kandidaten	29
5.5	Human-in-the-loop und Export von Trainingsdaten	29
5.6	Konzept für das Fine-Tuning	30
5.6.1	Metriken	31
5.7	Ergebnisdarstellung	33
5.8	Rückbindung an das IFC-Modell	33
5.9	Abgrenzungen und konzeptionelle Entscheidungen	34
5.10	Konzeptionelle Herausforderungen	34
5.11	Zusammenfassung des Konzepts	35
6	Implementierung des Prototyps	36
6.1	Softwarearchitektur und Datenfluss	36
6.1.1	Komponenten des Prototyps	36
6.1.2	Datenformate und Übergaben	38
6.2	IFC-Extraktion und Datenaufbereitung	39
6.2.1	Exportierte IFC-Informationen	39
6.2.2	Aggregationen, Coverings und Bewehrung	40
6.2.3	Materialschichten, Mengen und Einheiten	41
6.3	KBOB-Datenbasis und Umweltberechnung	42
6.3.1	Datenbasis und Umweltkennwerte	42
6.3.2	Berechnung der Umweltindikatoren	42
6.4	Semantisches Matching	43
6.4.1	Query-Bildung und Kandidatensuche	43
6.4.2	Score, Kandidatenliste und manuelle Auswahl	44

6.4.3	Export korrigierter Zuordnungen	44
6.5	Fine-Tuning des Bi-Encoders	44
6.5.1	Trainingsdaten und Aufbereitung	45
6.5.2	Hard-Negative Mining	45
6.5.3	Training mit MultipleNegativesRankingLoss	45
6.6	Technische Grenzen und Zwischenfazit	47
7	Ergebnisse	48
7.1	Evaluationsaufbau	48
7.2	Trefferqualität auf kontrollierten Testdaten	49
7.2.1	Gesamtergebnis	49
7.2.2	Analyse nach Materialfamilie	49
7.2.3	Analyse nach Herstellungsart und Query-Klasse	50
7.2.4	Analyse der verbleibenden Fehler	50
7.3	Robustheit gegenüber fehlerhaften Eingaben	52
7.3.1	Erzeugung der degradierten Queries	52
7.3.2	Queries mit Schreibfehlern	52
7.3.3	Queries mit fehlenden Angaben	53
7.3.4	Queries mit fehlenden Angaben und Schreibfehlern	53
7.4	Vergleich mit dem Basismodell und weiteren Modellen	54
7.5	Anwendung auf reale Infrastrukturprojekte	55
7.6	Zusammenfassung der Ergebnisse	56
8	Diskussion und Ausblick	58
8.1	Fehleranalyse und Diskussion	58
8.2	Limitationen der Arbeit	59
8.3	Implikationen für Forschung und Praxis	60
8.3.1	Implikationen für die Praxis	60
8.3.2	Implikationen für die Forschung	61
8.4	Zusammenfassung der wichtigsten Erkenntnisse	61
8.5	Ausblick auf zukünftige Entwicklungen	62
8.5.1	Kurzfristige Verbesserungen	62
8.5.2	Mittelfristige Erweiterungen	62
8.5.3	Langfristige Forschungsrichtungen	62
	Literaturverzeichnis	63
	Abbildungsverzeichnis	66
	Tabellenverzeichnis	67
	Verwendete Hilfsmittel	68

Anhang	69
A.1 Auszug KBOB-Ökobilanzdaten	69
A.2 Übersicht LCA-Tools	70
A.2.1 Erläuterungen zu den LCA-Tools	71
A.3 Von der Forschungsfrage zum Suchstring	72
A.4 Detaillierter Systemablauf	74
A.5 Klassendiagramm IFC 4.3	75
A.6 Datenbank mit KBOB-Ökobilanzdaten	76
A.7 Bewertung offizielle SBERT-Modelle	77
A.8 Evaluation SBERT-Modelle	78
A.9 Verteilungen der Trainingsdaten	79
A.10 Mindestanforderungen an das IFC-Modell	84
A.11 Dashboard Prototyp	85
A.12 Vergleich der Modelle mit den Test-Queries	89
A.13 Validierung Prototyp	90

Eigenständigkeitserklärung

«Ich erkläre hiermit,

dass ich die vorliegende Master-Thesis mit dem Titel «Smart Semantics for Sustainable Infrastructure» selbst und selbständig verfasst habe,

dass ich sämtliche nicht von mir selbst stammenden Textstellen bzw. Bestandteile eines Werkes (Bilder, Grafiken, Codes, etc.) gemäss gängigen wissenschaftlichen Zitierregeln korrekt zitiert und die verwendeten Quellen gut sichtbar erwähnt habe;

dass ich in einem Verzeichnis alle verwendeten Hilfsmittel (KI-Assistenzsysteme wie Chatbots [z. B. ChatGPT], Übersetzungs- [z. B. DeepL] Paraphrasier- [z. B. Quillbot]) oder Programmierapplikationen [z. B. Github Copilot] deklariert und ihre Art der Verwendung offenlege und bei den entsprechenden Textstellen angegeben habe,

dass ich sämtliche immateriellen Rechte an von mir allfällig verwendeten Materialien wie Bilder oder Grafiken erworben habe oder dass diese Materialien von mir selbst erstellt wurden;

dass das Thema, die Arbeit oder Teile davon nicht bei einem Leistungsnachweis eines anderen Moduls verwendet wurden, sofern dies nicht ausdrücklich mit der Dozentin oder dem Dozenten im Voraus vereinbart wurde und in der Arbeit ausgewiesen wird;

dass ich mir bewusst bin, dass meine Arbeit auf Plagiate und auf Drittautorschaft menschlichen oder technischen Ursprungs (künstliche Intelligenz) überprüft werden kann;

dass ich mir bewusst bin, dass die Hochschule für Architektur, Bau und Geomatik einen Verstoss gegen diese Eigenständigkeitserklärung bzw. die ihr zugrundeliegenden Studierendenpflichten der Studien- und Prüfungsordnung der Hochschule für Architektur, Bau und Geomatik verfolgt und dass daraus disziplinarische (Verweis oder Ausschluss aus dem Studiengang) Folgen resultieren können.»

Colin Macdonald

Winterthur, 22.05.2026

Abstract

Im Schweizer Infrastrukturbau bilden die KBOB-Ökobilanzdaten die Grundlage für die Nachhaltigkeitsbewertung von Bauwerken. Die Zuordnung von Materialien aus digitalen Bauwerksmodellen zu den passenden Ökobilanzdaten erfolgt heute manuell. Dieser Prozess ist zeitaufwändig, fehleranfällig und wird, wenn überhaupt, erst in späten Projektphasen durchgeführt.

Diese Arbeit entwickelt und evaluiert ein KI-gestütztes Verfahren, das genau an diesem Problem ansetzt. Es verringert den manuellen Aufwand, unterstützt die Zuordnung durch konsistente, nachvollziehbare Vorschläge und ordnet Materialien und Bauteile aus IFC-Modellen automatisch den passenden KBOB-Einträgen zu. Dazu werden Informationen wie Bauteiltyp, Materialbezeichnung und Festigkeitsklasse zu einem Suchtext kombiniert, den ein domänenspezifisch trainiertes Sprachmodell mit 110 infrastrukturelevanten KBOB-Einträgen vergleicht. Der Prototyp arbeitet als Entscheidungsunterstützung: Er liefert eine nach Ähnlichkeit sortierte Rangliste, die Fachperson trifft die finale Entscheidung.

Auf Testdaten mit 389 Abfragen erreicht das trainierte Modell eine Trefferquote von 96,9 %. Es übertrifft damit das LLM «GPT-5.5 Thinking» um 10 Prozentpunkte. Auch bei fehlerhaften Eingaben oder fehlenden Attributen bleibt der korrekte Eintrag in 87,7 % der Fälle unter den ersten zehn Vorschlägen. Auf realen Brückenmodellen liegt die Genauigkeit zwischen 60 und 89 %, abhängig von der Vollständigkeit der Modelldaten. Die Analyse zeigt, dass nicht der Algorithmus, sondern die Qualität der im IFC-Modell hinterlegten Informationen der limitierende Faktor ist. Modellierungsrichtlinien, die vollständige Angaben sicherstellen, haben einen grösseren Einfluss als Verbesserungen am Matching-Verfahren selbst.

Keywords

Künstliche Intelligenz (KI), Ökobilanzierung, Natural Language Processing (NLP), Building Information Modeling (BIM), KBOB, Infrastrukturbau, Semantische Ähnlichkeit

Vorwort

Wer Brücken baut, entwickelt ein solides Vertrauen in Tragfähigkeit. Ich habe im Bauingenieurwesen früh gelernt, dass gute Ergebnisse auf sauberen Grundlagen beruhen: auf nachvollziehbaren Annahmen, geprüften Modellen und belastbaren Informationen. Diese Lektion endet für mich nicht bei Beton und Stahl. Sie beginnt dort von Neuem, wo Bauwerke digital beschrieben werden.

Diese Erkenntnis wurde mir spätestens bewusst, als begann mit IFC-Daten zu arbeiten. Was auf den ersten Blick nach digitaler Ordnung aussieht, entpuppt sich bei genauerem Hinsehen oft als Sammlung gut gemeinter Informationen von unterschiedlicher Qualität. Ein Material heisst hier so, dort anders und wird an dritter Stelle gar nicht benannt. Damit zeigen sich die Baustellen digitaler Modelle sehr konkret.

Genau an dieser Schnittstelle setzt der Masterstudiengang Virtual Design and Construction (MSc FHNW VDC) an. Digitales Bauen zeigt sich dabei nicht als Frage einzelner Programme, sondern als Zusammenspiel von Modellen, Informationen, Prozessen und Menschen. Diese Sichtweise hat meinen Blick auf Bauwerksdaten verändert. In der Praxis scheitert Digitalisierung selten an falschen Werkzeugen. Häufiger fehlen Informationen dort, in der Form oder in der Qualität, wie man sie für die nächste Entscheidung bräuchte.

Aus dieser Beobachtung entstand das Thema meiner Masterthesis. Die Zuordnung von Materialinformationen aus IFC-Modellen zu Ökobilanzdaten wirkt zunächst einfach: Material lesen, passenden KBOB-Eintrag finden, fertig. In der Realität sind Materialbezeichnungen uneinheitlich, Modellinformationen lückenhaft und passende Einträge nicht immer eindeutig.

Daraus ergab sich für mich die Frage, warum Materialien im Jahr 2026 noch immer manuell einer Datenbank zugeordnet werden, obwohl KI längst passende Vorschläge machen kann, ohne dass die Nachvollziehbarkeit verloren geht.

Wer mich kennt, weiss, dass ich bei komplexen Fragen gerne länger dranbleibe, ob am Modell, am Schachbrett oder an einer Zuordnung, die nicht ganz aufgehen will. Auch diese Arbeit war in dieser Hinsicht konsequent: Viele scheinbar kleine Entscheidungen führten zu grösseren Fragen über Datenqualität, Nachvollziehbarkeit und die Rolle künstlicher Intelligenz im digitalen Bauen.

Ein besonderer Dank gilt meinem Thesis-Experten Louis Trümpler für seine fachliche Schärfe und sein direktes Feedback. Seine Open-Source-Tools für BIM und nachhaltiges Bauen haben diese Arbeit mitgeprägt. Ebenso danke ich Prof. Lukas Schildknecht für die konstruktive Begleitung, die wertvollen Impulse und seine kritischen Fragen.

Abschliessend danke ich allen, die mich im Studium und dieser Arbeit unterstützt, ermutigt und geduldig begleitet haben. Besonders danke ich jenen, die zur richtigen Zeit für Abstand gesorgt haben, mit guten Gesprächen, kurzen Blitzschachpartien und gelegentlichen Pausen abseits des Schreibtischs.

Abkürzungsverzeichnis

SIA	Schweizerischer Ingenieur- und Architektenverein
BAFU	Bundesamt für Umwelt
BERT	Bidirectional Encoder Representations from Transformers
BIM	Building Information Modeling
CO₂-eq	Kohlendioxid-Äquivalente
DL	Deep Learning (Tiefes Lernen)
EPD	Environmental Product Declaration (Umweltproduktdeklaration)
IFC	Industry Foundation Classes
KBOB	Koordinationskonferenz der Bau- und Liegenschaftsorgane der öffentlichen Bauherren
KI	Künstliche Intelligenz
LCA	Life Cycle Assessment (Ökobilanz)
LLM	Large Language Model (Grosses Sprachmodell)
ML	Machine Learning (Maschinelles Lernen)
NLP	Natural Language Processing (Verarbeitung natürlicher Sprache)
NNBS	Netzwerk Nachhaltiges Bauen Schweiz
SNBS	Standard Nachhaltiges Bauen Schweiz
STS	Semantic Textual Similarity
UBP	Umweltbelastungspunkte

Glossar

BERT

BERT ist ein vortrainiertes, bidirektionales Sprachmodell auf Basis von Transformer-Architektur, das durch gleichzeitige Berücksichtigung von linker und rechter Kontextinformation in allen Schichten vielseitig für NLP-Aufgaben feinjustiert werden kann (Devlin et al., 2019, 1).

DL

«Deep Learning (DL) ist maschinelles Lernen in Künstlichen Neuronalen Netzen mit mehreren bis sehr vielen Schichten, die aus einer Vielzahl künstlicher Neuronen zusammengesetzt sind. Tiefes Lernen ist verantwortlich für die Erfolge in der Sprach- und Text-, Bild- und Videoverarbeitung.» (KI.NRW, 2025).

EPD

Eine Environmental Product Declaration (EPD) ist eine standardisierte Umweltdeklaration, die die Umweltwirkungen eines Produkts über seinen Lebenszyklus transparent und vergleichbar darstellt.

LCA

«Zusammenstellung und Beurteilung der Input- und Outputflüsse und der potentiellen Umweltwirkungen eines Produktsystems im Verlauf seines Lebensweges» (SN EN 15643, 2021, 17).

LLM

«Große Sprachmodelle (LLMs) sind Basismodelle oder Foundation Models, für die Verarbeitung natürlicher Sprache mit großen Mengen von Textdaten trainiert wurden. Die Modelle lernen, Texte fortzusetzen, indem sie statistische Beziehungen zwischen Wörtern herstellen, und damit Wissen über Syntax, Semantik und Ontologie der Sprache, aufbauen» (KI.NRW, 2025).

ML

«Machine Learning (ML) bezweckt die Generierung von Wissen aus Erfahrungswerten, indem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln. Das Modell kann anschließend auf neue, potenziell unbekannte Daten derselben Art angewendet werden. Damit kommt das Maschinelle Lernen ohne manuelle Wissensangabe oder explizite Programmierung eines Lösungswegs aus (KI.NRW, 2025).»

Nachhaltigkeit

«Zustand des Gesamtsystems, einschließlich der umweltbezogenen, sozialen und ökonomischen Aspekte, innerhalb dessen gegenwärtige Bedürfnisse erfüllt werden, ohne die Fähigkeit zukünftiger Generationen zur Erfüllung ihrer eigenen Bedürfnisse zu beeinträchtigen» (SN EN 15643, 2021, 24).

NLP

«Maschinelle Sprachverarbeitung (Natural Language Processing = NLP) umfasst Techniken zur Erkennung, Interpretation und Erzeugung von natürlicher Sprache in Wort und Schrift. Dazu gehören die Vertextung gesprochener Sprache, Stimmungserkennung, Informationsextraktion aus Texten, maschinelle Übersetzung und das Führen von Gesprächen» (KI.NRW, 2025).

1 Einleitung

1.1 Motivation und Problemstellung

Die Bauwirtschaft ist ein zentraler Treiber der CO₂-Emissionen in der Schweiz (BAFU, 2025). «Mit einem jährlichen Investitionsvolumen von rund 60 Milliarden Franken hat die Schweizer Baubranche ein enormes Potenzial, um nachhaltig zu handeln und damit auch positive Auswirkungen auf [...] das Klima [...] zu erzielen» (NNBS, 2025). Damit dieses Potenzial genutzt werden kann, braucht es verlässliche Werkzeuge zur Bewertung der Umweltwirkungen von Bauprojekten über den gesamten Lebenszyklus.

Die Ökobilanzierung, also die systematische Erfassung und Bewertung von Umweltwirkungen eines Bauwerks, ist das zentrale Instrument dafür. In der aktuellen Baupraxis hat sie jedoch «noch nicht den Stellenwert, der notwendig wäre, um z. B. graue Emissionen zu verringern» (Hermann et al., 2024, 128). Die Erstellung von Ökobilanzen im Infrastrukturbau ist zeitaufwändig, fehleranfällig und wird oft erst in späten Projektphasen durchgeführt. Ein BIM-basierter Einsatz in frühen Planungsphasen würde den Vergleich verschiedener Projektvarianten ermöglichen und datenbasierte, klimaschonende Entscheidungen unterstützen.

Mit dem Kriterienbeschrieb Infrastruktur (SNBS und NNBS, 2026) wurde in der Schweiz ein Bewertungsrahmen für die Nachhaltigkeit von Infrastrukturprojekten geschaffen. Bisher fehlen jedoch Werkzeuge, die diese Vorgaben effizient in BIM-Prozesse integrieren. Bestehende Ansätze zur BIM-basierten Ökobilanzierung verknüpfen zwar Modelldaten mit EPDs, setzen aber eine manuelle Zuordnung der Materialien voraus (Hermann et al., 2024, 129). Erste KI-gestützte Verfahren existieren für den deutschen Hochbau und die ÖKOBAUDAT (siehe Kapitel 3). Ein kontextbasiertes Mapping, das Materialbezeichnungen aus IFC-Infrastrukturmodellen automatisiert den passenden Schweizer KBOB-Ökobilanzdaten zuordnet, wurde bisher nicht umgesetzt.

Durch meine Tätigkeit bei AFRY habe ich Zugang zu realen IFC-Modellen aus Infrastrukturprojekten und kenne die praktischen Herausforderungen bei einer Ökobilanzierung im Planungsalltag. Die Verbindung von Nachhaltigkeit und KI eröffnet ein zukunftsorientiertes Arbeitsfeld, das mich motiviert, einen konkreten Beitrag zur digitalen Transformation im Schweizer Infrastrukturbau zu leisten. In der Integration einer KI sehe ich grosses Potenzial, die Datenverarbeitung zu automatisieren, präzisere Ökobilanzen zu ermöglichen und Planende gezielt bei klimaschonenden Entscheidungen zu unterstützen.

1.2 Zielsetzung und Forschungsfragen

Ziel dieser Masterarbeit ist die Entwicklung eines KI-gestützten Material-Mappings, das Materialbezeichnungen aus IFC-Modellen semantisch den passenden Einträgen der KBOB-Ökobilanzdaten zuordnet. Der Fokus des Prototyps liegt auf Brückenbauten und den zugehörigen Materialien.

Die Zuordnung erfolgt KI-gestützt, aber nicht vollautomatisch. Der Prototyp schlägt die wahrscheinlichsten Zuordnungen vor, die finale Auswahl trifft die nutzende Person im Sinne eines *Human-in-the-loop*-Ansatzes. Kontextmerkmale wie IFC-Entität, Bauteiltyp, Materialbezeichnung, Durchmesser, Herstellungsart und Festigkeitsklasse werden für die Zuordnung berücksichtigt. Für die Interaktion wird eine einfache Benutzeroberfläche erstellt. Die Einbettung in den Schweizer Normenrahmen (SIA, KBOB, SNBS) wird in den Grundlagen hergeleitet und im Prototyp umgesetzt.

Die Arbeit fokussiert auf die folgende Hauptfragestellung:

Wie kann ein KI-gestütztes Verfahren prototypisch umgesetzt werden, das Materialnamen aus IFC-Modellen semantisch den passenden Ökobilanzdaten zuordnet und dabei kontextabhängige Informationen berücksichtigt, um eine praxisnahe Anwendung im Schweizer Infrastrukturbau zu ermöglichen?

Aus der Hauptfragestellung ergeben sich folgende Teilfragen, gegliedert in Nachhaltigkeit (1–3) und Künstliche Intelligenz (4–6):

1. *Welche Besonderheiten ergeben sich bei der Anwendung im Schweizer Markt bezüglich Nachhaltigkeitsbewertung?*
2. *Wie können Schweizer Ökobilanzdaten, Normen und Richtlinien in die Ökobilanzierung integriert werden?*
3. *Welche Anforderungen stellen SNBS und KBOB an die Granularität und Genauigkeit der Materialklassifikation?*
4. *Welche Kontextinformationen (z. B. Bewehrungsgehalt, Beziehungen zu anderen Bauteilen, Lage im Modell) aus dem IFC-Modell können für das semantische Mapping verwendet werden um die Ergebnisse zu verbessern?*
5. *Welche NLP-Methoden bzw. -Modelle eignen sich zur automatisierten Verknüpfung von Materialbezeichnungen aus IFC-Modellen mit EPD- bzw. Ökobilanzdaten, insbesondere unter Berücksichtigung von Datenlücken und sprachlicher Uneinheitlichkeit?*
6. *Welche Mindestanforderungen muss ein IFC-Modell für die semantische Anreicherung und der daraus folgenden Ökobilanzierung erfüllen?*

Die Arbeit beschränkt sich bei der Nachhaltigkeitsbewertung auf die Umweltdimension, da nur diese direkt mit Ökobilanzdaten quantifiziert werden kann, und fokussiert auf Brückenbauwerke.

1.3 Aufbau der Arbeit

Kapitel 2 vermittelt die fachlichen Grundlagen zu Normen, KBOB-Ökobilanzdaten, zentralen Aspekten des IFC-Standards sowie relevanten Konzepten aus KI und NLP. Kapitel 3 analysiert bestehende Ansätze zur BIM-basierten Ökobilanzierung und zum semantischen Mapping und leitet daraus relevante Forschungslücken ab. Kapitel 4 beschreibt das methodische Vorgehen von der Literaturrecherche über die Entwicklung des Prototyps bis zur Evaluation. Kapitel 5 erläutert das Systemdesign, die Matching-Strategie sowie den Ansatz für das Fine-Tuning. Kapitel 6 dokumentiert die technische Umsetzung mit Softwarearchitektur, IFC-Extraktion, KBOB-Integration und Trainingspipeline. Kapitel 7 präsentiert die Resultate der Evaluation anhand kontrollierter Testdaten, fehlerhafter Eingaben und realer Infrastrukturprojekte. Kapitel 8 ordnet die Ergebnisse ein, diskutiert Limitationen und formuliert Empfehlungen für die Weiterentwicklung.

2 Grundlagen

Dieses Kapitel führt die fachlichen Grundlagen ein, auf denen die vorliegende Arbeit aufbaut. Zunächst werden die massgebenden Normen, Standards und Datenquellen für die Nachhaltigkeitsbewertung im Schweizer Infrastrukturbau vorgestellt. Anschliessend erläutert das Kapitel die Methoden der Ökobilanzierung und deren Anwendung im Infrastrukturbereich. Es folgt eine Einführung in den IFC-Standard als Grundlage für den Datenaustausch digitaler Bauwerksmodelle. Abschliessend werden die relevanten Konzepte aus dem Bereich der Künstliche Intelligenz (KI) und des Natural Language Processing (NLP) beschrieben, die für das semantische Matching von Materialbezeichnungen benötigt werden.

2.1 Normen, Standards und Daten

Die folgenden Abschnitte beschreiben den normativen Rahmen, den Bewertungsstandard SNBS-Infrastruktur sowie die Ökobilanzdaten der KBOB als zentrale Datenquelle für die umweltbezogene Bewertung.

2.1.1 Normen

Die Nachhaltigkeitsbewertung von Bauwerken in der Schweiz basiert auf einem normativen Rahmen, der die Erfassung, Bewertung und Kommunikation von Umwelt-, Sozial- und Wirtschaftsaspekten regelt (siehe Tab. 2.1). Zentrale Grundlagen bilden die SN EN 15643 (2021) als Rahmennorm sowie die darauf aufbauende SN EN 17472 (2022), welche die Nachhaltigkeitsbewertung speziell für Ingenieurbauwerke festlegt.

Nachhaltigkeitsbewertung			
	Umwelt	Sozial	Wirtschaft
Rahmen	EN 15643 Allgemeine Bedingungen zur Bewertung von Gebäuden und Ingenieurbauwerken		
Bauwerk	EN 17472 Nachhaltigkeitsbewertung von Ingenieurbauwerken		
Produkt	EN 15804 Umweltproduktdeklaration Grundregeln, EN 15941 & 22057 Datenqualität und -vorlage		

Tab. 2.1: Überblick der massgebenden Normen in Bezug auf Nachhaltigkeit von Bauwerken, eigene Grafik in Anlehnung an SN EN 17472 (2022, 8)

Die SN EN 15643 (2021) definiert die grundlegenden Prinzipien, Ziele und Anforderungen an eine lebenszyklusbasierte Nachhaltigkeitsbewertung von Bauwerken. Sie fordert eine transparente, nachvollziehbare und vergleichbare Bewertung der drei Nachhaltigkeitsdimensionen: Umwelt, Soziales

und Wirtschaft. Die technische und funktionale Qualität ist nicht Bestandteil der Bewertung, wird jedoch über das funktionale Äquivalent¹ berücksichtigt und bildet die Grundlage für den Variantenvergleich.

Die SN EN 17472 (2022) konkretisiert diesen Ansatz für Ingenieurbauwerke (siehe Tab. 2.2). Bewertet wird der gesamte Lebenszyklus, wobei die Ergebnisse den Modulen A0–5, B1–8, C1–4 sowie D zugeordnet werden. Damit werden Planung und Errichtung, Nutzung, Rückbau und Entsorgung sowie potenzielle Wirkungen ausserhalb der Systemgrenze abgebildet. Modul D wird separat ausgewiesen, da es die «potenziellen Ressourcen für eine künftige Nutzung einschließlich wiederverwendeter Produkte, recycelter Werkstoffe und Energierückgewinnung» beschreibt (ebd., 39). Ein zentrales Element der Bewertung ist die Systemgrenze. Beide Normen folgen dem Modularitätsprinzip, bei dem Prozesse dem jeweiligen Lebenszyklusmodul zugeordnet werden. Die SN EN 17472 (2022) beschreibt diese Abgrenzung für Ingenieurbauwerke detailliert, insbesondere für die Errichtungs-, Nutzungs- und Rückbauphase.

Informationen zur Bewertung von Ingenieurbauwerken					
Informationen innerhalb des Lebenszyklus					Informationen ausserhalb des Lebenszyklus
Planung A0	Herstellung A1-3	Errichtung A4-5	Nutzung B1-8	Ende C1-4	Vorteile und Belastungen D1-2
Land und zugehörige Kosten/Planungsaufwand	- Materialgewinnung - Transport - Herstellung	- Transport (A4) - Errichtung - Erschliessung - Rückbau Bestand - Installationen	- Nutzung - Instandhaltung - Instandsetzung - Ersatz - Modernisierung - Energie für Betrieb - Wasser für Betrieb - Nutzeraktivität	- Rückbau - Transport - Abfallverwertung - Entsorgung	- Wiederverwendung - Recycling - Energierückgewinnung - sonstige Rückgewinnung - an Dritte gelieferte Energie/Trinkwasser

Tab. 2.2: Phasen der Nachhaltigkeitsbewertung, eigene Grafik in Anlehnung an SN EN 17472 (2022, 32)

Die Normen verlangen konsistente und überprüfbare Daten. Für die umweltbezogene Bewertung sollen möglichst produktspezifische Daten nach SN EN 15804 (2022), insbesondere Umweltproduktdeklarationen (EPD), verwendet werden. Sind diese nicht verfügbar, dürfen unter definierten Voraussetzungen auch generische Datensätze eingesetzt werden. Die Datenqualität ist hinsichtlich Aktualität sowie geografischer und technologischer Repräsentativität zu prüfen. Ergänzend definieren die SN EN 15941 (2025) und die SN EN ISO 22057 (2022) Anforderungen an Datenqualität, Datenauswahl und Datenstruktur.

Die Bewertung erfolgt anhand festgelegter Indikatoren. Die SN EN 15643 (2021) definiert die relevanten Themenfelder der drei Nachhaltigkeitsdimensionen, während die SN EN 17472 (2022) diese für Ingenieurbauwerke mit spezifischen Indikatoren und Berechnungsverfahren konkretisiert. Die umweltbezogene Bewertung umfasst unter anderem Ressourcenverbrauch, Emissionen, Wasser- und Flächennutzung sowie Biodiversität. Die soziale Bewertung behandelt Aspekte wie Zugänglichkeit, Konnektivität, Gesundheit, Sicherheit und Auswirkungen auf die Umgebung. Die ökonomische

¹Das funktionale Äquivalent beschreibt die wesentlichen Anforderungen, die ein Bauwerk erfüllen muss, um mit anderen Varianten verglichen werden zu können. Es umfasst die vorgesehene Nutzung, die wesentlichen funktionalen Anforderungen, die massgebenden technischen Anforderungen, die Art und Primärfunktion des Bauwerks sowie den Betrachtungszeitraum. Damit bildet es die Grundlage für transparente und aussagekräftige Vergleiche unterschiedlicher Lösungen (SN EN 15643, 2021, 29).

Bewertung basiert insbesondere auf Lebenszykluskosten und gegebenenfalls Erträgen über den gesamten Lebenszyklus.

Die Ergebnisse sind in einem systematischen Bericht darzustellen (SN EN 17472, 2022, 101–102). Dieser enthält den Zweck und Anwendungsbereich der Bewertung, die Beschreibung des Bauwerks, das funktionale Äquivalent, die Systemgrenzen, die verwendeten Daten und Annahmen sowie die Ergebnisse für die betrachteten Module und Indikatoren. Nicht ausgewiesene Indikatoren/Module sind als «nicht deklariert» zu kennzeichnen und zu begründen.

Zusammengefasst verlangen die Normen eine lebenszyklusbasierte, methodisch konsistente und transparent dokumentierte Nachhaltigkeitsbewertung. Zentrale Elemente sind eine klare Systemabgrenzung, ein definiertes funktionales Äquivalent, belastbare Daten sowie geeignete Indikatoren, um vergleichbare und nachvollziehbare Bewertungen von Ingenieurbauwerken zu ermöglichen.

2.1.2 SNBS und NNBS

Der Kriterienbeschrieb SNBS-Infrastruktur (SNBS und NNBS, 2026, 4) beschreibt einen vom «Netzwerk Nachhaltiges Bauen Schweiz» (NNBS) angestossenen und entwickelten Standard für Infrastrukturbauten in der Schweiz. Der Kriterienbeschrieb sei «der erste umfassende Standard für die Nachhaltigkeitsbeurteilung von Schweizer Infrastrukturbauwerken und -projekten» (ebd.). Er baut auf Normen auf und «versteht sich als Instrument zur Beurteilungs-, Entscheidungs- und Planungshilfe für einen nachhaltigen Infrastrukturbereich» (ebd.).

Der SNBS-Kriterienbeschrieb ist für verschiedene Infrastrukturbereiche und -typen sowie für unterschiedliche Projektarten und -phasen anwendbar. Er kann unter anderem als Checkliste, als Kriterienkatalog zur Analyse und Priorisierung relevanter Themen, zum Vergleich von Projektvarianten sowie als Referenz für die Berücksichtigung nachhaltiger Entwicklung in Ausschreibungen verwendet werden (ebd., 5).

Rahmenbedingungen								
Gesellschaft			Wirtschaft			Umwelt		
Raumentwicklung und Siedlung	Gemeinschaft	Gesundheit und Sicherheit	Betriebswirtschaft	Volkswirtschaft	Finanzierung	Rohstoffe, Energie und Bodennutzung	Natur und Umwelt	Gefahrenprävention

Tab. 2.3: Bereiche und Themen der Nachhaltigkeitsbewertung gemäss SNBS, eigene Grafik in Anlehnung an SNBS und NNBS (2026, 10)

Der Standard gliedert sich in die vier Bereiche Rahmenbedingungen, Gesellschaft, Wirtschaft und Umwelt (siehe Tab. 2.3). Diese Bereiche umfassen insgesamt 10 Themen, 29 Kriterien und 75 Indikatoren (ebd., 5, 9). Die Bewertung erfolgt auf Ebene der Indikatoren und stützt sich «vorwiegend auf eine qualitative Beurteilung der Indikatoren» (ebd., 5).

Zentral ist, wie auch in den Normen (siehe Abschnitt 2.1.1), die Lebenszyklusperspektive. Der SNBS-Infrastruktur «betrachtet den ganzen Lebenszyklus von Infrastrukturbauten» (ebd., 6) und «gelangt in

allen SIA-Phasen zur Anwendung» (SNBS und NNBS, 2026, 22). Damit unterstützt er Auftraggebende und Planende dabei, nachhaltigkeitsrelevante Themen frühzeitig zu berücksichtigen, Zielkonflikte und Synergien zu identifizieren und geeignete Massnahmen in die Projektentwicklung zu integrieren (ebd., 22, 31).

Der SNBS-Kriterienbeschrieb dient damit als gemeinsame Grundlage für die strukturierte Auseinandersetzung mit Nachhaltigkeit in Infrastrukturprojekten. Er ist nicht primär auf Zertifizierung ausgerichtet, denn «der SNBS-Infrastruktur vergibt bewusst keine Zertifizierungen oder Labels» (ebd., 14). Im Vordergrund stehen vielmehr die projektspezifische Beurteilung und die Optimierung relevanter Nachhaltigkeitsthemen (ebd.).

Für die vorliegende Arbeit sind die Umweltindikatoren U 1.5.1 (Materialbedarf) und U 2.1.1 (Treibhausgasemissionen) relevant (ebd., 113, 117). Der SNBS definiert keine quantitative Genauigkeitsanforderung an die Materialklassifikation. Die geforderte Granularität ergibt sich implizit aus dem KBOB-Katalog: Eine korrekte Zuordnung ist Voraussetzung für belastbare Umweltindikatoren. Fehlzusammenordnungen zwischen ähnlichen Einträgen (z. B. Tiefbaubeton statt Bohrpfehlbeton) führen zu fehlerhaften UBP- und Treibhausgaswerten.

2.1.3 Ökobilanzdaten des KBOB

Die Ökobilanzdaten der Koordinationskonferenz der Bau- und Liegenschaftsorgane der öffentlichen Bauherren (KBOB) bilden eine zentrale Datengrundlage für die umweltbezogene Bewertung von Bauwerken in der Schweiz. Das BAFU bezeichnet Datenbanken mit «aktuellen, transparenten und nachvollziehbaren Hintergrunddaten» wie die KBOB-Ökobilanzdaten als «wesentliche Grundlage» für belastbare Ökobilanzen (BAFU, 2021, 7).

Die Datensätze umfassen wesentliche Umweltindikatoren (siehe Tab. 2.4 oder Anhang A.1) wie die Gesamtbewertung in UBP'21, Treibhausgasemissionen, den biogenen Kohlenstoffgehalt sowie die erneuerbare und nicht erneuerbare (graue Energie) Primärenergie (KBOB und ecobau, 2026, 1).

BAUMATERIALIEN	Rohdichte/ Flächen- masse	Bezug	UBP'21			Primärenergie										Treibhausgasemissionen			
			Total UBP	Herstellung UBP	Entsorgung UBP	Total kWh oil-eq	erneuerbar					nicht erneuerbar (Graue Energie)					Total kg CO ₂ -eq	Herstellung kg CO ₂ -eq	Entsorgung kg CO ₂ -eq
							Herstellung total kWh oil-eq	Herstellung energetisch genutzt kWh oil-eq	Herstellung stofflich genutzt kWh oil-eq	Entsorgung kWh oil-eq	Total kWh oil-eq	Herstellung total kWh oil-eq	Herstellung energetisch genutzt kWh oil-eq	Herstellung stofflich genutzt kWh oil-eq	Entsorgung kWh oil-eq				
Vorbereitungsarbeiten	-																		
Baugrubensicherung, Bohrpfehlwand, gesprisst	4884	m ²	1'300'000	1'140'000	167'000	132	123	123	0	9.21	2'120	1870	1870	0	248	861	801	60.5	
Baugrubensicherung, Spundwand, auskragend	153	m ²	287'000	286'000	1'690	45.7	45.6	45.6	0	0.089	732	728	728	0	4.16	169	168	1.06	
Beton	kg/m ³																		
Magerbeton (ohne Bewehrung)	2'150	kg	104	68.8	35.4	0.010	0.008	0.008	0	0.002	0.128	0.077	0.077	0	0.052	0.063	0.050	0.013	
Tiefbaubeton (ohne Bewehrung)	2'350	kg	163	128	35.1	0.016	0.014	0.014	0	0.002	0.189	0.138	0.138	0	0.051	0.109	0.096	0.013	
Bohrpfahlbeton (ohne Bewehrung)	2'325	kg	174	139	35.1	0.017	0.015	0.015	0	0.002	0.197	0.146	0.146	0	0.051	0.119	0.107	0.013	
Betonfertigteil, hochfester Beton, ab Werk	2'770	kg	440	409	31.0	0.076	0.074	0.074	0	0.002	0.697	0.652	0.652	0	0.045	0.276	0.265	0.011	
Betonfertigteil, Normalbeton, ab Werk	2'500	kg	282	248	33.9	0.095	0.093	0.093	0	0.002	0.432	0.383	0.383	0	0.050	0.177	0.165	0.012	
Mauersteine	kg/m ³																		
Backstein	900	kg	371	328	42.5	0.081	0.078	0.078	0	0.002	0.787	0.734	0.734	0	0.053	0.256	0.254	0.013	
Kalksandstein	1'400	kg	258	216	42.5	0.039	0.037	0.037	0	0.002	0.389	0.337	0.337	0	0.053	0.162	0.149	0.013	
Zementstein	1'700	kg	214	172	42.5	0.026	0.024	0.024	0	0.002	0.225	0.173	0.173	0	0.053	0.126	0.114	0.013	
Andere Massivbaustoffe	kg/m ³																		
Kies gebrochen ¹	1'400 - 1'500	kg	64.8	24.6	40.2	0.008	0.006	0.006	0	0.002	0.085	0.032	0.032	0	0.053	0.018	0.005	0.013	
Rundkies ¹	1'600 - 1'700	kg	59.4	19.2	40.2	0.004	0.002	0.002	0	0.002	0.068	0.016	0.016	0	0.053	0.016	0.003	0.013	
Sand ¹	1'500	kg	59.4	19.2	40.2	0.004	0.002	0.002	0	0.002	0.068	0.016	0.016	0	0.053	0.016	0.003	0.013	

Tab. 2.4: Auszug der Ökobilanzdaten der KBOB, Screenshot KBOB und ecobau (2026).

Die UBP'21 basieren auf der Methode der ökologischen Knappheit (siehe Abschnitt 2.2.1). Für Ökobilanzen mit Bezug zur Schweiz hält das BAFU fest, die UBP-Methode mit den Schweizer Ökofaktoren sei eine Referenzmethode und «bei Studien mit Bezug zur Schweiz stets eine der Bewertungen mit der UBP-Methode» durchzuführen (BAFU, 2021, 39).

Im Zusammenhang mit Umweltproduktdeklarationen verweist das BAFU darauf, dass «für das Bauwesen seit April 2012 die europäische Norm EN 15804 die Vorgaben für EPDs von Baumaterialien, -produkten» beschreibt (ebd., 34). Damit besteht eine fachliche Anschlussfähigkeit der KBOB-Daten an die im Bauwesen etablierten Wirkungsindikatoren und Deklarationslogiken. Diese Aussage ist aber nicht gleichbedeutend mit einer vollständigen inhaltlichen Gleichsetzung von KBOB-Daten und produktspezifischen EPD.

Die KBOB-Ökobilanzdaten werden als Excel-Datei öffentlich bereitgestellt und dienen in der Schweiz als praktische Grundlage für die Berechnung und Bewertung von Umweltwirkungen im Baubereich (KBOB und ecobau, 2026, 1).

2.2 Ökobilanzierung in der Schweiz

Dieser Abschnitt erläutert zunächst die Methode der ökologischen Knappheit als Bewertungsmethode für Ökobilanzen mit Bezug zur Schweiz. Anschliessend wird die Anwendung der Ökobilanzierung im Infrastrukturbau dargestellt.

2.2.1 Methode der ökologischen Knappheit

Die Methode der ökologischen Knappheit, auch UBP-Methode genannt, ist ein in der Schweiz angewendetes Bewertungsverfahren zur Wirkungsabschätzung in Ökobilanzen. «Zentrale Grösse der Methode sind die Ökofaktoren, welche die Umweltbeeinträchtigung in Umweltbelastungspunkten UBP pro Mengeneinheit ausdrücken» (BAFU, 2021, 6, 49).

Die Formel für die Anzahl UBP von 1g fossilem Methan (CH_4) ist (ebd., 43–44):

$$\text{UBP}_{\text{CH}_4} = K_{\text{CH}_4} \cdot \frac{1 \text{ UBP}}{F_n} \cdot \left(\frac{F}{F_k} \right)^2 \cdot c$$

$$\text{UBP}_{\text{CH}_4} = 30 \cdot \frac{1 \text{ UBP}}{61\,826\,000 \text{ t CO}_2\text{-eq/a}} \cdot \left(\frac{61\,826\,000 \text{ t CO}_2\text{-eq/a}}{7\,829\,000 \text{ t CO}_2\text{-eq/a}} \right)^2 \cdot 10^{12}/a \approx 30 \text{ UBP/g}$$

wobei

K = Charakterisierungsfaktor einer Emission oder Ressource

F_n = Normierungsfluss: aktuelle jährliche Menge (Emission oder Verbrauch), bezogen auf die Schweiz

F = Aktueller Fluss: aktuelle jährliche Menge (Emission oder Verbrauch), bezogen auf das Referenzgebiet

F_k = Kritischer Fluss: gesetzlicher Grenzwert, bezogen auf das Referenzgebiet

c = Konstante ($10^{12}/a$): dient dazu, einfach darstellbare Zahlengrössen zu erhalten

UBP = Umweltbelastungspunkt: Einheit der bewerteten Umweltwirkung

Die Wirkungsabschätzung erfolgt nach dem Prinzip «Verhältnis zur tolerierten Zielmenge» (BAFU, 2021, 8). Entsprechend «gewichtet die Methode der ökologischen Knappheit die Umwelteinwirkungen, das heisst die Schadstoffemissionen und die Ressourcennutzungen, mit sogenannten «Ökofaktoren»» (ebd., 56). Die Herleitung der Ökofaktoren basiert auf Charakterisierung, Normierung und Gewichtung. Dabei wird die aktuelle Belastung einer Emission oder Ressourcennutzung mit der tolerierten Zielmenge ins Verhältnis gesetzt (ebd., 43, 56–57).

Die Methode erfasst ein breites Spektrum an Umweltwirkungen, darunter Treibhausgase, Luft- und Wasserschadstoffe, Schwermetalle, Landnutzung, Süsswasserverbrauch, Abfälle und Verkehrslärm (ebd., 15, 39, 47–48). Durch die Aggregation der Einzelbeiträge in UBP lassen sich Produkte und Prozesse vergleichend bewerten (ebd., 14–15, 36).

Bei Beton basiert die Bewertung auf den Sachbilanzen der eingesetzten Betonbestandteile und Prozesse. Massgebend sind insbesondere Zement und Gesteinskörnungen, ergänzt um Energie-, Transport- und Entsorgungsaufwände. Im Hintergrundbericht «Ökobilanz von ausgewählten Betonsorten» von Tschümperlin et al. (2020) können dazu die Details entnommen werden.

Für die Schweiz ist die Methode besonders relevant. Gemäss Bundesamt für Umwelt (BAFU) «richtet sich die Methode nach den in der schweizerischen Gesetzgebung sowie in internationalen Vereinbarungen verankerten Umweltqualitätszielen und Grenzwerten» (BAFU, 2021, 15). Deshalb bezeichnet das BAFU die UBP-Methode mit den Schweizer Ökofaktoren als Referenzmethode für Ökobilanzen mit Bezug zur Schweiz (ebd., 39).

2.2.2 Ökobilanzierung im Infrastrukturbau

Im Infrastrukturbau dient die Ökobilanzierung dazu, Umweltwirkungen von Varianten, Bauweisen und Materialisierungen quantifizierbar zu machen und damit Planungsentscheide zu unterstützen. Besonders relevant ist ihr Einsatz in frühen Projektphasen, in denen der Gestaltungsspielraum noch gross ist und Variantenvergleiche die Projektoptimierung wirksam beeinflussen können (Buchheister et al., 2025, 10, 73). «Infrastrukturprojekte mit großen Ausdehnungen und hohem Materialbedarf produzieren besonders viele Emissionen [...], Anwendungen zur Ökobilanzierung kommen dabei aber kaum zum Einsatz» (Hermann, 2023, 2).

Die Anwendung im Infrastrukturbau ist mit erhöhten Anforderungen verbunden. Infrastrukturbauten unterscheiden sich stark hinsichtlich Funktion, Dimension, Nutzungsdauer und Erhaltungszyklen. Daraus folgt, dass Untersuchungsrahmen, Bezugsgrösse und Datentiefe jeweils projektspezifisch festgelegt werden müssen, wenn belastbare und vergleichbare Ergebnisse erreicht werden sollen (Buchheister et al., 2025, 25–27).

Eine zentrale Herausforderung liegt in der Datengrundlage. Während im Hochbau bereits zahlreiche standardisierte Datensätze vorliegen, bestehen im Tief- und Infrastrukturbau weiterhin Lücken bei Materialien, Bauverfahren, Maschinen und generischen Standardelementen. Gerade in frühen Projektphasen sind deshalb häufig Abschätzungen oder vereinfachte Datensätze notwendig. Die Aussagekraft der Ergebnisse hängt damit in besonderem Mass von der Verfügbarkeit, Qualität und

Transparenz der verwendeten Daten ab (Stettler und Vorhoff, 2023, 11–18). Aktuelle Auswertungen zeigen zudem, dass insbesondere Beton, Asphalt, Stahl und Natursteine einen wesentlichen Anteil an den Umweltwirkungen von Infrastrukturprojekten verursachen (Bleiker et al., 2025, 3–4, 16).

2.3 IFC-Standard

Der IFC-Standard ist «ein umfassendes und standardisiertes Datenformat für den herstellernerneutlichen Austausch von digitalen Bauwerksmodellen» (Borrmann et al., 2021, 95). Er beschreibt Bauwerke nicht nur über ihre Form, sondern auch über Klassen, Attribute und Beziehungen. Dadurch ist zum Beispiel erkennbar, ob ein Element eine Stütze oder ein Widerlager ist und wie dieses Element im Gesamtmodell eingeordnet ist (ebd., 100–106). Wichtig ist dabei die Trennung zwischen der fachlichen Beschreibung eines Objekts und seiner Geometrie. Ein Brückenelement ist im IFC-Modell nicht nur ein Körper, sondern ein Objekt mit einer klaren Bedeutung und definierten Informationen (ebd., 114–124).

Für die Struktur eines IFC-Modells sind Aggregationsbeziehungen wichtig. Sie bildet einen hierarchischen Baum räumlicher Elemente, der dem Projekt zugeordnet ist (siehe Abb. 2.1).

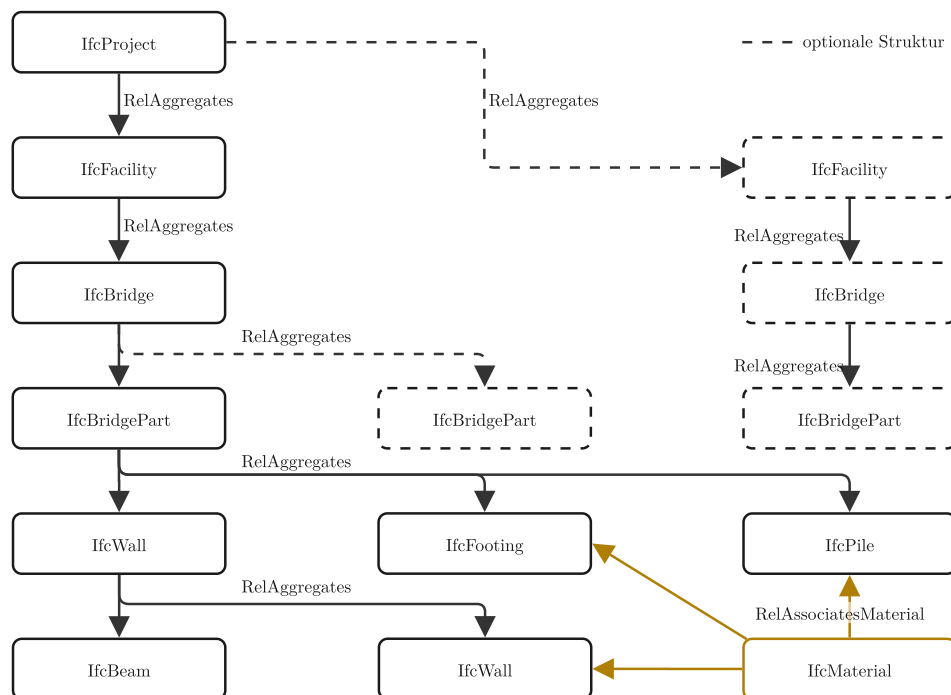


Abb. 2.1: Mögliche Struktur und Beziehungen von IFC-Elementen, eigene Grafik nach buildingSMART International (2023).

Bei einem Brückenmodell verläuft diese Hierarchie typischerweise von `IfcProject` über `IfcBridge` weiter zu `IfcBridgePart`, wobei die Zuordnung über `IfcRelAggregates` erfolgt.

Von der räumlichen Gliederung ist die Zerlegung eines Bauteils in Teilobjekte zu unterscheiden. Wird ein Element, z. B. eine (Rühl-)Wand, über `IfcRelAggregates` in eigene Teile wie Ausfachung (Wall) und Stahlprofile (Beam) zerlegt, beschreibt dies den inneren Aufbau und nicht die räumliche

Einordnung. In diesem Fall sieht das Konzept **Element Decomposition** standardmässig vor, dass das zusammengesetzte Element die gemeinsame Platzierung vorgibt, während die Materialzuordnung bei den Teilobjekten liegt und über **IfcRelAssociatesMaterial** erfolgt (buildingSMART International, 2023).

Insgesamt bietet der IFC-Standard eine strukturierte Grundlage, um Informationen aus IFC-Elementen auszulesen und mit KBOB-Ökobilanzdaten zu verknüpfen.

2.4 Künstliche Intelligenz und Natural Language Processing

Dieser Abschnitt führt die für diese Arbeit relevanten Konzepte aus dem Bereich der künstlichen Intelligenz (KI) ein. Er erläutert zunächst die Grundlagen des Natural Language Processing (NLP) und vortrainierter Sprachmodelle. Danach wird die semantische Textähnlichkeit als methodische Grundlage für den Vergleich von Materialbezeichnungen beschrieben.

2.4.1 Künstliche Intelligenz

KI bezeichnet im Allgemeinen Verfahren zur Entwicklung von Systemen, die Informationen verarbeiten und auf dieser Grundlage zweckmässig handeln oder Entscheidungen unterstützen. Die Verarbeitung natürlicher Sprache ist ein Teilgebiet der KI und wird als NLP bezeichnet. Moderne NLP-Verfahren nutzen sehr häufig maschinelles Lernen und Deep Learning, insbesondere Transformer-Modelle. NLP ist aber nicht identisch mit Deep Learning, da es auch regelbasierte, statistische und linguistische Ansätze umfasst (Jurafsky und Martin, 2026, 17, 38 ff., 63, 120, 173).

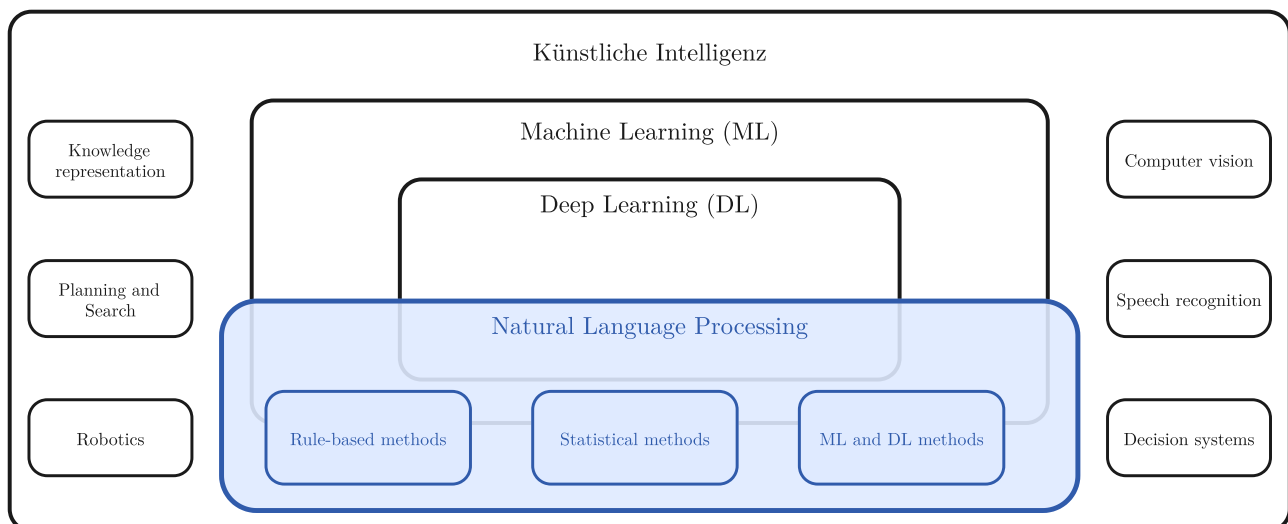


Abb. 2.2: Schematische Einordnung von NLP in die Teilgebiete von KI, eigene Grafik in Anlehnung an Jurafsky und Martin (2026).

Für die vorliegende Arbeit sind vor allem NLP-Aufgaben relevant wie die Zerlegung von Texteingaben in Tokens² (Tokenization³, in Abb. 2.3 werden die Wörter in Tokens aufgeteilt, welche farbig markiert sind.), die Bildung semantischer Repräsentationen von Wörtern und kurzen Texten sowie die Berechnung von Ähnlichkeiten zwischen unterschiedlich formulierten Bezeichnungen. Damit bildet NLP die Grundlage für Anwendungen, in denen technische Benennungen trotz abweichender Wortwahl inhaltlich einander zugeordnet werden können (Jurafsky und Martin, 2026, 4 ff.).

IfcWall Standard Concrete C30/37 --> IfcWall Standard Concrete C30 / 37

Abb. 2.3: Beispielhafte Tokenisierung eines Queries mit dem Modell BGE-M3 (J. Chen et al., 2025), eigene Darstellung.

Die Entwicklung des NLP wurde in den letzten Jahren stark durch vortrainierte Sprachmodelle geprägt. Ein wichtiger Meilenstein ist BERT, ein bidirektionales Transformer-Modell, das auf grossen Textmengen vortrainiert wird und anschliessend für unterschiedliche Aufgaben angepasst werden kann. Der zentrale Unterschied zu früheren Ansätzen liegt darin, dass BERT den linken und rechten Kontext eines Tokens gemeinsam berücksichtigt und dadurch sprachliche Zusammenhänge deutlich besser erfassen kann (Devlin et al., 2019, 1).

Für viele Anwendungen genügt es nicht, einzelne Wörter oder Tokens zu betrachten. Wenn ganze Sätze oder kurze Texte semantisch verglichen werden sollen, werden Repräsentationen benötigt, die den Inhalt eines gesamten Textes erfassen. Auf BERT aufbauende Modelle wie Sentence-BERT haben gezeigt, dass sich solche semantisch aussagekräftige Satz-Embeddings⁴ (Reimers und Gurevych, 2019, 1) effizient für Ähnlichkeitsvergleiche nutzen lassen. Spätere Embedding-Modelle wurden für zusätzliche Anforderungen wie mehrsprachige Anwendungen und Retrieval-Aufgaben weiterentwickelt. Zugleich zeigen neuere Arbeiten, dass auch die Verarbeitung längerer Texte ein wichtiges Entwicklungsziel solcher Modelle ist (Ajallouda et al., 2025, 539 ff.).

Insgesamt zeigt sich, dass moderne KI im Sprachbereich weit über die reine Textgenerierung hinausgeht. Sie ermöglicht es, sprachliche Informationen strukturiert zu erfassen, semantisch zu vergleichen und für weitere Anwendungen nutzbar zu machen. Solche Verfahren schaffen die Grundlage dafür, textuelle Informationen systematisch zu analysieren und in maschinenverarbeitbare Repräsentationen (Vektoren) zu überführen (Jurafsky und Martin, 2026, 96 ff.).

²Tokens können Wörter, Teile von Wörtern oder einzelne Zeichen sein. Zur Generierung von Tokens gibt es verschiedene Algorithmen wie das häufig verwendete byte-pair encoding (BPE) (Jurafsky und Martin, 2026, 13–17), WordPiece (BertTokenizer) (Hugging Face, 2025) oder SentencePiece (Unigram LM) (Kudo et al., 2026).

³«Tokenization, the first stage of natural language processing, is the process of segmenting the running input text into tokens» (Jurafsky und Martin, 2026, 13).

⁴Mit semantisch aussagekräftig ist gemäss Reimers und Gurevych (2019, 1) gemeint, dass semantisch ähnliche Sätze nahe sind im Vektorraum.

2.4.2 Semantic Textual Similarity und Embedding-Modelle

Ein zentrales Konzept im NLP ist die Semantic Textual Similarity (STS), also die Bestimmung, wie ähnlich sich zwei Texte in ihrer Bedeutung sind. STS erfasst dabei nicht nur identische Begriffe, sondern vergleicht den inhaltlichen Zusammenhang unterschiedlich formulierter Ausdrücke (Cer et al., 2017, 1–2).

Methodisch wird STS über semantische Textrepräsentationen umgesetzt. Dazu werden kurze Texte oder Bezeichnungen in dichte Vektor-Darstellungen (Embeddings) überführt, deren räumliche Nähe im Vektorraum die inhaltliche Ähnlichkeit widerspiegelt. Der Vergleich erfolgt häufig mittels Kosinus-Ähnlichkeit, da sie eine effiziente Bewertung semantischer Nähe erlaubt (Reimers und Gurevych, 2019, 1).

Für die vorliegende Arbeit ist STS relevant, da Bezeichnungen aus IFC-Modellen und Einträge aus Ökobilanzdaten fachlich übereinstimmen können, obwohl sie unterschiedlich formuliert sind. Ziel ist nicht die exakte Zeichenkettenübereinstimmung, sondern die Erkennung semantisch naher Materialbezeichnungen. So kann etwa «*Concrete Strength*» mit «*Strength of Concrete*» miteinander abgeglichen werden (Bharti et al., 2025, 2).

STS bildet damit die methodische Grundlage für den semantischen Vergleich technischer Benennungen.

3 Stand der Technik

Dieses Kapitel ordnet die vorliegende Arbeit in den aktuellen Forschungsstand ein. Zunächst wird das BIM-LCA-Matching als semantisches Problem beschrieben. Danach werden bestehende NLP- und KI-basierte Ansätze zur automatisierten Zuordnung vorgestellt und hinsichtlich ihrer Übertragbarkeit auf den Schweizer Infrastrukturbau eingeordnet. Abschliessend werden die identifizierten Forschungslücken zusammengefasst.

3.1 BIM-LCA-Integration als semantisches Matching-Problem

Ein wesentlicher Engpass in der BIM-basierten Ökobilanzierung liegt in der Zuordnung von BIM-Elementen zu passenden Ökobilanzdaten. Petrosa et al. (2025, 2) beschreiben diese Zuordnung als zeit- und arbeitsintensive Aufgabe, die bisher nur teilweise automatisiert ist.

BIM-Modelle und Ökobilanzdaten enthalten viele textuelle Bezeichnungen, Kategorien und Beschreibungen. Diese Informationen folgen jedoch selten einer einheitlichen Terminologie. Besonders in frühen Planungsphasen sind Modellinformationen oft kurz, allgemein oder projektspezifisch benannt (H. Chen et al., 2025, 2). Damit wird die BIM-LCA-Integration zu einem semantischen Matching-Problem: Modellinformationen müssen nicht nur gelesen, sondern inhaltlich mit externen Ökobilanzdaten abgeglichen werden.

3.2 Semantische Anreicherung von IFC-Modellen

Eine frühe Richtung der semantischen Modellanreicherung arbeitet regelbasiert. Belsky et al. (2013) schlagen dafür ein Regelwerk vor, mit dem fehlende semantische Konstrukte aus bestehenden IFC-Informationen abgeleitet und ergänzt werden. Solche Verfahren sind nachvollziehbar, setzen aber voraus, dass Regeln, Zielkonzepte und Modellierungsannahmen im Voraus bekannt sind.

Diese Voraussetzung ist bei realen IFC-Modellen nur begrenzt gegeben, da sich Modelle je nach Autorensoftware und Benennung stark unterscheiden. Neuere Arbeiten nutzen deshalb verstärkt NLP-Verfahren, Embedding-Modelle, Ontologien, Wissensgraphen und LLMs, um uneinheitlich beschriebene Informationen maschinell nutzbar zu machen.

3.3 NLP- und LLM-basierte Matching-Verfahren

Die folgenden Abschnitte stellen drei Richtungen vor: Embedding-basierte Verfahren mit strukturierten Wissensdatenbanken, LLM- und RAG¹-basierte Ansätze sowie hybride Methoden, die Textähnlichkeit mit zusätzlichen Materialeigenschaften kombinieren.

3.3.1 Semantic Model Healing und Embedding-basierte Verfahren

Forth et al. (2023, 1–4) beschreiben *Semantic Model Healing* als Prozess, mit dem unvollständige BIM-Informationen in frühen Planungsphasen für die Ökobilanzierung semantisch ergänzt werden. Dafür werden IFC-Informationen mit einer LCA-Wissensdatenbank abgeglichen, die Bauteile, Materialien und Umweltindikatoren enthält (ebd., 15–18).

Das Matching kombiniert Klassifikation und semantische Ähnlichkeit. IFC-Elemente werden zuerst über ihre Klassifikation gefiltert, dann werden Bezeichnungen über Vektorisierung und Kosinus-Ähnlichkeit mit den Datenbankeinträgen verglichen. Im Vergleich mehrerer Sprachmodelle liefert BERT die besten Ergebnisse (ebd., 19–22, 27–31).

Die Evaluation mit fünf realen Bürogebäudemodellen ergibt 78.1 % korrekt zugeordnete Elemente (ebd., 31–35). Fehlerursachen sind vor allem fehlende Klassifikationen und unzureichende Modellinformationen (ebd., 32–33).

3.3.2 Fine-Tuning, LLMs und RAG

Die Trefferqualität hängt stark davon ab, ob Sprachmodelle bauspezifische Begriffe richtig erfassen. Forth, Berggold et al. (2025, 2–5) untersuchten deshalb das domänenspezifische Fine-Tuning eines deutschen BERT-Modells für das Matching von BIM-Informationen zu Materialpass-Datensätzen. Ohne Fine-Tuning erreichte das Modell 44.2 % korrekte Zuordnungen, mit kombinierten Strategien stieg die Genauigkeit auf 80.3 % (ebd., 6–8).

Petrosa et al. (2025) entwickelten ein Tool, das verschiedene IFC-Informationen extrahiert und mit *GPT-4o mini* passenden ÖKOBAUDAT-Datensätzen zuordnet (ebd., 7). Der automatische Ansatz reduzierte die Bearbeitungszeit von etwa drei Stunden auf rund elf Minuten, erreichte aber nur 66 % korrekte Zuordnungen (ebd., 14).

H. Chen et al. (2025) schlagen ein RAG¹-basiertes Framework vor, das semantische Ähnlichkeit und Zeichenähnlichkeit kombiniert. ÖKOBAUDAT-Einträge werden mit SBERT eingebettet, über FAISS² gesucht und zusätzlich mit Levenshtein-Distanz bewertet. GPT-2 erzeugt Erklärungen zu den Treffern. In der Fallstudie schwankt die Genauigkeit je nach Bauteilkategorie zwischen 27.7 % und 78.1 % (ebd., 1).

¹RAG steht für *Retrieval-Augmented Generation*. Dabei wird ein generatives Sprachmodell mit einer vorgelagerten Suche kombiniert. Das Modell erzeugt seine Antwort nicht nur aus intern gelerntem Wissen, sondern nutzt zusätzlich abgerufene externe Informationen.

²FAISS steht für *Facebook AI Similarity Search*. Es ist eine Softwarebibliothek zur effizienten Ähnlichkeitssuche in grossen Mengen von Vektoren.

3.3.3 Hybride Ansätze zur Absicherung semantischer Matches

Forth, Kaltenegger et al. (2025, 4–6) kombinieren *Semantic Textual Similarity* mit ontologiebasierter Bewertung und Dichtewerten. Dazu werden IFC-Modelle und LCA-Daten in Wissensgraphen überführt. Die Ergebnisse zeigen, dass zusätzliche Materialeigenschaften wie Dichtewerte die Zuordnung verbessern können (ebd., 8–10). Wissensgraphen sind damit eine mögliche Erweiterung, aber keine Voraussetzung für ein textbasiertes Matching.

3.4 LCA-Tools

Für die BIM-basierte Ökobilanzierung existieren verschiedene Werkzeuge, die sich in Datenbasis, Anwendungsbereich und Automatisierungsgrad unterscheiden. Für den Schweizer Hochbau gibt es Plattformen mit KBOB-Anbindung (etwa IfcLCA, Vyzn, viride). Internationale Werkzeuge wie One Click LCA sind nicht auf Schweizer Infrastrukturdaten ausgerichtet. Im Infrastrukturbau deckt Eco2nstruct zwar Baustoffe und Bauprozesse ab, bietet aber keine BIM-Anbindung. ORIS unterstützt IFC-Auswertungen für Infrastrukturprojekte, integriert jedoch keine KBOB-Daten und keine Bewertung nach UBP. Eine detaillierte Beschreibung der einzelnen Tools findet sich im Anhang A.2.

3.5 Forschungslücken und Herausforderungen

Die ausgewerteten Arbeiten zeigen, dass KI- und NLP-Methoden die Zuordnung von BIM-Informationen zu Ökobilanzdaten unterstützen können. Besonders tragfähig erscheinen Ansätze, die textbasierte Ähnlichkeit mit weiteren Informationen wie Klassifikationen, Materialeigenschaften, Wissensdatenbanken oder Nutzerentscheidungen kombinieren. Reine Textähnlichkeit reicht oft nicht aus, weil Material- und Bauteilbezeichnungen in BIM- und IFC-Modellen häufig kurz, uneinheitlich, projektspezifisch oder unvollständig sind.

Für die vorliegende Arbeit ergeben sich drei Forschungslücken. Erstens wurden die meisten BIM-LCA-Ansätze an Hochbaumodellen und mit der ÖKOBAUDAT entwickelt. Die Übertragbarkeit auf Schweizer Infrastrukturprojekte und KBOB-Ökobilanzdaten ist nicht gegeben. Zweitens setzen die Ansätze auf unterschiedlichen Matching-Ebenen an, was die Vergleichbarkeit erschwert. Drittens ist offen, welche IFC-Informationen in Infrastrukturmodellen stabil und vollständig genug sind, um ein semantisches Matching zu tragen.

Für die Verknüpfung von IFC-Infrastrukturmodellen mit KBOB-Daten muss daher geprüft werden, welche Modellinformationen zuverlässig auswertbar sind und wie gut sich KBOB-Einträge über Bezeichnungen, Kategorien und weitere Strukturinformationen abgleichen lassen. Neben der Trefferquote ist auch die Nachvollziehbarkeit wichtig. Bei unsicheren Zuordnungen bleibt eine fachliche Prüfung

notwendig. Die von Hermann et al. (2024, 129) geforderte Automatisierungsfähigkeit in der BIM-basierten Ökobilanzierung bleibt eine zentrale Herausforderung. Für diese Arbeit steht die semantische Zuordnung von IFC-Informationen zu KBOB-Ökobilanzdaten im Vordergrund.

4 Methodik

Dieses Kapitel beschreibt das methodische Vorgehen. Entwickelt und geprüft wird ein Prototyp, der (Material-)Informationen aus IFC-Modellen passenden Einträgen der KBOB-Ökobilanzdaten zuordnet. Der Prototyp erzeugt Vorschläge mit Konfidenzwert, die endgültige Auswahl bleibt jedoch bei der nutzenden Person.

Das Vorgehen ist in fünf Phasen gegliedert: Recherche, Anforderungsanalyse, Konzeption, prototypische Umsetzung und Evaluation. Die Phasen bauen aufeinander auf, werden jedoch nicht als strikt linear verstanden. Erkenntnisse aus Tests und aus der laufenden Entwicklung können dazu führen, dass Anforderungen, Datenmodell oder Mapping-Logik angepasst werden (siehe Abb. 4.1).

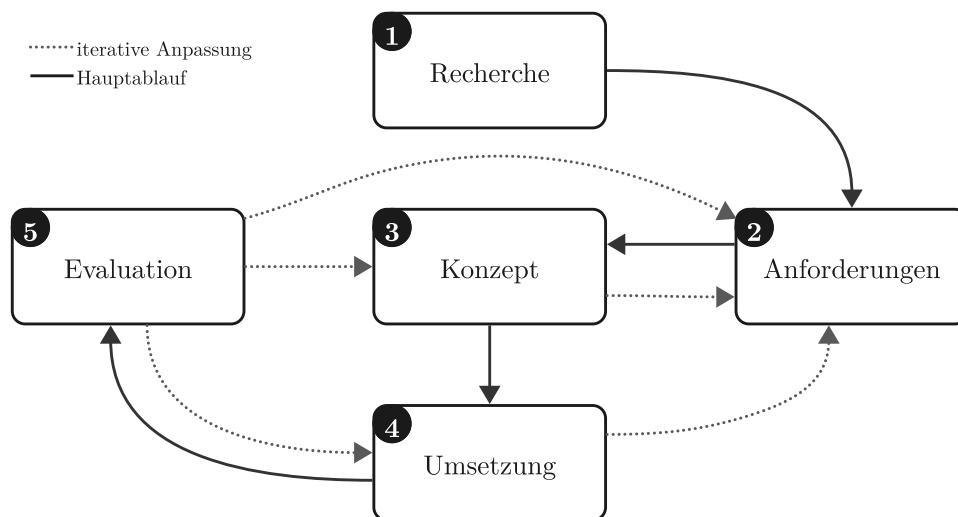


Abb. 4.1: Methodisches iteratives Vorgehen, eigene Darstellung.

4.1 Literatur- und Dokumentenanalyse

Zur Beantwortung der Forschungsfragen erfolgte die Recherche in *Web of Science*, *Scopus* und *Swisscovery* sowie auf spezifischen Webseiten wie denen des BAFU und des SIA. Dadurch konnten sowohl internationale Fachliteratur und Konferenzbeiträge als auch Schweizer Quellen wie Normen und Richtlinien berücksichtigt werden. Die Suchstrategie basierte auf einer Begriffstabelle mit zentralen Schlagwörtern und Synonymen aus den Forschungsfragen (siehe Abb. 4.2 oder Anhang A.3). Die Suchstrings wurden je nach Datenbank angepasst und mit Booleschen Operatoren kombiniert. Die Literatursuche erfolgte nach der Recherche mit den Suchstrings schrittweise: Zunächst wurden Titel, Abstracts und wenn das Thema relevant war Volltexte gesichtet und danach relevante Quellen zusätzlich über Rückwärts- und Vorwärtssuche ergänzt.

Inhaltlich lag der Fokus auf NLP-Verfahren im BIM-Kontext, BIM-basierter Ökobilanzierung, Schweizer Grundlagen wie SNBS und SIA sowie den technischen Grundlagen des IFC-Standards. Bevorzugt wurden Arbeiten ab 2015. Ausgeschlossen wurden Quellen ohne Bezug zum Thema sowie nicht-wissenschaftliche Quellen ohne nachvollziehbare Grundlage.

Forschungsfrage

Wie kann ein KI-gestütztes Verfahren prototypisch umgesetzt werden, das **Materialnamen** aus **IFC-Modellen** **semantisch** den passenden **Ökobilanzdaten** zuordnet und dabei kontextabhängige Informationen berücksichtigt, um eine **praxisnahe Anwendung im Schweizer Infrastrukturbau** zu ermöglichen?

Teilfragestellungen

Welche Besonderheiten ergeben sich bei der Anwendung im Schweizer Markt bezüglich **Nachhaltigkeitsbewertung**?

Wie können **Schweizer Ökobilanzdaten, Normen und Richtlinien** in die **Ökobilanzierung** integriert werden?

Welche Anforderungen stellen **SNBS und KBOB** an die Granularität und Genauigkeit der Materialklassifikation?

Wie und welche Kontextinformationen (z.B. Bewehrungsgehalt, Beziehungen zu anderen Bauteilen, Lage im Modell) aus dem **IFC-Modell** können für das **semantische Mapping** verwendet werden um die Ergebnisse zu verbessern?

Welche **NLP-Methoden** bzw. -Modelle eignen sich zur automatisierten Verknüpfung von Materialbezeichnungen aus **IFC-Modellen** mit **EPD- bzw. Ökobilanzdaten**, insbesondere unter Berücksichtigung von Datenlücken und sprachlicher Uneinheitlichkeit?

Welche Mindestanforderungen muss ein **IFC-Modell** für die **semantische** Anreicherung und der daraus folgenden **Ökobilanzierung** erfüllen?

Schlüsselbegriff	Synonyme/Varianten
Materialnamen	Material name, Materialbezeichnung, material classification, material mapping, material identification
IFC-Modell	IFC, Building Information Modeling, BIM, IFC file, IFC schema
Ökobilanzdaten	Life Cycle Assessment, LCA, Environmental Product Declaration, EPD, Ökobilanz, KBOB
semantisches Mapping	semantic mapping, semantic matching, semantic linking, semantische Zuordnung
NLP-Methoden	Natural Language Processing, NLP, Large Language Model, LLM, BERT, embedding, entity linking

Suchstring

("IFC" OR "Building Information Modeling" OR "BIM") AND ("semantic" OR "semantisch" OR "semantic mapping" OR "semantic similarity") AND ("Natural Language Processing" OR "NLP" OR "Large Language Model" OR "LLM" OR "BERT" OR "embedding")

Abb. 4.2: Von der Forschungsfrage und den Teilfragestellungen zum Suchstring, eigene Darstellung.

4.2 Anforderungsanalyse

In der zweiten Phase wurden die Erkenntnisse aus der Recherche in Anforderungen überführt. Funktional muss der Prototyp Informationen aus IFC-Modellen auslesen, Materialbezeichnungen bereinigen, KBOB-Daten einbinden, Kandidaten suchen, diese bewerten und die besten Treffer ausgeben. Neben Materialnamen werden Kontextmerkmale wie **IfcEntity**, **PredefinedType**, Eigenschaften und Weiteres berücksichtigt (Details siehe Kapitel 5.3.1).

Nichtfunktional stehen Nachvollziehbarkeit, Robustheit, Erweiterbarkeit und Integrationsfähigkeit im Vordergrund. Der Prototyp soll auch mit kurzen, uneinheitlichen oder unvollständigen Materialbezeichnungen und Bauteileigenschaften umgehen können. Da die Ergebnisse fachlich prüfbar bleiben müssen, wird ein *Human-in-the-loop*-Ansatz (siehe Abschnitt 5.5) gewählt. Der Prototyp unterstützt die Zuordnung, ersetzt aber nicht die Entscheidung der nutzenden Person.

Die Entwicklung des Prototyps erfolgte iterativ. Erkenntnisse aus der Konzeption, der prototypischen Umsetzung und der Evaluation wurden genutzt, um Anforderungen, Datenaufbereitung und Mapping-Logik bei Bedarf anzupassen.

4.3 Konzeption des Prototyps

In der dritten Phase wurden Datenmodell, Schnittstellen und Mapping-Logik festgelegt. Das Datenmodell bestimmt, welche Informationen aus dem IFC-Modell und aus den KBOB-Daten für das Matching verwendet werden. Materialbezeichnungen werden dabei nicht isoliert verarbeitet, sondern mit Kontextinformationen kombiniert. So kann ein allgemeiner Materialname wie «Beton» oder «Stahl» durch ergänzende Bauteileigenschaften besser zugeordnet werden.

Für die Mapping-Logik werden regelbasierte, embeddingbasierte und hybride Verfahren geprüft. Die Auswahl richtet sich nach Datenverfügbarkeit, Nachvollziehbarkeit und Umsetzbarkeit. Ziel ist kein möglichst komplexes Modell, sondern ein Verfahren, das unter realen Bedingungen brauchbare Vorschläge liefert. Die Ausgabe erfolgt als Rangliste mit mehreren KBOB-Kandidaten und Ähnlichkeitsbeziehungsweise Konfidenzwert.

4.4 Prototypische Umsetzung

Die Umsetzung als vierte Phase erfolgt schrittweise. Zuerst werden die IFC-Extraktion und die KBOB-Datenintegration aufgebaut. Danach folgen Vorverarbeitung, Kandidatensuche, semantisches Ranking und Ergebnisdarstellung. Eine einfache Demo-Oberfläche macht die Vorschläge sichtbar und ermöglicht die fachliche Prüfung. Entwicklungsschritte, Annahmen und Anpassungen werden dokumentiert.

Als Datengrundlage dient das IFC-Schema mit den möglichen Entitäten des Schwerpunkts von Brückenbauteilen wie Fahrbahnplatte, Abdichtung, Belag, Widerlager, Pfeiler oder Geländer. Betrachtet werden typische Materialgruppen wie Beton, Bewehrungsstahl, Stahl, Asphalt, Abdichtungen, Beschichtungen, Naturstein, Kunststoffe und Spannstahl.

4.5 Evaluation

Die Evaluation kombiniert quantitative und qualitative Auswertungen. Quantitativ werden die Vorschläge des Prototyps mit einer manuellen Referenzklassifikation verglichen. Zentrale Kennzahl ist, ob die korrekte KBOB-Zuordnung auf dem besten Vorschlag ist. Dieses Top-1-Kriterium passt zu einer möglichen zukünftigen vollständigen Automatisierung, wobei die nutzende Person auch tiefer eingestufte Einträge auswählen kann.

Qualitativ werden Fehlzusammenordnungen, Grenzfälle und der Nutzen von Kontextinformationen untersucht. Dabei wird geprüft, in welchen Fällen zusätzliche IFC-Informationen das Matching verbessern und wo weiterhin eine manuelle Beurteilung nötig bleibt. Bewertet werden auch die Verständlichkeit der Vorschläge und die Anwendbarkeit im Schweizer Kontext.

Eine Evaluation des Prototyps mit realen IFC-Modellen soll zeigen, ob der gewählte Ansatz unter praxisnahen Bedingungen brauchbare und nachvollziehbare Zuordnungsvorschläge liefert. Vorgesehen ist die Anwendung auf mindestens zwei Modelle.

4.6 Gütekriterien und Abgrenzung

Die methodische Qualität wird durch eine dokumentierte Recherche, transparente Anforderungsableitung, nachvollziehbare Entwicklung und Evaluation mit realen Projektdaten abgesichert. Die manuelle Referenzklassifikation dient als Vergleichsbasis für die Trefferqualität.

Es wird kein vollautomatisches Entscheidungssystem entwickelt. Elemente ohne verwertbare Materialinformationen können je nach Kontextinformation nur eingeschränkt oder nicht zugeordnet werden. Die Berechnung von UBP oder CO₂-eq setzt Mengen- oder Volumeninformationen im IFC-Modell voraus. Der Prototyp konzentriert sich auf ausgewählte Bauwerkstypen und Materialgruppen. Das begrenzt die Übertragbarkeit, macht die Umsetzung im Rahmen der Masterarbeit aber prüfbar.

5 Konzept

Dieses Kapitel beschreibt das fachliche und technische Konzept des entwickelten Prototyps. Ziel ist ein Verfahren, das Informationen aus IFC-Modellen mit den KBOB-Ökobilanzdaten verknüpft und der nutzenden Person nachvollziehbare Zuordnungsvorschläge liefert. Im Zentrum steht nicht eine vollautomatische Entscheidung, sondern ein *Human-in-the-loop*-Ansatz: Das System erzeugt priorisierte Kandidaten, die fachlich geprüft, bestätigt oder korrigiert werden können.

Das Konzept verbindet drei Ebenen. Erstens werden relevante Informationen aus dem IFC-Modell extrahiert und in eine für das Matching geeignete Struktur überführt. Zweitens werden die KBOB-Daten als Datenbank aufbereitet (siehe Anhang A.6). Drittens werden IFC-Kontextinformationen und KBOB-Einträge mit einem hybriden Matching-Ansatz verglichen, der regelbasierte Prüfungen und semantische Embeddings kombiniert. Die bestätigten Zuordnungen bilden die Grundlage für die Berechnung von Umweltindikatoren und können zugleich als Trainingsdaten für eine spätere Modellverbesserung genutzt werden.

5.1 Zielbild des Prototyps

Der Prototyp soll IFC-Modelle so aufbereiten, dass Bauteile zu passenden KBOB-Ökobilanzdaten zugeordnet werden können. Dafür müssen Informationen aus unterschiedlichen Quellen (siehe Abb. 5.1) zusammengeführt werden: Das IFC-Modell liefert Bauteile, Materialien, Mengen, Einheiten und Property-Set-Informationen. Die KBOB-Daten liefern die möglichen Zielmaterialien bzw. Bauteile/Vorbereitungsarbeiten mit den entsprechenden Umweltindikatoren wie UBP, Treibhausgasemissionen und Primärenergie. Die zentrale Aufgabe des Prototyps besteht darin, zwischen diesen beiden Datenquellen eine belastbare Zuordnung herzustellen.

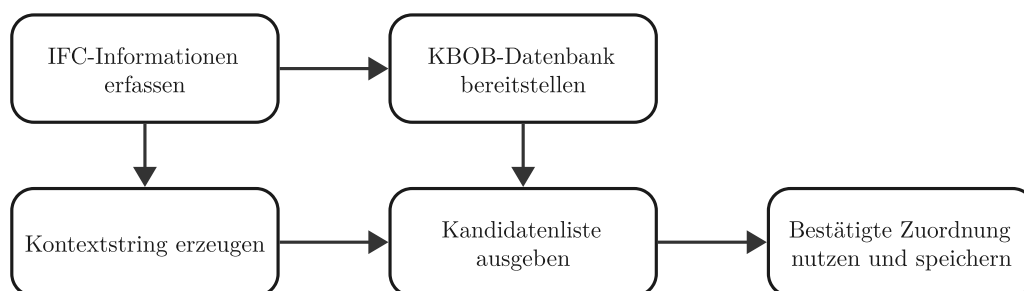


Abb. 5.1: Zentrale Funktionen des Prototyps, eigene Grafik.

Die Zuordnung wird als semantisches Matching verstanden. Dies ist notwendig, weil IFC-Modelle und KBOB-Daten selten dieselbe Terminologie verwenden. Ein IFC-Material kann beispielsweise als «Beton», «Stahlbeton», «Beton C30/37» oder projektspezifisch mit einer Kurzbezeichnung benannt sein. In den KBOB-Daten stehen dagegen festgelegte Material- oder Bauteilbezeichnungen. Ein rein

zeichenbasierter Vergleich wäre dafür zu eng. Das Konzept nutzt deshalb eine Kombination aus strukturierten Modellinformationen, regelbasierten Prüfungen und semantischer Ähnlichkeit.

Der Prototyp soll damit zwei Aufgaben verbinden: Erstens die Zuordnung von IFC-Informationen zu KBOB-Daten, zweitens die anschliessende Verwendung dieser Zuordnung für eine modellbasierte Umweltbewertung. Die Umweltberechnung ist jedoch nur möglich, wenn neben der Zuordnung zum KBOB-Eintrag auch Mengen vorhanden sind. Das Matching-Konzept und das Berechnungskonzept sind deshalb miteinander verbunden, aber nicht identisch.

5.2 Konzeptioneller Systemablauf

Der Systemablauf beginnt mit dem Import einer IFC-Datei (siehe Abb. 5.2). Daraus werden relevante Informationen der Bauteile wie Materialinformationen, Property Sets, Mengen und Einheiten extrahiert. Diese Informationen werden nicht direkt mit den KBOB-Daten verglichen, sondern zuerst in eine strukturierte Zwischenform überführt. Für jedes Element beziehungsweise für jede relevante Materialschicht entsteht ein Datensatz, der als Grundlage für das Matching dient.

Parallel dazu werden die KBOB-Ökobilanzdaten als Datenbank aufbereitet (siehe Anhang A.6). Jeder Eintrag davon enthält eine eindeutige Identifikation, die Bezeichnung, Bezugseinheit und zugehörigen Umweltindikatoren (KBOB und ecobau, 2026). Damit entsteht die Datenbasis, gegen welche die aus dem IFC-Modell erzeugten Kontextinformationen verglichen werden.

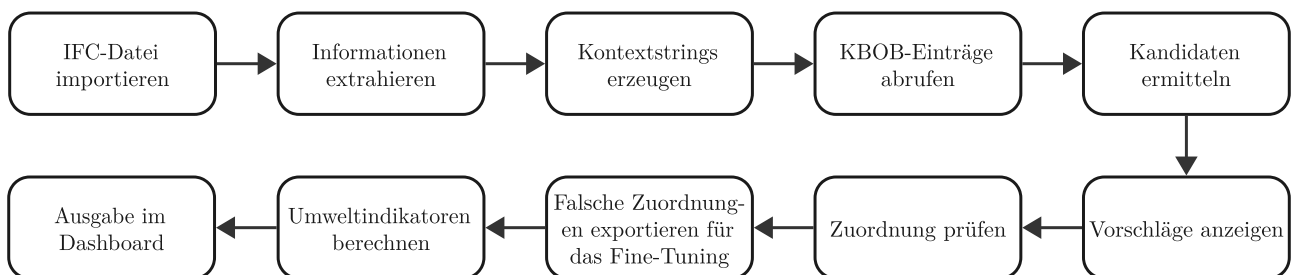


Abb. 5.2: Konzeptioneller Systemablauf des Prototyps, eigene Grafik.

Der eigentliche NLP-basierte Matching-Prozess besteht aus drei Schritten. Zuerst wird aus den IFC-Informationen ein Kontextstring gebildet. Danach werden die KBOB-Kandidaten mit diesem verglichen und abschliessend werden die besten Treffer als Rangliste ausgegeben. Die nutzende Person prüft diese Vorschläge und bestätigt oder korrigiert die Zuordnung. Erst danach wird das finale Mapping für die Berechnung der Umweltindikatoren verwendet.

Der Ablauf trennt bewusst zwischen automatischer Vorverarbeitung und fachlicher Entscheidung. Dadurch bleibt das Verfahren auch bei unvollständigen oder mehrdeutigen Modellinformationen kontrollierbar. Der detaillierte Ablauf ist im Anhang A.4 abgebildet.

5.3 Datenmodell für IFC- und KBOB-Informationen

Das Datenmodell legt fest, welche IFC-Informationen für das semantische Matching verwendet werden. Es bildet nicht das vollständige IFC-Schema ab, sondern reduziert ein Modellobjekt auf die Angaben, die für die Materialzuordnung relevant sind (siehe Abb. 5.3). Aus diesen Angaben wird der Kontext abgeleitet, mit dem ein IFC-Objekt einem passenden KBOB-Eintrag zugeordnet werden kann. Das detaillierte Klassendiagramm ist im Anhang A.5 dargestellt.

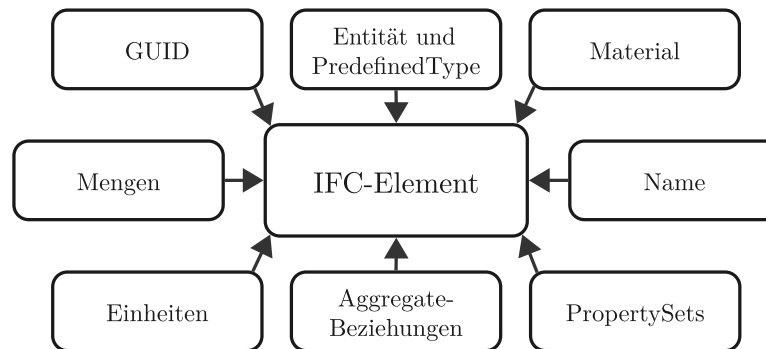


Abb. 5.3: Relevante Informationen für die IFC-Elemente, eigene Grafik.

Die KBOB-Daten dienen als Zielkatalog. Sie enthalten die möglichen Zielzuordnungen mit Bezeichnung, Bezugseinheit und Umweltindikatoren.

Die Verbindung zwischen KBOB-Daten und IFC-Element entsteht im Mapping-Ergebnis. Es enthält alle Informationen, die für Prüfung und Berechnung notwendig sind: IFC- und KBOB-Informationen, Matching-Score und Prüfstatus.

5.3.1 IFC-Informationen

Für das Konzept wird das IFC-Modell als fachlich strukturierte Informationsquelle verstanden. Vor allem die semantische Einordnung eines Bauteils, sein Material, die verfügbaren Mengen und seine Beziehungen im Modell sind relevant. Diese Informationen bestimmen, ob ein KBOB-Eintrag nicht nur sprachlich ähnlich, sondern auch fachlich passend ist. Aus ihnen wird der Kontext für das spätere Matching abgeleitet.

Ausgangspunkt sind konkrete Unterklassen von `IfcElement` (buildingSMART International, 2026). `IfcElement` beschreibt in IFC allgemein physisch vorhandene Komponenten eines Bauwerks, ist selbst aber abstrakt. Für den betrachteten Brückenkontext sind deshalb vor allem konkrete Bauteilklassen wie `IfcWall`, `IfcSlab`, `IfcBeam`, `IfcPile`, `IfcFooting` usw. sowie Bewehrung und Spannglieder relevant. Die IFC-Entität liefert dabei, zusammen mit dem `PredefinedType`, eine erste fachliche Einordnung des Modellobjekts und wird deshalb als Kontextinformation für das Matching verwendet.

Materialinformationen sind in IFC über `IfcRelAssociatesMaterial` mit Elementen verknüpft. Die Materialdefinition kann unterschiedlich granular sein: als einzelnes Material, als Schichtaufbau, als

Profilmaterial oder als Bestandteil eines zusammengesetzten Elements. Für das Konzept bedeutet dies, dass ein Bauteil nicht zwingend genau einem Material entspricht. Ein Bauteil kann mehrere Materialschichten oder Bestandteile enthalten, die für die Zuordnung zu KBOB-Einträgen getrennt betrachtet werden müssen. Materialinformationen auf Elementebene sind dabei spezifischer als Materialinformationen auf Typebene und müssen entsprechend höher gewichtet werden.

Mengeninformationen bilden die zweite zentrale Grundlage. Sie sind notwendig, damit Umweltindikatoren berechnet werden können. In IFC werden solche Mengen über `IfcElementQuantity` beschrieben und über Property-Beziehungen einem Element zugeordnet. Für das Konzept sind insbesondere Volumen, Fläche, Länge, Gewicht und Stückzahl relevant.

Zusätzlich werden Property Sets und Modellbeziehungen als Kontextinformationen verstanden. Property Sets können fachlich relevante Angaben wie Betongüte (C30/37), Herstellungsart (insitu/precast) oder Bewehrungsgrad (XX kg/m³) enthalten. Beziehungen wie räumliche Einordnung und Aggregation helfen, ein Element im Modellzusammenhang zu verstehen. Die räumliche Einordnung erfolgt in IFC über `IfcRelContainedInSpatialStructure` und Aggregationen beschreiben Bauteil-Beziehungen über `IfcRelAggregates`. Für das Konzept ist dies wichtig, weil Materialinformationen in realen Modellen nicht immer am übergeordneten Bauteil liegen, sondern auch auf Unterelementen, Coverings oder Bewehrungsobjekten vorkommen können.

Damit werden vier IFC-Informationsarten unterschieden:

- Bauteilinformationen wie IFC-Entität, `PredefinedType`, Name und Typisierung,
- Materialinformationen wie Einzelmaterial, Schicht, Profil oder Bestandteil,
- Mengen- und Eigenschaftsinformationen wie Volumen, Fläche, Länge, Dichte, Betongüte oder Bewehrungsgrad,
- Strukturinformationen wie räumliche Einordnung, Aggregation, Coverings und modellierte Bewehrung.

Material- und Bauteilinformationen bilden den Kern des semantischen Matchings. Mengeninformationen sind vor allem für die spätere Berechnung relevant. Strukturinformationen dienen dazu, den Modellkontext zu erfassen.

5.3.2 Kontextstring für das semantische Matching

Die im Datenmodell festgelegten IFC-Informationen werden für das Matching nicht einzeln, sondern als kompakter Kontextstring (Query) verwendet. Damit wird vermieden, dass der Materialname allein über die Zuordnung entscheidet. Dies ist nötig, da in der KBOB-Liste z.B. 14 Einträge mit dem Begriff «Beton» (ohne hochbauspezifische Einträge) vorkommen. Der Kontextstring beschreibt das Analyseobjekt so knapp wie möglich und so genau wie nötig (siehe Tab. 5.1).

Welche Informationen in den Kontextstring übernommen werden, richtet sich nach den verfügbaren IFC-Angaben und der regelbasierten Prüfung in Abschnitt 5.4.1. Massgebend sind Angaben, die

den KBOB-Zielkatalog fachlich unterscheiden können, etwa Bauteilart, Material, Stärkeklasse oder Herstellungsart (vgl. Abschnitt 5.3.1 und Abschnitt 5.3.3).

IfcEntity	PredefinedType	Name	Material	StrengthClass	Durchmesser	CastingMethod
IfcSlab	TRACKSLAB	-	Beton NPK G	C30/37	-	INSITU
IfcWall	RETAININGWALL	Bohrpfahl	Beton	-	-	INSITU
IfcBeam	BEAM	-	Stahl	S235	-	-
IfcPile	BORED	-	Beton	C25/30	900	Ortbeton
IfcPile	COHESION	-	Stahlbeton	-	500	PRECAST

Tab. 5.1: Beispiele möglicher Kontextstrings aus den IFC-Informationen, jeweils ein Kontextstring pro Zeile, eigene Tabelle.

Der Mehrwert liegt nicht in der vollständigen Abbildung des IFC-Objekts, sondern in der fachlichen Eingrenzung des Suchraums. Im Beispiel wird nicht nur die Materialbezeichnung «Beton» oder «Stahl» übergeben, sondern mit Bauteilart, Stärkeklasse und Herstellungsart kombiniert. Dadurch lassen sich KBOB-Einträge besser unterscheiden, die ähnliche Bezeichnungen haben, aber unterschiedliche Anwendungen beschreiben.

5.3.3 KBOB-Informationen

Die KBOB-Daten bilden den Zielkatalog des Matchings. Die Originaldatei enthält insgesamt 347 Einträge, von denen jedoch ein grosser Teil hochbauspezifisch ist (Türen, Fenster, Bodenbeläge, Dämmstoffe) oder herstellerspezifische Werte enthält, die nicht dem Schweizer Durchschnitt entsprechen. Für diese Arbeit wurden 110 Einträge als infrastrukturelevant identifiziert. Die bereinigte Liste wird als Datenbank¹ aufgesetzt (siehe Anhang A.6). Für das Matching ist ausschliesslich die Bezeichnung (in der Tabelle unter «Vorbereitungsarbeit/Baumaterial») relevant. Für die spätere Berechnung werden zusätzlich die Umweltindikatoren und die Bezugseinheit benötigt (siehe Tab. 5.2).

Vorbereitungsarbeit/ Baumaterial	Rohdichte/ Flächenmasse	Bezug	UBP'21	Primärenergie [kWh oil-eq]		Treibhausgasemissionen [kg CO2-eq]	Biogener Kohlenstoff [kg C]
				erneuerbar	nicht erneuerbar		
Tiefbaubeton	2'350	kg	163	0.016	0.189	0.109	0

Tab. 5.2: Relevante Informationen aus der KBOB-Liste am Beispiel Tiefbaubeton. Die vollständige Tabelle befindet sich im Anhang A.6, eigene Tabelle.

Die Bezugseinheit ist für die Berechnung zentral. Ein IFC-Element kann beispielsweise ein Volumen enthalten, während ein KBOB-Eintrag auf Kilogramm, Quadratmeter oder Laufmeter bezogen ist. Das Matching allein löst diese Umrechnung nicht. Für die spätere Berechnung sind in der Implementierung

¹Die Datenbank ist auf [GitHub](#) (Macdonald, 2026) zugänglich.

Regeln im Python-Code nötig, welche die Einheiten abgleichen und bei Bedarf mithilfe der Rohdichte umrechnen.

5.3.4 Mapping-Datensatz

Der Mapping-Datensatz hält fest, welcher KBOB-Eintrag einem aus dem IFC-Modell abgeleiteten Datensatz zugeordnet wird. Er enthält den gebildeten IFC-Kontextstring (siehe Kapitel 5.3.2), die vorgeschlagenen KBOB-Kandidaten mit Ähnlichkeitswerten und den ausgewählten KBOB-Eintrag.

Der Ähnlichkeitswert dient zur Rangierung der Vorschläge. Die nutzende Person kann den vorgeschlagenen Eintrag übernehmen oder durch einen anderen KBOB-Eintrag ersetzen. Die zugehörigen KBOB-Ökobilanzdaten werden anschliessend über diesen Eintrag aus der Datenbank abgerufen.

Für die Berechnung der Ökobilanz wird die gewählte Zuordnung mit den verfügbaren Mengeninformationen aus dem IFC-Modell kombiniert. Korrigierte Zuordnungen können zudem für Evaluation und spätere Modellverbesserungen genutzt werden.

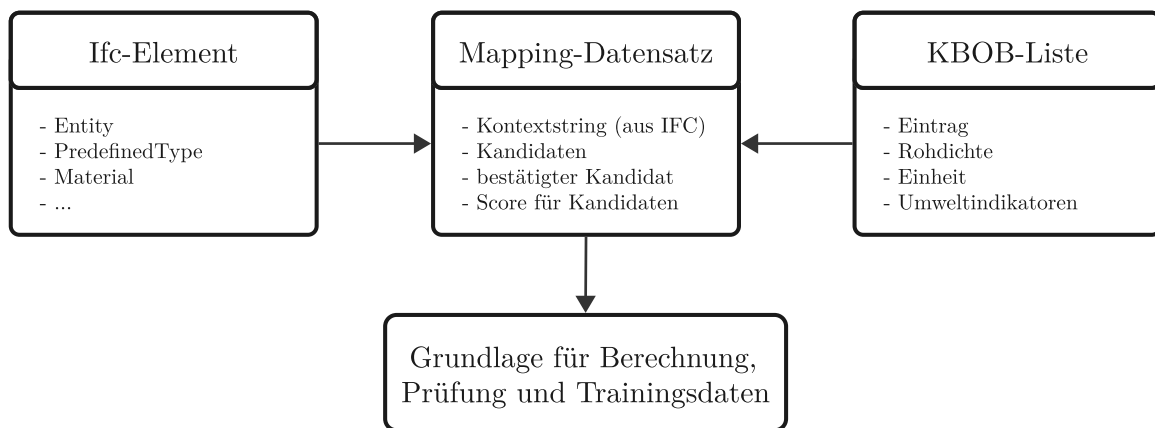


Abb. 5.4: Der Mapping-Datensatz als zentrales Objekt für den Prototyp, eigene Grafik.

5.4 Matching-Strategie

Die Matching-Strategie beschreibt, wie ein aus dem IFC-Modell abgeleiteter Query einem passenden Eintrag der KBOB-Daten zugeordnet wird. Die dafür verwendeten IFC-Informationen sind in Abschnitt 5.3.2 beschrieben. Die Strategie unterscheidet zwischen fachlicher Prüfung, semantischer Kandidatensuche und abschliessender Auswahl durch die nutzende Person.

Die Zuordnung erfolgt nicht vollautomatisch. Das System schlägt mehrere passende KBOB-Einträge vor und ordnet sie nach einem Ähnlichkeitswert. Die nutzende Person prüft die Vorschläge und bestätigt oder korrigiert die Zuordnung. Damit bleibt die fachliche Verantwortung bei der prüfenden Person, während das Matching die Suche im Zielkatalog unterstützt.

5.4.1 Regelbasierte Prüfung

Die regelbasierte Prüfung legt fest, welche IFC-Informationen in den Query übernommen werden. Ihre Funktion besteht darin, einzelne Modellinformationen nur dann zu verwenden, wenn sie für die Zuordnung fachlich aussagekräftig sind.

Dies betrifft die Elementnamen. Sie werden nicht grundsätzlich als Matching-Information verwendet. Sie werden nur dann in den Query übernommen, wenn sie auf eine fachlich unterscheidbare Bauweise oder einen KBOB-nahen Anwendungsfall hinweisen, etwa Spundwand, Schlitzwand oder Bohrpfahlwand sowie entsprechende Wortteile oder Übersetzungen. Alle anderen Elementnamen werden nicht berücksichtigt, da sie die semantische Zuordnung verzerren können.

Die Auswahl und Rangierung der KBOB-Kandidaten erfolgt anschliessend über das embedding-basierte Matching.

5.4.2 Embedding-basiertes Matching

Das embedding-basierte Matching ist der Kern des Prototyps. Der IFC-Query und die Bezeichnungen der KBOB-Einträge werden in Vektoren überführt und anhand ihrer Nähe (semantische Ähnlichkeit) verglichen. So können Query und KBOB-Einträge verglichen werden, die nicht dieselbe Zeichenfolge verwenden, aber fachlich ähnlich sind.

Die Kandidatensuche erfolgt mit einem Bi-Encoder². Dieser vergleicht den IFC-Query mit den KBOB-Einträgen und erzeugt eine Rangliste möglicher Zuordnungen. Damit wird der gesamte Zielkatalog effizient durchsucht und eine Auswahl fachlich prüfbarer Kandidaten bereitgestellt.

Als Basismodell für den Bi-Encoder wird BGE-M3 von J. Chen et al. (2025) gewählt. Die Vorauswahl erfolgte anhand der Kriterien Suchqualität, Mehrsprachigkeit, Geschwindigkeit, Kurztext-Robustheit und Fine-Tuning-Eignung (siehe Anhang A.7) mit den offiziellen sbert.net-Modellen (Reimers und Gurevych, 2025) sowie weiteren verbreiteten Sentence-Transformer-Modelle auf Hugging Face. In Vergleichstests erzielte BGE-M3 sehr starke Ergebnisse (siehe Anhang A.8). Für die Wahl sprechen zudem Multilingualität (über 100 Sprachen, gemischte deutsch-englische Queries), Fine-Tuning über `sentence-transformers` ohne Neutraining der Architektur, eine Kontextlänge von 8'192 Tokens, starke Zero-Shot-Performance sowie die MIT-Lizenz.

Ein zusätzliches Re-Ranking mit einem Cross-Encoder³ wäre als Erweiterung möglich. Dabei würden die bereits gefundenen Kandidaten nochmals gemeinsam mit dem IFC-Query bewertet und neu sortiert. Im vorliegenden Konzept wird dieser Schritt jedoch nicht als Bestandteil des umgesetzten Matching-Ablaufs vorgesehen, sondern als mögliche Weiterentwicklung betrachtet.

²Ein Bi-Encoder kodiert Query und Kandidaten mit zwei getrennten Encodern (daher *Bi*) unabhängig in je einen Vektor. Die Kandidatenvektoren können vorberechnet werden, was eine effiziente Suche über grosse Kataloge ermöglicht.

³Ein Cross-Encoder verarbeitet Query und Kandidat gemeinsam in einem einzigen Encoder und erzeugt direkt einen Ähnlichkeitswert. Er ist genauer als ein Bi-Encoder, aber langsamer, da jedes Paar einzeln bewertet werden muss.

5.4.3 Bewertung und Sortierung der Kandidaten

Die Kandidaten werden anhand eines Scores sortiert. Der Score beschreibt im Konzept die semantische Ähnlichkeit zwischen dem IFC-Query und einem KBOB-Eintrag. Er dient dazu, die gefundenen Kandidaten in eine prüfbare Reihenfolge zu bringen. Der bestbewertete Kandidat steht damit zuerst in der Liste, wird aber nicht automatisch als richtige Zuordnung übernommen.

Der Score ist kein absoluter Wahrheitswert. Er zeigt nicht, ob ein KBOB-Eintrag fachlich korrekt ist, sondern wie ähnlich der Eintrag dem Query nach dem verwendeten Matching-Verfahren ist. Ein sprachlich ähnlicher Eintrag kann fachlich ungeeignet sein, etwa wenn die Anwendung oder die Materiallogik nicht zum Modellobjekt passt.

Die Kandidatenliste dient deshalb als Entscheidungsgrundlage für die nutzende Person. Erst der bestätigte KBOB-Eintrag wird mit den Mengeninformationen aus dem IFC-Modell kombiniert und für die Berechnung der Umweltindikatoren verwendet. Konkrete Berechnung, Skalierung und allfällige Normalisierung des Scores werden in der prototypischen Umsetzung beschrieben.

5.5 Human-in-the-loop und Export von Trainingsdaten

Das Matching erzeugt eine Kandidatenliste, aber noch kein finales Mapping. Das Mapping entsteht erst, wenn die nutzende Person einen KBOB-Eintrag auswählt und die Auswahl speichert. Grundlage dafür sind die Kandidatenliste von dem embedding-basierten Matching (Abschnitt 5.4.2), die verwendeten IFC-Informationen aus Abschnitt 5.3 bzw. der daraus erstellte Query (Abschnitt 5.3.2) und die Darstellung des betroffenen Modellelements im Viewer. Der Score dient dabei zur Rangierung der automatisch vorgeschlagenen KBOB-Einträge.

Im Dashboard kann ein vorgeschlagener Eintrag übernommen, ein anderer KBOB-Eintrag gewählt oder die Auswahl leer gelassen werden (siehe Abb. 5.5). Gespeichert werden die gewählte Zuordnung und der zugehörige Score.

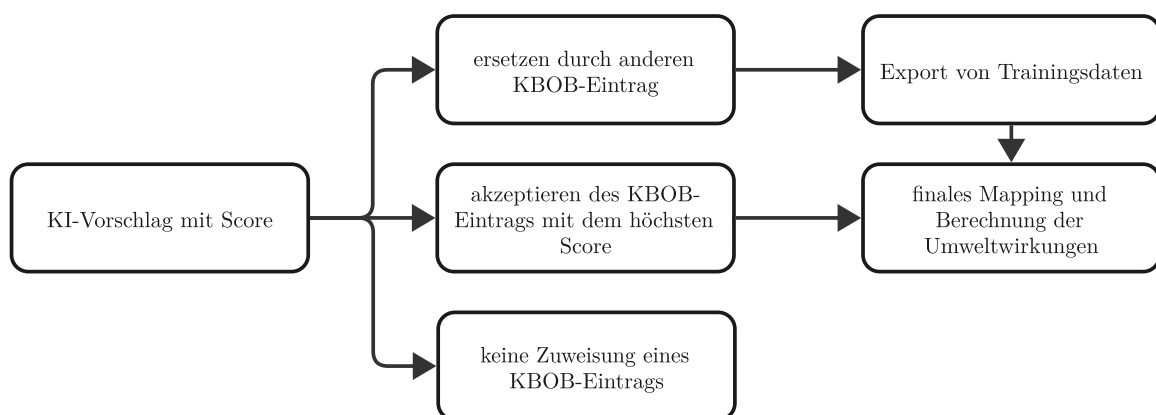


Abb. 5.5: Der Human-in-the-loop Ansatz für den Prototyp, eigene Grafik.

Nur gespeicherte Zuordnungen werden in die Ergebnisdaten übernommen und für die Berechnung der Umweltindikatoren verwendet. Falls die Auswahl manuell leer gelassen wird, wird für das betroffene Analyseobjekt keine Berechnung der Umweltwirkung durchgeführt. Die für die Berechnung erforderlichen Mengen sind Voraussetzung. Die nutzende Person bekommt eine Fehlermeldung, falls die Mengen fehlen.

Manuell geänderte Zuordnungen können als Trainingsdaten exportiert werden. Dabei wird der aus den IFC-Feldern gebildete Query mit dem gewählten KBOB-Eintrag als Zielwert gespeichert. Der Export erfolgt als zeilenweise ausgerichtetes Query- und Expected-Paar. Bereits vorhandene Paare werden nicht erneut gespeichert. Dadurch können die exportierten Paare in einer späteren Trainingsphase des Embedding-Modells verwendet werden.

5.6 Konzept für das Fine-Tuning

Für das KI-Training (Fine-Tuning) werden eigens erstellte Trainingspaare vorgesehen (siehe Abb. 5.6). Diese bilden typische Zuordnungsfälle im Kontext von Brückenbauwerken. Der Kontextstring folgt der in Abschnitt 5.3.2 beschriebenen Struktur und enthält die für das Matching relevanten IFC-Informationen. Der Zielwert ist ein Eintrag aus dem KBOB-Katalog gemäss Abschnitt 5.3.3. Das Training zielt darauf ab, die Distanz der Vektoren von IFC-basierten Queries und KBOB-Einträgen zu minimieren im Kontext des Schweizer Infrastrukturbaus.

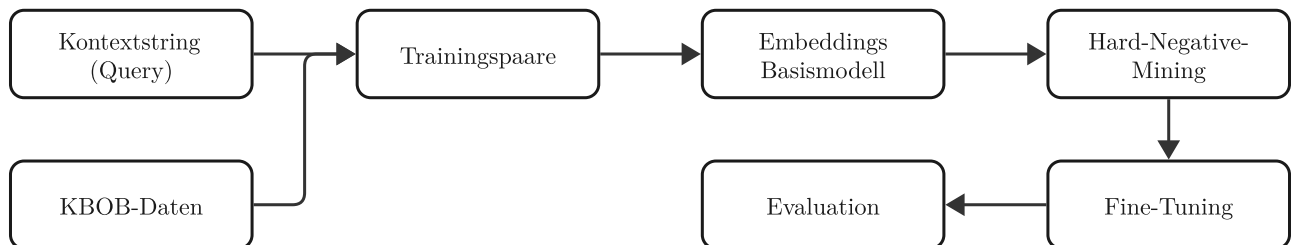


Abb. 5.6: Genereller Ablauf für das Fine-Tuning, eigene Grafik.

Der Datenkatalog soll unterschiedliche Materialgruppen, Bauteilarten und Formulierungsvarianten enthalten. Entscheidend ist, dass nicht nur eindeutige Zuordnungen abgebildet werden, sondern auch Fälle mit ähnlichen KBOB-Einträgen. Dazu werden neben positiven Paaren auch Hard Negatives vorgesehen. Hard Negatives sind fachlich oder sprachlich nahe Einträge, die nicht als korrekte Zuordnung gelten. Beispiele für das Query «IfcBeam Beton C40/50 Fertigteil» sind «Betonfertigteil normalfest», «Tiefbaubeton» und «Magerbeton». Diese Einträge gehören zur gleichen Materialgruppe, unterscheiden sich aber in Anwendung und Bedeutung für die Zuordnung. Hard Negatives sollen das Modell deshalb darin unterstützen, nicht nur den Oberbegriff «Beton» zu erkennen, sondern nahe beieinanderliegende Zielkatalogeinträge zu unterscheiden.

Die Hard Negatives werden aus der Evaluation eines Basismodells⁴ abgeleitet. Aus der Evaluation dieses Basismodells werden Kandidaten übernommen, die weit oben eingestuft werden, aber falsch sind. Dabei werden korrekte KBOB-Einträge derselben Query ausgeschlossen, damit keine zulässigen Zuordnungen als Negativbeispiele verwendet werden.

5.6.1 Metriken

Für die Evaluation des trainierten Modells stehen die folgenden Metriken im Vordergrund. Damit die Bewertung nachvollziehbar ist, werden zuerst die verwendeten Begriffe und Referenzgrößen definiert. Nachdem die Metriken generell erklärt sind, erfolgt ein Beispiel (siehe Abschnitt 5.6.1.1).

Es werden N Queries ausgewertet. Zu jeder Query i wird vorgängig festgelegt, welche KBOB-Einträge als fachlich korrekt gelten. Die KBOB-Einträge werden in der Menge R_i zusammengefasst. Das Modell erzeugt für dieselbe Query eine sortierte Kandidatenliste. Ein Kandidat gilt als korrekt, wenn sein KBOB-Eintrag in R_i enthalten ist. Sind für eine Query mehrere KBOB-Einträge fachlich zulässig, enthält R_i mehrere davon. Dies kann bei ungenauen Queries wie «IfcBeam Beton» der Fall sein, welche z. B. «Tiefbaubeton», «Betonfertigteil hochfest» oder «Betonfertigteil normalfest» als fachlich passende KBOB-Einträge haben.

Der Rang $rank_i$ bezeichnet die Position des ersten korrekten Kandidaten in der Kandidatenliste.

$$\text{Hit@1} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(rank_i \leq 1)$$

Hit@1 misst den Anteil der Queries, bei denen ein korrekter KBOB-Eintrag auf Rang 1 steht. Die Kennzahl ist streng, weil nur der erste Vorschlag zählt.

$$\text{Hit@K} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(rank_i \leq K)$$

Hit@K misst den Anteil der Queries, bei denen mindestens ein korrekter KBOB-Eintrag innerhalb der ersten K Vorschläge enthalten ist. Für den *Human-in-the-loop*-Ablauf ist diese Kennzahl wichtig, weil nicht nur der erste Vorschlag, sondern eine kurze Kandidatenliste geprüft werden kann.

⁴Als Basismodell wird hier das noch nicht auf den Anwendungsfall feinabgestimmte Sprachmodell verstanden, das als Ausgangspunkt für das Fine-Tuning dient.

$$\text{MRR@K} = \frac{1}{N} \sum_{i=1}^N \begin{cases} \frac{1}{\text{rank}_i}, & \text{falls } \text{rank}_i \leq K \\ 0, & \text{sonst} \end{cases}$$

MRR@K berücksichtigt die Position des ersten korrekten Treffers. Ein Treffer auf Rang 1 erhält den Wert 1, ein Treffer auf Rang 2 den Wert $\frac{1}{2}$ und ein Treffer auf Rang 5 den Wert $\frac{1}{5}$. Dadurch wird ein Modell besser bewertet, wenn der erste korrekte KBOB-Eintrag weiter oben in der Kandidatenliste steht.

5.6.1.1 Beispiel der Metriken

Als Beispiele werden vier Queries betrachtet. Für jede Query ist bekannt, auf welchen Rängen korrekte KBOB-Einträge in der Kandidatenliste erscheinen:

- Query 1: korrekte Treffer auf den Rängen [1, 3, 8]
- Query 2: korrekte Treffer auf den Rängen [2, 12]
- Query 3: korrekte Treffer auf den Rängen [7, 11, 13, 18]
- Query 4: korrekter Treffer auf Rang [6]

Für Hit@K und MRR@K zählt jeweils der erste korrekte Treffer. Damit ergeben sich die ersten korrekten Ränge: $\text{rank} = [1, 2, 7, 6]$.

$$\text{Hit@1} = \frac{1 + 0 + 0 + 0}{4} = 0.25$$

Hit@1 beträgt 0.25, weil nur bei Query 1 ein korrekter KBOB-Eintrag auf Rang 1 steht.

$$\text{Hit@5} = \frac{1 + 1 + 0 + 0}{4} = 0.50$$

Hit@5 beträgt 0.50, weil bei Query 1 und Query 2 ein korrekter KBOB-Eintrag innerhalb der ersten fünf Ränge vorkommt. Query 3 und Query 4 werden nicht gezählt, da ihre ersten korrekten Treffer erst auf Rang 7 beziehungsweise Rang 6 liegen.

$$\text{Hit@10} = \frac{1 + 1 + 1 + 1}{4} = 1.00$$

Hit@10 beträgt 1.00, weil bei allen vier Queries mindestens ein korrekter KBOB-Eintrag innerhalb der ersten zehn Ränge enthalten ist.

$$\text{MRR@10} = \frac{1}{4} \left(1 + \frac{1}{2} + \frac{1}{7} + \frac{1}{6} \right) = 0.4524$$

Query 1 trägt mit 1 am stärksten zum Ergebnis von MRR@10 bei, weil der erste korrekte Treffer auf Rang 1 liegt. Query 2 trägt mit $\frac{1}{2}$ bei. Query 3 und Query 4 tragen mit $\frac{1}{7}$ beziehungsweise $\frac{1}{6}$ weniger bei, weil die ersten korrekten Treffer weiter unten in der Rangliste liegen.

5.7 Ergebnisdarstellung

Nach der bestätigten Zuordnung können die Umweltindikatoren berechnet werden. Dazu wird die aus dem IFC-Modell ermittelte Menge mit der Bezugseinheit des gewählten KBOB-Eintrags abgeglichen. Stimmen Menge und Bezugseinheit nicht überein, ist eine Umrechnung erforderlich, beispielsweise von Volumen zu Masse über die Rohdichte. Fehlen dafür notwendige Angaben, muss die Berechnung als unvollständig gekennzeichnet oder mit einer dokumentierten Annahme ergänzt werden.

Die Ergebnisse werden im Dashboard grafisch dargestellt. Neben Summenwerten soll ersichtlich bleiben, welcher KBOB-Eintrag verwendet wurde und welche Menge zugrunde liegt. Dadurch können hohe Umweltbeiträge gezielt geprüft werden. Die Ergebnisdarstellung dient damit auch der Qualitätssicherung. Auffällige Beiträge können auf tatsächlich hohe Umweltwirkungen hinweisen, aber auch auf Fehler bei Menge, Einheit oder Materialzuordnung.

5.8 Rückbindung an das IFC-Modell

Konzeptionell kann das IFC-Modell nicht nur als Eingabe, sondern auch als Träger berechneter Umweltinformationen verstanden werden. IFC 4.3 sieht dafür die Property Sets `Pset_EnvironmentalImpactIndicators` und `Pset_EnvironmentalImpactValues` vor (buildingSMART International, 2026). Ersteres beschreibt Umweltindikatoren bezogen auf eine funktionale Einheit, letzteres Umweltwirkungswerte eines konkreten Elements.

Nach bestätigter KBOB-Zuordnung könnten ausgewählte Ergebnisse der Umweltberechnung dem ursprünglichen Modellelement zugeordnet werden. Treibhausgasemissionen und Primärenergie lassen sich grundsätzlich in diese Logik einordnen. Für UBP'21 besteht dagegen kein eindeutig vorgesehenes Standardfeld, weshalb eine projektspezifische Ergänzung oder eine dokumentierte Referenz auf die KBOB-Datengrundlage nötig wäre.

Diese Rückbindung wird in der Arbeit nicht umgesetzt, sondern als möglicher konzeptioneller Anschluss betrachtet.

5.9 Abgrenzungen und konzeptionelle Entscheidungen

Das Konzept grenzt den Prototyp in vier Punkten ein. Erstens wird kein autonomes Entscheidungssystem vorgesehen. Das Matching erzeugt eine sortierte Kandidatenliste, das finale Mapping entsteht jedoch erst durch die Auswahl der nutzenden Person. Diese Abgrenzung ist notwendig, weil der Score nur die Ähnlichkeit zwischen IFC-Query und KBOB-Eintrag beschreibt und keine fachlich richtige Zuordnung garantiert.

Zweitens beschränkt sich das Konzept auf ausgewählte Infrastrukturbauteile und Materialgruppen. Es wird keine vollständige Abdeckung aller IFC-Entitäten oder aller KBOB-Einträge angestrebt. Der Fokus liegt auf einem prüfbaran Anwendungsbereich mit typischen Brückenbauteilen, häufigen Materialien und KBOB-Einträgen, die für diesen Kontext verwendbar sind.

Drittens ersetzt der Prototyp keine vollständige Nachhaltigkeitsbewertung und keine vollständige Ökobilanz eines Bauwerks. Betrachtet wird nur die ökologische Dimension der Nachhaltigkeit (vgl. Tab. 2.1 und Tab. 2.3) anhand der KBOB-Umweltindikatoren. Soziale und wirtschaftliche Aspekte werden nicht bewertet. Innerhalb der ökologischen Dimension wird keine eigene lebenszyklusbasierte Prozesskette aufgebaut. Der Prototyp übernimmt die Umweltkennwerte des ausgewählten KBOB-Eintrags und verknüpft sie mit den verfügbaren IFC-Mengen. Damit werden nur jene Herstellungs-, Entsorgungs- oder Prozessanteile berücksichtigt, die im jeweiligen KBOB-Datensatz bereits enthalten sind. Der Prototyp ergänzt keine projektspezifische Prozesskette für das konkrete Bauwerk. Nicht zusätzlich modelliert werden insbesondere Baustelleneinrichtung, Bauhilfsmassnahmen, Schalung, Einbauprozesse, Verdichtung, projektspezifische Transporte, Nutzung, Instandhaltung sowie konkrete Rückbauabläufe, sofern sie nicht bereits Bestandteil des gewählten KBOB-Eintrags sind.

Viertens wird zwischen Matching-Fehlern und Berechnungsgrenzen unterschieden. Falsche oder fehlende Ergebnisse können aus unvollständigen Materialangaben, fehlenden oder falschen Mengen, Annahmen zum Bewehrungsgehalt oder der Verzinkung sowie aus einer fachlich unpassenden KBOB-Auswahl entstehen. Die Rückbindung berechneter Umweltinformationen in das IFC-Modell wird nicht umgesetzt, sondern in Abschnitt 5.8 als möglicher konzeptioneller Anschluss beschrieben.

5.10 Konzeptionelle Herausforderungen

Die erste Herausforderung liegt nicht in der Zuordnungsmethode selbst, sondern in der Qualität und Struktur der Modellinformationen. Für das Matching stehen nur die in Abschnitt 5.3.1 beschriebenen Informationen zur Verfügung. Fehlende oder nicht eindeutig ableitbare Angaben können fachlich nicht zuverlässig ergänzt werden.

Eine zweite Herausforderung betrifft die Auswahl der Informationen für den Kontextstring. Mehr verfügbare Modellinformationen führen nicht automatisch zu besseren Vorschlägen. Projektspezifische Namen oder allgemeine Beschreibungen können die semantische Suche verfälschen, wenn sie keinen

Bezug zu unterscheidbaren KBOB-Einträgen haben. Deshalb wird der Kontextstring auf Merkmale beschränkt, die den Suchraum fachlich eingrenzen können (siehe Abschnitt 5.3.2).

Eine weitere Herausforderung entsteht durch die unterschiedliche Granularität von Modell und Zielkatalog. Das IFC-Modell kann Angaben enthalten, die im KBOB-Katalog nur generisch abgebildet sind. Umgekehrt kann der KBOB-Katalog Varianten unterscheiden, die im Modell nicht enthalten sind. Solche Fälle bleiben fachlich mehrdeutig und müssen über die Kandidatenliste geprüft werden (siehe Abschnitt 5.5).

Nach der bestätigten Zuordnung erfolgt die Berechnung der Umweltindikatoren. Der ausgewählte KBOB-Eintrag kann nur ausgewertet werden, wenn seine Bezugseinheit mit einer passenden Menge aus dem Modell verbunden werden kann. Fehlen dafür notwendige Mengen oder Umrechnungsgrößen, ist die Berechnung unvollständig, auch wenn das Matching einen passenden Kandidaten liefert (siehe Abschnitt 5.7).

Schliesslich darf der Score nicht als fachlicher Wahrheitswert verstanden werden. Er beschreibt die Ähnlichkeit zwischen Kontextstring und KBOB-Eintrag und dient der Sortierung der Vorschläge. Ob ein Eintrag hinsichtlich Anwendung, Materiallogik und Bezugseinheit passt, bleibt Teil der fachlichen Prüfung. Die Grenzen des Scores ergeben sich aus der Matching-Strategie in Abschnitt 5.4. Nichtsdestotrotz erhält die nutzende Person durch den Score ein Eindruck wie sicher sich das Modell ist. Wenn die Differenz zum zweiten Eintrag gross ist, ist sich das KI-Modell sicher, bei einer Mehrdeutigkeit, werden die ersten Einträge einen ähnlichen Score haben.

5.11 Zusammenfassung des Konzepts

Das Konzept verknüpft IFC-Informationen und KBOB-Ökobilanzdaten über einen materialbezogenen Mapping-Datensatz. Aus den in Abschnitt 5.3.1 beschriebenen Bauteil-, Material- und Eigenschaftangaben wird ein Kontextstring gebildet. Mengeninformationen werden für die spätere Berechnung verwendet. Der KBOB-Katalog dient dabei doppelt: Die Bezeichnung wird für das Matching genutzt, Bezugseinheit und Umweltindikatoren für die Berechnung.

Die Matching-Strategie kombiniert eine fachliche Prüfung der verwendeten IFC-Angaben mit einem embedding-basierten Vergleich zwischen Kontextstring und KBOB-Bezeichnungen. Das Ergebnis ist eine sortierte Kandidatenliste mit Score. Der Score beschreibt die semantische Ähnlichkeit im gewählten Matching-Verfahren, ist jedoch keine Garantienachweis für eine fachlich richtige Zuordnung. Ein KBOB-Eintrag wird deshalb erst durch die Auswahl der nutzenden Person zum Mapping.

Bestätigte Zuordnungen bilden die Grundlage für die Berechnung der Umweltindikatoren, sofern passende Mengen vorhanden sind. Korrigierte Zuordnungen können für die spätere Modellverbesserung verwendet werden. Das Konzept ist auf ausgewählte Infrastrukturbauteile und Materialgruppen begrenzt. Es ersetzt keine vollständige Nachhaltigkeitsbewertung und ergänzt keine projektspezifische Lebenszyklus-Prozesskette ausserhalb der im gewählten KBOB-Eintrag enthaltenen Anteile.

6 Implementierung des Prototyps

Dieses Kapitel beschreibt die prototypische Umsetzung des in Kapitel 5 beschriebenen Ansatzes. Im Fokus stehen die technische Struktur und der Datenfluss von der IFC-Datei bis zur Ergebnisdarstellung.

6.1 Softwarearchitektur und Datenfluss

Dieser Abschnitt gibt einen Überblick über die Laufzeitstruktur des Prototyps und das durchgehende Datenobjekt. Zunächst werden die Komponenten und ihr Zusammenspiel vorgestellt, anschliessend das gemeinsame JSONL-Datenformat erläutert. Detailregeln zur IFC-Extraktion, zur KBOB-Datenbasis, zum Matching und zum Fine-Tuning vertiefen die Abschnitte 6.2 bis 6.5.

6.1.1 Komponenten des Prototyps

Der Prototyp ist als Python-Anwendung mit Streamlit-Dashboard umgesetzt und kann auf [Hugging Face](#)¹ ausprobiert werden. Die Verarbeitungskette besteht aus vier aufeinanderfolgenden Schritten: IFC-Verarbeitung, semantische Kandidatensuche, fachlicher Nutzerentscheid und Berechnung der Umweltindikatoren. Verbunden werden diese Schritte über eine gemeinsame JSONL-Arbeitsdatei (siehe Abschnitt 6.1.2).

Nach dem Upload (siehe Abb. 6.1 oder Anhang A.11) wird die IFC-Datei in zwei Laufzeitpfade überführt: für die Viewer-Anzeige in ein statisches Verzeichnis und für die Extraktion in eine temporäre Datei. Die Extraktion läuft als separater Prozess und erzeugt daraus die Arbeitsdatei. Das Matching ergänzt anschliessend dieselbe Datei um eine sortierte Kandidatenliste möglicher KBOB-Einträge.

Im Tab «AI-Mapping» (siehe Abb. 6.2) erfolgt die fachliche Auswahl: Vorgeschlagene Kandidaten können übernommen, ersetzt oder offen gelassen werden. Gespeicherte Zuordnungen dienen als Eingabe für die Berechnung.

Die Berechnung liest die gespeicherten Zuordnungen, verbindet sie mit den verfügbaren Mengenangaben und führt die Resultate in die Dashboard-Datenstruktur zurück. Im Tab «Charts» (siehe Abb. 6.3) werden diese Resultate aggregiert dargestellt.

Zusätzlich ist im Tab «AI-Mapping» der IFCLite-Viewer von Trümpler (2026) eingebunden. Die GUID-basierte Kopplung zwischen Auswahlliste und 3D-Ansicht unterstützt die fachliche Prüfung der Zuordnungen am Modell, ist aber nicht Teil der Berechnungslogik.

¹Der Prototyp wurde als Hugging Face Space veröffentlicht. Die Plattform kann gehostete Anwendungen aus technischen oder organisatorischen Gründen vorübergehend einschränken. Falls der primäre Space nicht erreichbar ist, steht unter [Hygrokopisch/ifc-kbob-ai-matcher](#) eine alternative Adresse zur Verfügung. Sollten beide Spaces nicht verfügbar sein, kann der Prototyp lokal ausgeführt werden. Der Quellcode ist auf [GitHub](#) zugänglich.

IFC-material mapping with AI

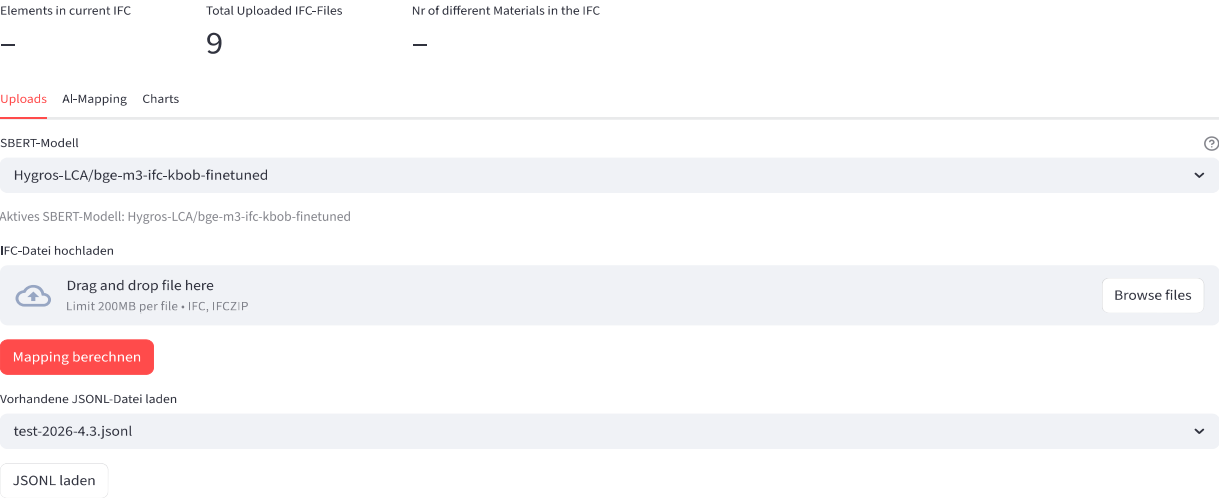


Abb. 6.1: Tab «Uploads» des Prototyps, Screenshot Benutzeroberfläche des Prototyps.

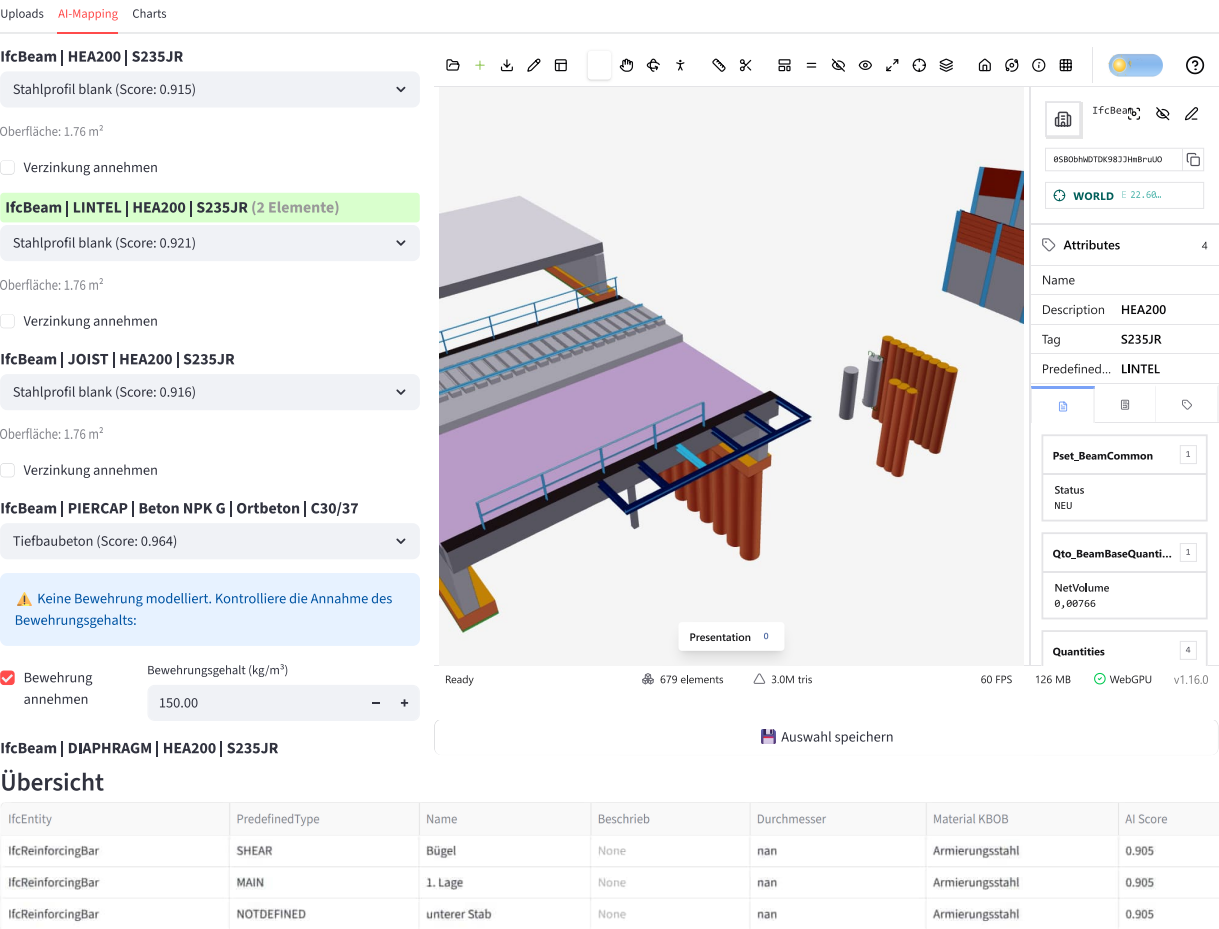


Abb. 6.2: Tab «AI-Mapping» des Prototyps mit selektiertem Stahlträger, Screenshot Benutzeroberfläche des Prototyps.

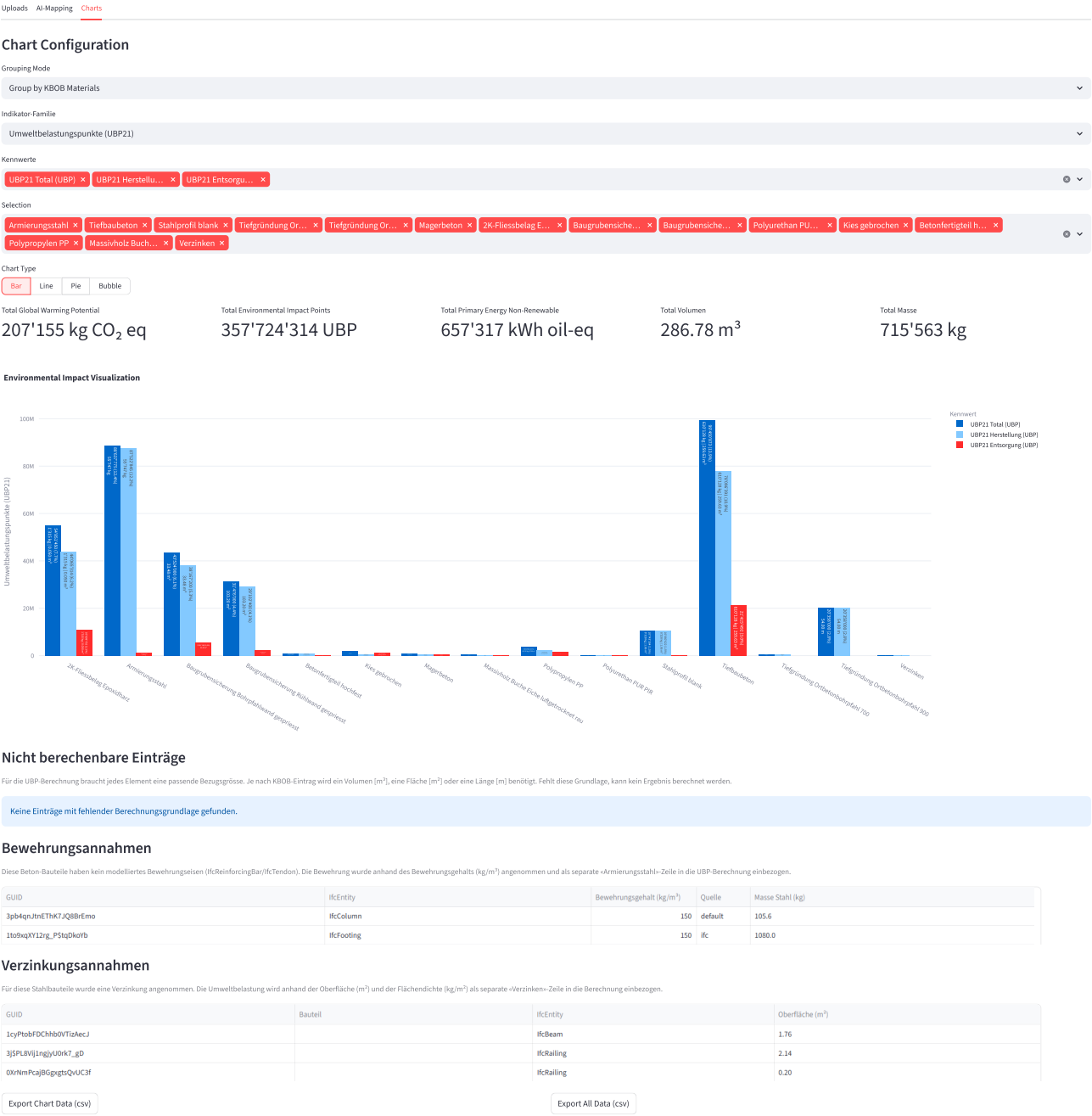


Abb. 6.3: Tab «Charts» des Prototyps mit Säulen-Diagramm der berechneten UBP'21, Screenshot Benutzeroberfläche des Prototyps.

6.1.2 Datenformate und Übergaben

Das durchgehende Zwischenformat ist JSONL (siehe Abb. 6.4) nach Ward (2026). Jede Zeile repräsentiert einen aus der IFC-Datei abgeleiteten Datensatz. Ein Datensatz kann ein Element oder eine einzelne Materialschicht beschreiben. Bei mehrschichtigen Materialien erfolgt die Trennung über `MaterialLayerIndex`. Welche IFC-Felder exportiert und wie Mengen und Einheiten aufbereitet werden, ist in Abschnitt 6.2 beschrieben.

```
{
  "IfcEntity": "IfcBeam",
  "PredefinedType": "EDGEBEAM",
  "Name": " ",
  "Description": " ",
  "GUID": "3ILRKS6L0kv6e2GuvGSD1",
  "Status": ["NEU"],
  "CastingMethod": "Ortbeton",
  "StrengthClass": "C30/37",
  "ExposureClass": "XC4, XD3, XF4",
  "NetVolume": 5.625,
  "GrossVolume": 5.625,
  "NetSurfaceArea": 35.9,
  "GrossSurfaceArea": 35.9,
  "ReinforcementVolumeRatio": 200.0,
  "HasModeledRebar": true,
  "AggregateParentGUID": "35yzdeMxD6e904ecV6WIKY",
  "Material": ["Beton NPK G"],
  "MaterialLayerIndex": 1,
  "top_k_matches": [
    {"material": "Tiefbaubeton", "score": 0.962},
    {"material": "2K-Fliessbelag Epoxidharz", "score": 0.698},
    {"material": "Polycarbonat PC", "score": 0.601},
    {"material": "Betonfertigteil normalfest", "score": 0.593},
    {"material": "2-Komponenten Klebstoff", "score": 0.502},
    {"material": "Armierungsstahl", "score": 0.499},
    {"material": "Polyethylen PE", "score": 0.496}
  ],
  "selected_kbob_material": "Tiefbaubeton",
  "selected_ai_score": 0.962,
  "selected_on": "2026-04-12T16:00:21.397683+00:00",
  "reinforcement_accepted": false,
  "galvanization_accepted": false
}
```

Abb. 6.4: Eintrag eines IfcBeam in der JSONL-Datei (nicht alle `top_k_matches` dargestellt), eigene Darstellung.

Nach dem Matching enthält derselbe Datensatz das Feld `top_k_matches` (in Abb. 6.4 blau markiert). Darin werden vorgeschlagene KBOB-Einträge mit Score in sortierter Reihenfolge gespeichert. Der Score dient der Rangierung und ist kein fachlicher Richtigkeitsnachweis. Die Query-Bildung und Kandidatensuche folgen in Abschnitt 6.4.

Beim Speichern im Dashboard werden `selected_kbob_material`, `selected_ai_score` und `selected_on` in dieselbe JSONL-Datei zurückgeschrieben (in Abb. 6.4 orange markiert). Als Zuordnungsschlüssel dienen GUID und `MaterialLayerIndex`, sodass Extraktion, Matching, Nutzerentscheid und Berechnung konsistent verknüpft bleiben.

6.2 IFC-Extraktion und Datenaufbereitung

Dieser Abschnitt beschreibt, wie aus einer IFC-Datei die JSONL-Arbeitsdatei erzeugt wird. Zunächst werden die exportierten Felder und Filterregeln erläutert, dann die Aggregationslogik für zusammengesetzte Bauteile und schliesslich die Aufbereitung von Materialschichten, Mengen und Einheiten.

6.2.1 Exportierte IFC-Informationen

Die Extraktion erfolgt mit `IfcOpenShell` (`IfcOpenShell Contributors`, 2026) über ein klar sequenziertes Verfahren. Zuerst wird das Modell mit `ifcopenshell.open(...)` geladen. Danach werden mit `model.by_type>IfcElement` alle `IfcElement`-Objekte gelesen und gegen die Exportregeln gefiltert. Ausgeschlossen werden für den Brückenbau nicht relevante Entitäten: `IfcOpeningElement`, `IfcElementAssembly`, `IfcCurtainWall`, `IfcFlowTerminal`, `IfcDoor`, `IfcWindow`, `IfcFurnishingElement`, `IfcAnnotation`, `IfcSpace`, `IfcVirtualElement` sowie `IfcCovering` mit `PredefinedType CEILING`.

Für jedes exportierte Objekt werden `IfcEntity` und GUID übernommen. `PredefinedType` wird direkt aus dem Modell gelesen. Wenn der Wert `USERDEFINED` für `PredefinedType` definiert ist, wird ein vorhandener `ObjectType` als typisierender Wert verwendet. Beispielsweise kann so in der Autorensoftware für nicht implementierte `IfcEntities` wie `IfcRail` der `ObjectType` «RAIL» definiert werden in Kombination mit der `IfcEntity IfcBuildingElementProxy`. Dadurch können zusätzliche wichtige Kontextinformationen aus der IFC-Datei ausgelesen werden für das Matching. Property

Sets werden über `ifcopenshell.util.element.get_psets(...)` ausgelesen. Leere Angaben sowie Platzhalter wie `NOTDEFINED` werden nicht exportiert. Dadurch enthält die JSONL-Datei nur fachlich verwertbare Felder.

Numerische Grössen werden mit den projektspezifischen IFC-Einheiten in SI-kompatible Grössen überführt. Dazu werden die Skalierungsfaktoren für `LENGTHUNIT`, `AREAUNIT` und `VOLUMEUNIT` über `ifcopenshell.util.unit.calculate_unit_scale(...)` bestimmt. Auf dieser Basis werden Längen, Flächen und Volumen skaliert. Zusätzlich werden abgeleitete Grössen gebildet, welche für das Matching zur KBOB-Datenbank bzw. für die spätere Umweltberechnung wichtig sind: Für `IfcPile` wird aus Länge und Volumen ein Durchmesser berechnet, für `IfcWall` wird aus Länge und Höhe eine `Ansichtsfläche` gebildet und für `IfcCovering` wird `NetArea` als `Ansichtsfläche` übernommen. Fehlt `NetVolume` (Nettovolumen mit Abzug möglicher Öffnungen und Aussparungen), wird ersatzweise `GrossVolume` (Bruttovolumen ohne Abzüge von z. B. Aussparungen) verwendet.

Die Exportstruktur enthält mehr Informationen als die Bi-Encoder Query. Welche Felder in den Query-Text einfließen, beschreibt Abschnitt 6.4.1. `Description` bleibt für die Anzeige und Nachvollziehbarkeit im Dashboard erhalten, fließt aber nicht in die Bi-Encoder Query ein.

6.2.2 Aggregationen, Coverings und Bewehrung

Die Aggregationslogik ist notwendig, weil zusammengesetzte Bauelemente im IFC-Modell häufig in mehrere Teilobjekte zerlegt sind. Das betrifft beispielsweise eine Rühlwand, bei der Wandanteile und Trägerprofile als getrennte Objekte modelliert sein können (siehe Abb. 6.5), für die Zuordnung zu KBOB-Datensätzen aber als ein Bauteil betrachtet werden müssen. Deshalb wird die Parent-Child-Struktur explizit geführt.

Technisch werden zwei Beziehungspfade ausgelesen: reguläre Aggregationen über das inverse Attribut `IsDecomposedBy` auf `IfcRelAggregates` sowie Covering-Beziehungen über `HasCoverings` und `IfcRelCoversBldgElements`. Beide Pfade werden in einer gemeinsamen internen Childliste zusammengeführt. `IfcRelCoversBldgElements` wird auch berücksichtigt, obwohl die Beziehung in IFC 4.3 als veraltet gekennzeichnet ist.

Die Exportselektion nutzt nur reguläre Aggregationschildren. Hat ein Parent genau ein solches Child, wird entlang dieser 1:1-Kette auf das Child weitergegangen. Das ist sinnvoll, weil bei einer IFC-Elementzerlegung objektbezogene Informationen wie Geometrie oder Material schema-konform an den Teilobjekten liegen können. Bei Verzweigungen oder ohne reguläres Aggregations-Child bleibt das Parent als eigenes Exportobjekt bestehen. `IfcCovering`, `IfcReinforcingBar`, `IfcReinforcingMesh` und `IfcTendon` bleiben unabhängig davon als eigenständige Analyseobjekte erhalten.

Im Export entstehen daraus die Strukturfelder `AggregateChildGUIDs` (rekursiv erreichbare Nachkommen), `AggregateParentGUID` (Bezug zum Parent) und `HasModeledRebar` (Bewehrung in derselben Aggregation vorhanden).

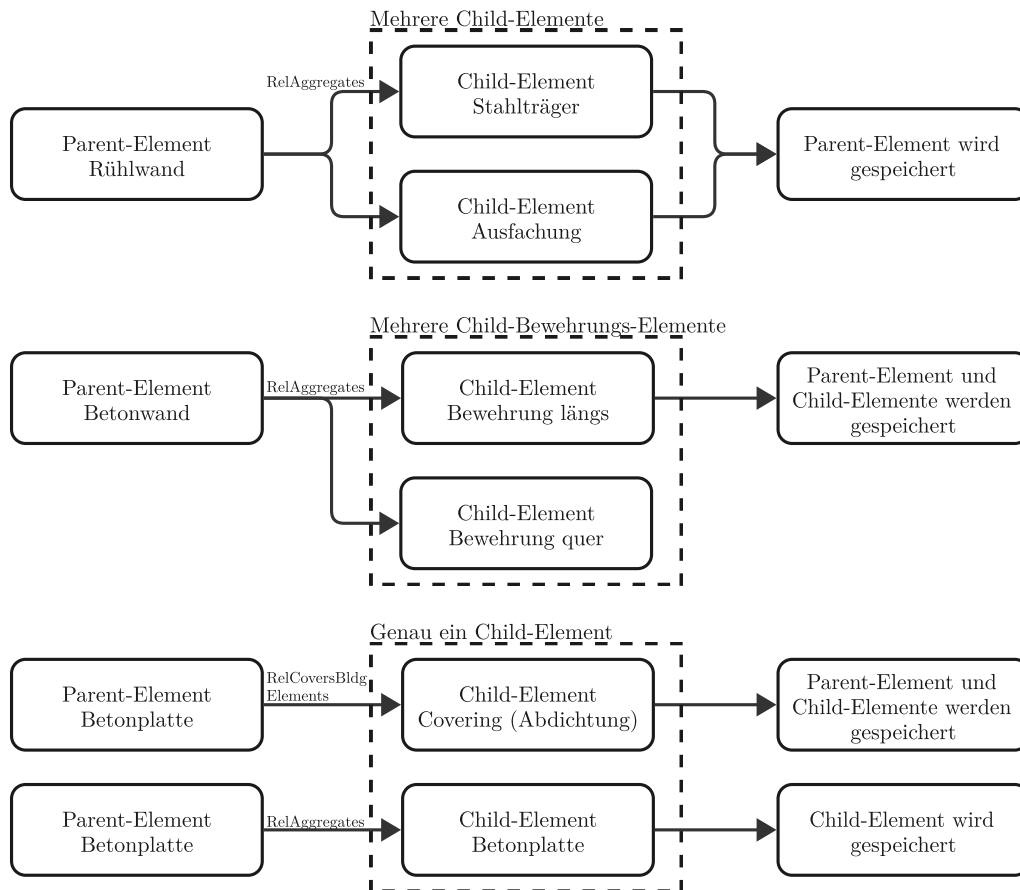


Abb. 6.5: Aggregationslogik des Prototyps, eigene Darstellung.

6.2.3 Materialschichten, Mengen und Einheiten

Die Materialzuordnung wird zuerst mit `ifcopenshell.util.element.get_material` ausgelesen. Reicht das nicht aus, wird geprüft, ob das Material direkt dem Bauteil oder seinem Typ zugeordnet ist. Dabei werden Materialzuordnungen über `IfcRelAssociatesMaterial` und Typbezüge über `IfcRelDefinesByType` berücksichtigt. Verarbeitet werden `IfcMaterial`, `IfcMaterialLayer`, `IfcMaterialLayerProfile`, `IfcMaterialConstituentSet` und die jeweiligen `Set`-Varianten. Einträge mit identischem Namen werden zusammengeführt.

Liefert diese Auswertung Materialinformationen, erzeugt der Export für jeden erkannten Materialeintrag einen eigenen JSONL-Datensatz. Die laufende Nummer wird in `MaterialLayerIndex` gespeichert. Ist eine Schichtdicke vorhanden, wird sie zusätzlich als `MaterialLayerThickness` ausgegeben. Sind für alle erkannten Einträge Schichtdicken vorhanden, werden `NetVolume` und `GrossVolume` proportional zu diesen Dicken aufgeteilt. Fehlt jede Materialzuordnung, bleibt ein einzelner elementbezogener Datensatz ohne Materialfeld bestehen.

Die Einheitenskalierung erfolgt wie in Abschnitt 6.2.1 beschrieben. Die Extraktion stellt damit die für Matching und Berechnung nutzbaren Material- und Mengeninformatoren in einer einheitlichen JSONL-Struktur bereit. Welche Grösse für die spätere Umweltberechnung massgebend ist, wird in

diesem Schritt noch nicht entschieden. Diese Auswahl erfolgt erst nach der gespeicherten Zuordnung zum KBOB-Eintrag im Berechnungsschritt (vgl. Abschnitt 6.3.2).

6.3 KBOB-Datenbasis und Umweltberechnung

Dieser Abschnitt beschreibt zunächst die KBOB-Datenbasis, die sowohl dem Matching als auch der Berechnung zugrunde liegt, und anschliessend die Ermittlung der Umweltindikatoren aus gespeicherten Zuordnungen.

6.3.1 Datenbasis und Umweltkennwerte

Die KBOB-Daten liegen im Prototyp als SQLite-Datenbank mit der Tabelle `Oekobilanzdaten` vor. Für das semantische Matching wird daraus ausschliesslich die Spalte `Material` gelesen. Dieselbe Datenbasis speist im Dashboard im Tab «AI-Mapping» auch die vollständige Materialliste für manuelle Korrekturen. Das Matching verarbeitet damit nur die Bezeichnung der Einträge.

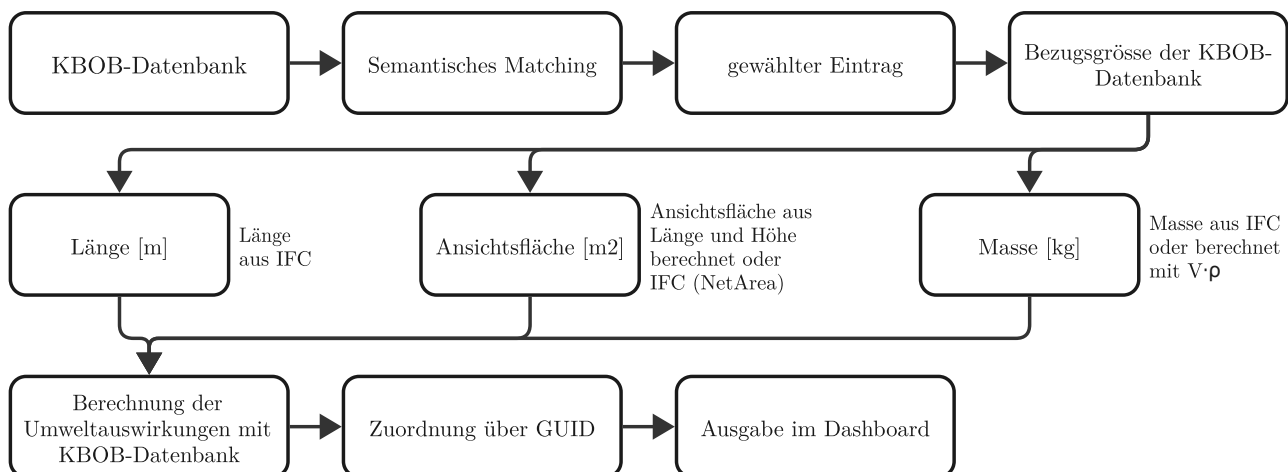


Abb. 6.6: Ablauf vom KBOB-Eintrag zur Ausgabe im Dashboard, eigene Darstellung.

Für die Umweltberechnung werden aus derselben Tabelle zusätzlich `Rohdichte`, `Bezug`, `Einheit` sowie alle numerischen Kennwertspalten gelesen. Die Implementierung behandelt `Material`, `Rohdichte`, `Bezug` und `Einheit` als Metadaten und verwendet die übrigen Spalten als berechenbare Umweltkennwerte. In der vorhandenen Datenbank betrifft dies UBP'21, Treibhausgasemissionen und mehrere Primärenergiekennwerte (siehe Abb. 6.6).

6.3.2 Berechnung der Umweltindikatoren

Die Berechnung verarbeitet die JSONL-Arbeitsdatei zeilenweise. Für jede Zeile wird der gewählte KBOB-Eintrag mit den Kennwertspalten der Datenbank verknüpft. Die in Abschnitt 6.3.1 definierten Metadaten dienen als Steuerinformationen. Alle übrigen numerischen Spalten werden für die Berechnung direkt aus der Datenbank abgerufen.

Der zentrale Schritt ist die Bestimmung einer passenden Bezugsgrösse pro Datensatz (siehe Abb. 6.6). Je nach Bezug des KBOB-Eintrags wird Länge (m), Ansichtsfläche (m²), Masse (kg) oder Volumen (m³) herangezogen. Bei kg-Bezug werden vorhandene Massenfelder bevorzugt. Fehlen diese, wird die Masse aus Volumen und Rohdichte berechnet. Für Volumen dient **NetVolume**. Fehlt dieser Wert, wird **GrossVolume** herangezogen. Sonderfälle betreffen **IfcCovering** (synthetisches Volumen aus Fläche und Schichtdicke) sowie Bewehrungselemente (Masse aus **Weight** und **Count**).

Der Berechnungswert wird mit den Kennwertfaktoren multipliziert. Parallel werden Diagnostikfelder («Bezugsgrösse», «Berechnungswert» und «Fehlende Berechnungsgrundlage») geschrieben, damit nachvollziehbar bleibt, ob ein Ergebnis berechnet werden konnte.

Für zwei fachliche Zusatzentscheide aus dem Tab «AI-Mapping» erzeugt die Berechnung zudem synthetische Zusatzzeilen: Bei angenommener Bewehrung wird aus **NetVolume** und Bewehrungsgehalt eine Armierungsstahl-Masse berechnet. Bei bestätigter Verzinkung wird aus **NetSurfaceArea/GrossSurfaceArea** eine Verzinken-Masse gebildet. Diese Beiträge fliessen wie reguläre Datensätze in die Summen ein. Damit folgt die Umsetzung dem Konzept aus Abschnitt 5.7: Berechnet wird nur, was über Bezugseinheit und verfügbare Mengen fachlich belastbar bestimmt werden kann.

6.4 Semantisches Matching

Dieser Abschnitt behandelt die Umsetzung der Bi-Encoder-Suche: Zunächst wird erläutert, wie aus jedem JSONL-Datensatz ein Query-Text entsteht und wie die Kandidatenliste erzeugt wird. Danach folgen die Darstellung im Dashboard mit den fachlichen Zusatzentscheiden sowie der Export korrigierter Zuordnungen als Trainingsdaten.

6.4.1 Query-Bildung und Kandidatensuche

Die Matching-Stufe liest die JSONL-Arbeitsdatei zeilenweise und ergänzt jeden Datensatz um **top_k_matches**. Dieses Feld enthält 30 vorgeschlagene KBOB-Einträge mit Score. Der Suchkorpus besteht aus den Materialbezeichnungen der KBOB-Datenbank.

Für den Bi-Encoder-Vergleich wird aus jedem Datensatz ein Query-Text gebildet. Regulär übernommen werden **IfcEntity**, **PredefinedType**, **Material**, **Durchmesser**, **CastingMethod** und **StrengthClass**. **Name** ist kein reguläres Query-Feld (siehe Abb. 6.7). Es wird nur für zusätzliche Disambiguierungs-Tokens genutzt, wenn eine **IfcWall** erkannte Begriffe zu Baugrubensicherungen enthält (vgl. Abschnitt 6.2.1).

Die KBOB-Materialbezeichnungen werden als Korpus codiert und mit den codierten Query-Texten verglichen. Verwendet wird die Cosinus-Ähnlichkeit mit normalisierten Embeddings. Die Treffer werden als **material-score**-Paare gespeichert. Der Score wird für die Ausgabe linear auf den Bereich von 0 bis 1 skaliert.

IfcPile BORED Beton NPK L 600 Ortbeton C20/25
IfcReinforcingBar B500B
IfcBuildingElementProxy RAIL Stahl
IfcRailing HANDRAIL S235JR
IfcWall RETAININGWALL Fertigteil Rühlwand

Abb. 6.7: Beispiel-Queries aus der selbst erstellten IFC-Datei, eigene Darstellung.

6.4.2 Score, Kandidatenliste und manuelle Auswahl

Im Tab «AI-Mapping» zeigt das Dashboard die Kandidatenliste als bearbeitbare Auswahl an. Vor der Anzeige werden JSONL-Datensätze zu Gruppen zusammengefasst. Die Gruppierung verwendet `IfcEntity`, `PredefinedType`, `Name`, `Material`, `Durchmesser` und `MaterialLayerIndex`. Bewehrungselemente werden gesondert nach ihrer IFC-Entität gruppiert.

Die nutzende Person kann einen Vorschlag übernehmen, einen anderen KBOB-Eintrag aus der vollständigen Materialliste auswählen oder keine Zuordnung speichern.

Für ausgewählte KBOB-Materialien bietet das Dashboard zusätzliche fachliche Entscheide: Bei «Tiefbaubeton» und «Bohrpfahlbeton» ohne modellierte Bewehrung (`HasModeledRebar: false`) kann ein Bewehrungsgehalt in kg/m^3 erfasst werden. Die Quelle des Werts («ifc», «default» oder «user») wird mitgespeichert. Bei «Stahlprofil blank» kann eine Verzinkung bestätigt werden, sofern eine Oberfläche (`NetSurfaceArea/GrossSurfaceArea`) im Datensatz vorhanden ist.

Beim Speichern schreibt das Dashboard `selected_kbob_material`, `selected_ai_score` und `selected_on` in die JSONL-Arbeitsdatei zurück. Wird ein KBOB-Eintrag ausserhalb der Kandidatenliste gewählt, bleibt `selected_ai_score` leer.

6.4.3 Export korrigierter Zuordnungen

Das Dashboard exportiert manuell geänderte Zuordnungen als Trainingspaare für ein zukünftiges Fine-Tuning. Erfasst werden nur Fälle mit gesetztem KBOB-Eintrag, in denen die Zuordnung gegenüber der aktuellen Vorauswahl geändert wurde. Für jedes Trainingspaar wird der Query-Text aus dem JSONL-Datensatz mit derselben Feldlogik wie im Matching rekonstruiert. Der gewählte KBOB-Eintrag bildet den Zielwert. Der Export schreibt zwei zeilenweise aufeinander abgestimmte TXT-Dateien. Bereits vorhandene Query-Zielwert-Paare werden nicht noch einmal exportiert.

6.5 Fine-Tuning des Bi-Encoders

Dieser Abschnitt beschreibt die Trainingspipeline für das domänenspezifische Fine-Tuning des in Abschnitt 5.4.2 begründeten BGE-M3. Zuerst wird die Aufbereitung der Trainingsdaten erläutert, dann das Hard-Negative Mining und schliesslich der Trainingsprozess mit `MultipleNegativesRankingLoss`.

6.5.1 Trainingsdaten und Aufbereitung

Die Trainingsdaten bestehen aus Regeln abgeleiteten Query-KBOB-Paaren in zwei zeilenweise abgestimmten TXT-Dateien: eine Datei mit Queries (in derselben Feldlogik wie im Matching) und eine mit den zugehörigen erwarteten KBOB-Bezeichnungen. Pro Query können mehrere gültige Zielwerte durch Trennzeichen angegeben werden. Die Trainingsdaten bestehen aus ca. 11'400 Paaren. Die Details zu den Trainingsdaten sind im Anhang A.9 ersichtlich.

Die Aufbereitung der Trainingsdaten entfernt zunächst doppelte Query-KBOB-Paare. Optional begrenzt sie zudem, wie oft derselbe KBOB-Eintrag im Datensatz vorkommen darf, damit häufige Zielwerte das Training nicht dominieren. Anschliessend werden Queries und Positives vereinheitlicht: Leerzeichen werden getrimmt, Gross-/Kleinschreibung per *casefold* normalisiert und mehrfache Leerzeichen eliminiert (z.B. wird «IfcBeam GIRDER_SEGMENT Beton NPK G C30/37 insitu» zu «ifcbeam girder_segment beton npk g c30/37 insitu»).

6.5.2 Hard-Negative Mining

Hard Negatives sind KBOB-Einträge, die für eine Query plausibel erscheinen, aber nicht zu den korrekten Zielmaterialien gehören. Sie dienen als schwierige Gegenbeispiele, damit der Bi-Encoder eng verwandte KBOB-Einträge besser unterscheiden kann.

Hard Negatives werden aus den Fehlern der vorangehenden Evaluation des Basismodells BGE-M3 mit den Trainingsdaten gewonnen. Berücksichtigt werden standardmässig nur Zeilen, in denen der Top-1-Treffer falsch ist. Als Hard-Negative-Kandidaten dienen die in `top10_materials` gespeicherten Vorhersagen. Diese Kandidaten werden verworfen, wenn sie bereits als korrektes Material derselben Query enthalten sind. Dadurch werden falsche Hard-Negatives verhindert. Aus den verbleibenden Kandidaten werden in Reihenfolge der Top-10-Liste sechs Hard Negatives pro Query übernommen.

6.5.3 Training mit MultipleNegativesRankingLoss

Das Fine-Tuning wurde auf «Calculon» durchgeführt, dem HPC-Cluster der Hochschule für Informatik der FHNW. Der Cluster wird vom Institute for Data Science betrieben und verfügt über zehn Rechenknoten mit NVIDIA-GPUs. Für das Training wurde eine RTX A4500 mit 20 GB Arbeitsspeicher reserviert.

Das Training wurde iterativ durchgeführt. In jedem Durchgang wurden zunächst Trainingsdaten erstellt, dann das Modell trainiert und anschliessend ausgewertet, welche Queries nach dem Training noch falsch zugeordnet werden (Hit@1-Fehler). Auf dieser Basis wurden die Trainingsdaten gezielt verbessert und das Training erneut gestartet. Dieser Zyklus wurde wiederholt, bis die Trefferquote > 90 % war.

Innerhalb jedes Trainingslaufs wurde der Datensatz auf Query-Ebene in einen Trainings- und einen Evaluationsanteil aufgeteilt, sodass die Evaluation im Training nur Queries enthält, die das Modell während des Trainings nicht gesehen hat.

Parameter	Wert
Basismodell	BAAI/bge-m3
Maximale Tokenlänge	128
Epochen	2
Batchgrösse	32
Lernrate	2×10^{-5}
Warmup-Anteil	0.1
Entwicklungsanteil	0.1
Scheduler	Linear mit Warmup
Mixed Precision	FP16
Hard-Negative-Modus	fallback
Hard-Negative-Selektion	random_preselected
Anzahl Hard Negatives	2
Evaluationsmetrik	Hit@5 + MRR@10
Seed	42

Tab. 6.1: Verwendete Trainings-Hyperparameter des Fine-Tunings auf «Calculon», eigene Tabelle.

Als Verlustfunktion diente `MultipleNegativesRankingLoss` von `sentence-transformers`. Pro Query wurden zwei Hard Negatives aus den vorab ermittelten Kandidaten (siehe Abschnitt 6.5.2) entnommen. Für jedes Hard Negative entstand ein Triplet aus Query, korrektem KBOB-Eintrag und dem Hard Negative. Das Modell lernt, die Query näher an den korrekten Eintrag und weiter weg vom Hard Negative zu verschieben. Da das Training in Batches von 32 Triplets (siehe Tab. 6.1) erfolgt, wirken zusätzlich die korrekten Materialien aller anderen Batch-Beispiele als *in-batch negatives*. Um widersprüchliche Gradienten zu vermeiden, stellt die Batch-Zusammenstellung sicher, dass jedes Material innerhalb eines Batches höchstens einmal als korrekte Antwort vorkommt.

Beispiel: Enthält ein Batch Triplets für Query A (korrekt: «Tiefbaubeton») und Query B (korrekt: «Stahlprofil blank»), dann dient «Stahlprofil blank» automatisch als Negativbeispiel für Query A und umgekehrt. Hätten beide Queries dasselbe korrekte Material, würde es gleichzeitig als Positiv und als in-batch Negative wirken, was ein widersprüchliches Lernsignal wäre. Deshalb stellt die Batch-Zusammenstellung sicher, dass jedes Material pro Batch höchstens einmal als korrekte Antwort vorkommt.

Nach jeder Epoche, also nachdem das Modell einmal alle Trainingsbeispiele durchlaufen hat, wurde es auf dem Evaluationsanteil bewertet. Als Metrik für den besten Modellstand diente Hit@5 (siehe Abschnitt 5.6.1.1). Bei Gleichstand entschied MRR@10. Nur der beste Modellstand wurde als Endergebnis übernommen und als neues Modell gespeichert.

6.6 Technische Grenzen und Zwischenfazit

Zum Abschluss dieses Implementierungskapitels werden die technischen Grenzen des Prototyps zusammenfassend eingeordnet.

Die Implementierung ist ein Prototyp und auf den Kontext Brückenbau ausgelegt. Die Filterregeln der Extraktion, die regelbasierten Ergänzungen im Query, die Annahmen zur Bewehrung und Verzinkung sowie der verwendete Ausschnitt des KBOB-Katalogs sind auf diesen Anwendungsbereich abgestimmt. Für andere Infrastrukturtypen wären Anpassungen an Filterregeln, Query-Bildung und Trainingsdaten erforderlich. Weitere Grenzen entstehen durch die Modellqualität und durch Annahmen in der Berechnungslogik. Diese Punkte werden getrennt von den konzeptionellen Abgrenzungen in Abschnitt 5.9 aufgeführt, da sie aus der konkreten Umsetzung und den verwendeten Daten entstehen.

Eine grundlegende Grenze ergibt sich daraus, dass der Prototyp nur verarbeiten kann, was im IFC-Modell vorhanden ist. Bauteile oder Materialien, die nicht modelliert wurden, erscheinen weder in der Extraktion noch in der Berechnung. Dazu gehören beispielsweise Schalung, Fugenbänder, Umgebungsarbeiten oder provisorische Hilfskonstruktionen. Die Ergebnisse bilden deshalb immer nur den modellierten Anteil eines Bauwerks ab.

Eng damit verbunden ist die Detailtiefe der Modellierung. Ein grob modelliertes Element ohne Materialschichten, ohne Property-Set-Angaben und ohne differenzierte Mengen liefert weniger Kontextinformationen für das Matching und weniger Berechnungsgrundlagen als ein detailliert ausgearbeitetes Modell. Die Qualität der Ergebnisse hängt damit direkt von der Informationstiefe ab, die das IFC-Modell bereitgestellt hat.

Weiter setzt die Berechnung voraus, dass im IFC-Modell eine Menge als Volumen oder Masse vorhanden ist (vgl. Abschnitt 6.3.2). Fehlen diese Grössen, kann das betroffene Element nicht ausgewertet werden.

Schliesslich kann das Matching nur Zuordnungen zu den im KBOB-Katalog enthaltenen Einträgen vorschlagen. Ist ein konkretes Material im Katalog nicht abgebildet, bleibt die beste Übereinstimmung eine Annäherung. Die nutzende Person muss in solchen Fällen fachlich entscheiden, ob der vorgeschlagene Eintrag als Stellvertreter geeignet ist.

Die verbleibenden technischen Grenzen betreffen nicht modellierte Bauteile, die Informationstiefe des IFC-Modells, die Verfügbarkeit passender Mengen zur KBOB-Bezugseinheit, die Abdeckung des Katalogs sowie den auf Brückenbau begrenzten Anwendungsbereich. Eine vertiefte Einordnung folgt in Abschnitten 8.2 und 8.5.

7 Ergebnisse

Dieses Kapitel präsentiert die Ergebnisse des in Kapitel 5 konzeptionierten und in den Abschnitten 6.4 und 6.5 umgesetzten Matching-Verfahrens. Die Auswertung erfolgt in drei Schritten. Zuerst wird die Trefferqualität des feinabgestimmten Modells auf kontrollierten Testdaten untersucht. Danach wird die Robustheit gegenüber fehlerhaften und unvollständigen Eingaben geprüft. Abschliessend wird der Prototyp auf reale IFC-Modelle aus Infrastrukturprojekten angewendet und mit dem nicht feinabgestimmten Basismodell verglichen.

7.1 Evaluationsaufbau

Der Testdatensatz besteht aus 389 Queries aus einem eigens mit Tekla Structures erstellten IFC-Modell und zusätzlich manuell definierten Einträgen. Jeder Query ist im selben Format aufgebaut wie die Queries des produktiven Matchings (vgl. Abschnitt 6.4.1). Sie enthält mindestens eine IFC-Entität, je nach Query kommen eine Materialbezeichnung sowie optional ein vordefinierter Typ, eine Stärkeklasse und/oder eine Herstellungsart hinzu. Zu jeder Query ist mindestens ein fachlich korrekter KBOB-Eintrag als Referenz hinterlegt. Bei mehrdeutigen Queries (161/389) können mehrere Einträge als korrekt gelten. Darunter sind auch fast gleichwertige Einträge wie «Stahlblech verzinkt» und «Stahlblech blank». Bei den verschiedenen Holz-Einträgen¹ wird nicht differenziert, da die Umweltbelastung einen nur sehr geringen Anteil ausmacht im Vergleich zu Stahlbeton. Aufgrund des begrenzten Katalogs der KBOB-Daten gibt es auch Queries die eigentlich keinen passenden Eintrag haben. Für diese wird der jeweils fachlich nächstliegende KBOB-Eintrag als korrekte Zuordnung gewertet.

Für die Bewertung werden die in Abschnitt 5.6.1 eingeführten Metriken verwendet.

Die Evaluation verwendet vier Testszenarien, um die Praxistauglichkeit des trainierten Modells abzuschätzen (die Erzeugung und genaue Beschreibung der degradierten Szenarien 2–4 erfolgt in Abschnitt 7.3):

1. **Normale Queries:** Alle Angaben soweit vorhanden vollständig und korrekt geschrieben.
2. **Queries mit Schreibfehlern:** Pro Query werden ein bis zwei Wörter mit Tippfehlern versehen.
3. **Queries mit fehlenden Angaben:** Pro Query wird ein semantisch relevantes Token entfernt.
4. **Kombination:** Schreibfehler und fehlende Angaben treten gleichzeitig auf.

Zusätzlich werden die Ergebnisse nach vier Dimensionen aufgeschlüsselt. Die *Materialfamilie* unterscheidet Gruppen wie Beton, Stahl, Holz, Kunststoff und Asphalt. Die *Herstellungsart* trennt zwischen Ortbeton, Fertigteilen und unbekannter Herstellung. Die *Query-Klasse* unterscheidet

¹Beispielsweise «Konstruktionsvollholz», «Massivholz Buche Eiche kammergetrocknet gehobelt», «Massivholz Buche Eiche kammergetrocknet rau», «Massivholz Buche Eiche luftgetrocknet rau», «Massivholz Fichte Tanne Lärche kammergetrocknet gehobelt», «Massivholz Fichte Tanne Lärche luftgetrocknet gehobelt» oder «Massivholz Fichte Tanne Lärche luftgetrocknet rau» für die Query «IfcMember RAFTER Holz»

eindeutige Queries mit genau einem korrekten Zielwert von mehrdeutigen Queries mit mehreren zulässigen KBOB-Einträgen. Schliesslich wird ausgewertet, ob die Angabe einer Stärkeklasse vorhanden ist oder fehlt.

7.2 Trefferqualität auf kontrollierten Testdaten

Dieser Abschnitt untersucht die Trefferqualität des feinabgestimmten Modells auf den 389 normalen Test-Queries. Zunächst wird das Gesamtergebnis vorgestellt, danach folgt eine Aufschlüsselung nach Materialfamilie sowie nach Herstellungsart und Query-Klasse. Abschliessend werden die verbleibenden Fehler systematisch analysiert.

7.2.1 Gesamtergebnis

Das feinabgestimmte Modell erreicht auf den 389 normalen Test-Queries eine Hit@1-Rate von 96,9 %. Das bedeutet: In knapp 97 von 100 Fällen steht der fachlich korrekte KBOB-Eintrag auf Rang 1 der Vorschlagsliste. Hit@10 liegt bei 99,0 %, der korrekte Eintrag befindet sich also in nahezu allen Fällen unter den ersten zehn Vorschlägen. MRR@10 beträgt 0,979, was einer mittleren Trefferposition nahe Rang 1 entspricht. Insgesamt treten 12 Fehler auf.

7.2.2 Analyse nach Materialfamilie

Tab. 7.1 zeigt die Ergebnisse der Evaluation nach Materialfamilie aufgeschlüsselt. Stahl, Bewehrung/Spannstahl und Holz erreichen jeweils 100 % Hit@1. Alle drei besitzen eindeutige sprachliche Merkmale in den Queries: die Stahlgüte S235 oder S355, die Bewehrungssorte B500B beziehungsweise die Holz-Festigkeitsklasse C24. Diese Angaben genügen dem Modell für eine korrekte Zuordnung. Ebenso fehlerfrei sind Aluminium, Elastomer, Schotter/Kies und Strassenoberbau-Materialien (hydraulisch gebundene und ungebundene Tragschichten, die auf Hartbeton oder Kies abgebildet werden).

Beton/Stahlbeton zeigt mit 96,6 % leicht tiefere Werte, weil der KBOB-Katalog mehrere semantisch nahe Betoneinträge enthält (Tiefbaubeton, Bohrpahlbeton, Magerbeton, Betonfertigteil u. a.).

Kunststoff erreicht 80,0 % Hit@1. Die drei Fehler betreffen PTFE-Gleitlager. Abdichtung weist mit 60,0 % die tiefste Trefferquote auf – hier fehlen geeignete Trainingsbeispiele und die sprachliche Nähe konkurrierender KBOB-Einträge erschwert die Unterscheidung. Asphalt/Bitumen verzeichnet einen Fehler durch sprachliche Nähe zum falschen Eintrag. Eine systematische Analyse aller 12 Fehler nach Fehlermustern folgt in Abschnitt 7.2.4.

Materialfamilie	Queries	Hit@1	Fehler
Stahl & Baustahl	120	100%	0
Beton & Stahlbeton	117	96.6%	4
Bewehrung & Spannstahl	59	100%	0
Holz	23	100%	0
Kunststoff	15	80.0%	3
Aluminium	10	100%	0
Übrige ²	9	77.8%	2
Asphalt/Bitumen	8	87.5%	1
Schotter/Kies/Erdbau	8	100%	0
Elastomer	6	100%	0
Abdichtung	5	60.0%	2
Stein/Mauerwerk	5	100%	0
Strassenoberbau (Tragschichten)	4	100%	0
Gesamt	389	96.9%	12

Tab. 7.1: Hit@1 nach Materialfamilie auf den 389 evaluierten Queries, eigene Tabelle.

7.2.3 Analyse nach Herstellungsart und Query-Klasse

Die Herstellungsart beeinflusst die Trefferquote. Fertigteile (precast, 8 Queries) erreichen 100 % Hit@1, weil die Materialangabe bei Fertigteilen spezifisch ist und wenig Varianz aufweist. Queries ohne Herstellungsangabe (310 Queries) erreichen 97,4 %. Ortbetonteile (insitu, 71 Queries) liegen bei 94,4 %. Der tiefere Wert erklärt sich dadurch, dass Ortbeton-Queries häufig Tiefbauelemente wie Bohrpfähle betreffen, bei denen das Modell zwischen verschiedenen Betonarten unterscheiden muss.

Eindeutige Queries, für die genau ein korrekter KBOB-Eintrag definiert ist, erreichen 97,5 % Hit@1 (237 Queries). Mehrdeutige Queries mit mehreren zulässigen Zielwerten erreichen 96,1 % (152 Queries). Die Differenz ist gering, weil das Modell auch bei mehrdeutigen Queries meist einen der zulässigen Einträge auf Rang 1 platziert.

7.2.4 Analyse der verbleibenden Fehler

Tab. 7.2 zeigt die 12 fehlerhaften Queries mit dem vorgeschlagenen und dem erwarteten KBOB-Eintrag sowie dem Rang, auf dem der korrekte Eintrag tatsächlich erscheint. Die Trennlinien in der Tabelle markieren die vier nachfolgend beschriebenen Fehlermuster.

²Die Kategorie «Übrige» umfasst 9 Queries, die keiner spezifischen Materialfamilie zugeordnet werden: drei Jetgrouting-Mörtel-Varianten (Mörtel, Zementmörtel, Spezialmörtel), zwei Verkleidungen (IfcCovering CLADDING mit Metall bzw. Faserzement), einen Feinmörtel-Belag (IfcCovering TOPPING), einen Gummipuffer (IfcBuildingElementPart VEHICLESTOP) sowie zwei identische Rühlwand-Fertigteil-Queries (IfcWall RETAININGWALL).

Query	Vorgeschlagen	Erwartet	Rang
IfcPile BORED Beton C20/25 insitu	Tiefgründung Rüttelstopfsäule	Bohrpfahlbeton	2
IfcPile BORED Stahlbeton C20/25 insitu	Tiefgründung Rüttelstopfsäule	Bohrpfahlbeton	2
IfcPile COHESION Beton C20/25 insitu	Tiefgründung Rüttelstopfsäule	Tiefbaubeton	2
IfcPile JETGROUTING Mörtel	Tiefgründung Vorgefertigter Betonpfahl	Baukleber Einbettmörtel mineralisch	3
IfcSlab FLOOR Beton NPK D C20/25 insitu	2K-Fliessbelag Epoxidharz	Tiefbaubeton	2
IfcBearing GUIDE PTFE	Kautschukdichtungsmasse	Polyethylen PE	23
IfcBearing GUIDE Polytetrafluoroethylene	Kautschukdichtungsmasse	Polyethylen PE	23
IfcBearing GUIDE Teflon	Kautschukdichtungsmasse	Polyethylen PE	21
IfcCovering MEMBRANE Flüssigkunststoff	Dichtungsbahn Gummi EPDM	2K-Fliessbelag Epoxidharz	42
IfcCovering MEMBRANE Abdichtung	Kautschukdichtungsmasse	Dichtungsbahn bituminös	2
IfcPavement FLEXIBLE Polymermodifiziertes Bitumen	Dichtungsbahn bituminös	Gussasphalt	2
IfcBuildingElementPart VEHICLESTOP Gummipuffer	Gussasphalt	Dichtungsbahn Gummi EPDM	2

Tab. 7.2: Die 12 fehlerhaften Queries des feinabgestimmten Modells auf kontrollierten Testdaten, gruppiert nach Fehlermuster, eigene Tabelle.

1. **Verwechslung ähnlicher Tiefbau-Einträge** (5 Fehler): vier der 12 Fehler betreffen Tiefgründungen. Die KBOB-Datenbank enthält mit «Tiefgründung Rüttelstopfsäule», «Bohrpfahlbeton» und «Tiefbaubeton» drei semantisch nahe Einträge. Das Modell ordnet die Pile-Queries der Rüttelstopfsäule zu, obwohl der vordefinierte Typ BORED auf Bohrpfähle hinweist. Rüttelstopfsäulen bestehen aus Kies, nicht aus Beton, weshalb diese Zuordnung fachlich falsch ist. Der fünfte Fehler betrifft eine Betonplatte (IfcSlab FLOOR) mit der Betonbezeichnung NPK D, die das Modell fälschlicherweise einem Epoxidharz-Belag statt Tiefbaubeton zuordnet. Der korrekte Eintrag steht in allen fünf Fällen auf Rang 2 oder 3.
2. **Fehlende Zuordnung für PTFE-Gleitlager** (3 Fehler): die drei IfcBearing-GUIDE-Queries mit PTFE, Polytetrafluoroethylene und Teflon werden als «Kautschukdichtungsmasse» statt «Polyethylen PE» klassifiziert. Der korrekte Eintrag liegt auf Rang 21–23. PTFE ist im Trainingsdatensatz zwar vertreten, doch das Modell assoziiert die Lager-Queries mit Dichtungsmaterialien statt mit PE-Kunststoffen.
3. **Seltene Materialbezeichnung ohne Trainingsbeispiel** (1 Fehler): «Flüssigkunststoff» ist im Trainingsdatensatz nicht vertreten. Das Modell findet den korrekten Eintrag erst auf Rang 42. Dieser Fall lässt sich durch gezielte Ergänzung im Trainingsdatensatz beheben.
4. **Sprachliche Nähe zwischen falschen und richtigen Einträgen** (3 Fehler): bei «Abdichtung», «Polymermodifiziertes Bitumen» und «Gummipuffer» unterscheiden sich der vorgeschlagene und der korrekte Eintrag nur in einer Nuance. Beispielsweise verwechselt das Modell «Kautschukdichtungsmasse» mit «Dichtungsbahn bituminös», weil beide Einträge eine bituminöse Abdichtung beschreiben. Der korrekte Eintrag liegt jeweils auf Rang 2.

Acht der 12 Fehler haben den korrekten Eintrag auf Rang 2 oder 3. Im Praxisbetrieb mit dem *Human-in-the-loop*-Ansatz kann die nutzende Person diese Fehler beim Durchsehen der Vorschlagsliste schnell erkennen und korrigieren. Die vier verbleibenden Fehler (PTFE-Lager und Flüssigkunststoff) mit Rängen über 10 erfordern entweder zusätzliche oder korrigierte Trainingsdaten oder eine Regelergänzung.

7.3 Robustheit gegenüber fehlerhaften Eingaben

Reale IFC-Modelle enthalten häufig Schreibfehler oder unvollständige Angaben. Um die Robustheit des Modells unter diesen Bedingungen zu prüfen, wurden die 389 Test-Queries systematisch degradiert. Zunächst wird die Erzeugung der degradierten Queries beschrieben. Danach werden die Ergebnisse für drei Szenarien einzeln vorgestellt: Schreibfehler, fehlende Angaben sowie die Kombination beider Effekte.

7.3.1 Erzeugung der degradierten Queries

Für das Szenario *Schreibfehler* werden pro Query ein bis zwei Wörter mit jeweils einem Tippfehler versehen. Die simulierten Fehler entsprechen typischen Eingabefehlern: Buchstabendreher («Cylidnrical» statt «Cylindrical»), fehlende Zeichen («initu» statt «insitu»), verdoppelte Buchstaben («BBeton» statt «Beton») oder benachbarte Tasten auf der Tastatur («Asühalt» statt «Asphalt»). Die IFC-Entität selbst bleibt unverändert, weil sie in der Praxis maschinell erzeugt wird und nicht von Tippfehlern betroffen ist.

Für das Szenario *fehlende Angaben* wird pro Query genau ein semantisch relevantes Token entfernt. Entfernt werden können die Materialbezeichnung, der vordefinierte Typ, die Stärkeklasse oder die Herstellungsart.

Das dritte Szenario *Kombination* wendet beide Degradationen gleichzeitig an: Zuerst wird ein Token entfernt, danach werden ein bis zwei Tippfehler in die verbleibenden Tokens eingefügt.

7.3.2 Queries mit Schreibfehlern

Schreibfehler reduzieren Hit@1 um 9 Prozentpunkte auf 87,9 % (siehe Tab. 7.3). Das Modell toleriert zeichenbasierte Störungen vergleichsweise gut, weil es auf Subwort-Ebene arbeitet. Ein Tippfehler verändert oft nur ein einzelnes Subwort-Token (also ein Wortfragment, das das Modell intern als Einheit verarbeitet), während die übrigen Tokens der Query weiterhin korrekt auf den richtigen KBOB-Eintrag hinweisen. Ein Beispiel: Die Query «IfcBeam PIERCAP BBeton NPK G C30/37 insitu» enthält einen verdoppelten Buchstaben in «BBeton». Die Stärkeklasse C30/37 und die Herstellungsart «insitu» bleiben lesbar, sodass das Modell trotzdem den korrekten Eintrag auf Rang 1 platziert.

Von den 47 Fehlern entfallen 36 auf mehrdeutige Queries, bei denen bereits ohne Tippfehler mehrere KBOB-Einträge in Frage kommen. Eindeutige Queries behalten mit 95,4 % Hit@1 eine hohe Treffsicherheit. Queries mit einer vorhandenen Stärkeklasse (z. B. C30/37 oder S235) erreichen 95,1 % Hit@1, weil die alphanumerische Kennung auch bei Tippfehlern in anderen Tokens als starker Hinweis erhalten bleibt.

Szenario	Hit@1	Hit@10	MRR@10	Fehler
Normale Queries	96.9%	99.0%	0.979	12
Nur Schreibfehler	87.9%	94.3%	0.904	47
Nur fehlende Angaben	74.8%	92.3%	0.798	98
Fehlende Angaben & Schreibfehler	67.6%	87.7%	0.734	126

Tab. 7.3: Vergleich der vier Evaluationsszenarien auf 389 Test-Queries, eigene Tabelle.

7.3.3 Queries mit fehlenden Angaben

Fehlende Angaben wirken deutlich stärker: Hit@1 sinkt um 22 Prozentpunkte auf 74,8 %. Der Grund liegt darin, dass ein fehlendes Token dem Modell eine zentrale Information vollständig entzieht. Wenn beispielsweise aus «IfcBearing CYLINDRICAL S235JR» die Stahlgüte S235JR entfernt wird, muss das Modell allein aus der Entität und dem vordefinierten Typ auf das Material schliessen, was bei generischen Elementen oft nicht eindeutig möglich ist.

Besonders betroffen sind Queries ohne Stärkeklasse: Diese erreichen nur 68,4 % Hit@1, während Queries mit Stärkeklasse bei 93,1 % verbleiben. Die Stärkeklasse (z. B. C30/37 für Beton, S355 für Stahl, B500B für Bewehrungsstahl) ist das wichtigste Unterscheidungsmerkmal für das Modell. Fehlt sie, steigt die Verwechslungsgefahr zwischen ähnlichen Materialfamilien erheblich.

Am stärksten betroffen sind Queries, deren Materialbezeichnung (z. B. «Beton», «Stahl», «Aluminium») durch die Token-Entfernung verloren geht. Bei 131 der 389 degradierten Queries fehlt nach der Entfernung das Material (z. B. das Wort «Aluminium» aus «IfcRailing HANDRAIL Aluminium») und Hit@1 sinkt auf 38,2 %. Das Modell muss dann allein aus Entität und vordefiniertem Typ auf das Material schliessen, was bei einem generischen Element wie einem Geländer nicht möglich ist. Selbst eine vorhandene Stärkeklasse wie C30/37 kann den Verlust des Materialworts nicht vollständig kompensieren: 22 dieser 131 Queries enthalten zwar noch eine Stärkeklasse, werden aber dennoch häufig falsch zugeordnet, weil C30/37 zwar auf die Materialfamilie Beton hinweist, ohne die weiteren fehlenden Kontextinformationen aber nicht zwischen verwandten Einträgen wie «Tiefbaubeton», «Betonfertigteil normalfest» oder «Bohrpfahlbeton» unterscheiden kann.

7.3.4 Queries mit fehlenden Angaben und Schreibfehlern

Das kombinierte Szenario stellt den härtesten Test dar. Hit@1 fällt auf 67,6 % (126 Fehler), MRR@10 auf 0,734 und Hit@10 auf 87,7 %. Die Kombination wirkt stärker als die Summe der Einzeleffekte, weil ein fehlendes Token die Redundanz der Query reduziert (vgl. Abschnitt 7.3, Szenario fehlende Angaben) und ein Tippfehler die verbleibenden Hinweise zusätzlich verzerrt. Ein Beispiel: Aus «IfcMember STUD Holz» wird «Holz» entfernt und STUD zu «STTUD» verfälscht. Dem Modell bleibt kein verwertbarer Hinweis auf das Material, und es ordnet die Query einem Stahlprofil zu (korrekter Eintrag auf Rang 15). Insgesamt betreffen 105 der 126 Fehler Queries, die nach der Token-Entfernung nur noch aus der IFC-Entität und einem einzigen weiteren, oft fehlerhaft geschriebenen Token bestehen.

7.3.4.1 Aufschlüsselung nach Einflussfaktoren

Die in Abschnitt 7.3 beschriebenen Schutzfaktoren wirken auch hier, allerdings abgeschwächt. Die Stärkeklasse bleibt der wichtigste Prädiktor: 89,3 % Hit@1 mit Stärkeklasse (84 Queries) gegenüber 61,6 % ohne (305 Queries). Bewehrung und Spannstahl bleiben bei 100 % (57 Queries), weil die Kennung B500B selbst als letztes verbleibendes Token eindeutig genug ist.

Eindeutige Queries halten sich bei 78,1 % Hit@1 (237 Queries, 52 Fehler). Mehrdeutige Queries fallen auf 51,3 % (152 Queries, 74 Fehler). Ein Beispiel: «IfcMember Hola» (ein Tippfehler in «Holz», gleichzeitig fehlt PredefinedType) soll auf Konstruktionsvollholz abgebildet werden, das Modell wählt stattdessen einen Beton-Eintrag, da es im Training mit deutlich mehr Beton als Holz Queries trainiert wurde.

7.3.4.2 Rangverteilung der Fehler

Auch im kombinierten Szenario liegt der korrekte Eintrag bei der Mehrheit der Fehler noch in der Vorschlagsliste. Bei 54 der 126 Fehler steht er auf Rang 2 bis 5, bei weiteren 24 auf Rang 6 bis 10. Zusammen sind das 78 Fehler (62 %), bei denen der richtige Eintrag in den ersten zehn Vorschlägen erscheint. Bei 34 Fehlern liegt er auf Rang 11 bis 30. Nur 14 Fehler betreffen Ränge über 30, vorwiegend die bereits im Normalszenario problematischen PTFE-Lager und Queries, bei denen das Materialwort vollständig verloren ging. Insgesamt bleibt Hit@10 selbst im schlechtesten Szenario bei 87,7 %: In knapp 9 von 10 Fällen erscheint der korrekte Eintrag unter den ersten zehn Vorschlägen.

7.4 Vergleich mit dem Basismodell und weiteren Modellen

Um die Wirkung des Fine-Tunings einzuordnen, wurde das feinabgestimmte Modell auf denselben 389 Test-Queries mit 13 weiteren Embedding-Modellen und einem generativen Sprachmodell (GPT-5.5) verglichen. Tab. 7.4 zeigt die Ergebnisse. Im Anhang A.12 sind die Resultate von allen Embedding-Modellen als Übersicht dargestellt.

Modell	Hit@1	Hit@10	Fehler
Hygros-LCA/bge-m3-ifc-kbob-finetuned	96.9%	99.0%	12
OpenAI/GPT-5.5 Thinking	86.6%	-	52
sentence-transformers/LaBSE	35.2%	54.0%	252
BAAI/bge-m3 (Basismodell)	32.7%	71.2%	262
sentence-transformers/multilingual-e5-large	32.7%	53.7%	262
google/EmbeddingGemma-300M	23.7%	72.0%	297

Tab. 7.4: Modellvergleich auf 389 Test-Queries. Die obere Gruppe zeigt aufgabenspezifische Ansätze, die untere die besten generischen Embedding-Modelle. Mit GPT-5.5 wurde nur die beste Zuordnung bestimmt, weshalb Hit@10 fehlt, eigene Tabelle.

Das feinabgestimmte Modell übertrifft alle Vergleichsmodelle deutlich. Die Feinabstimmung steigert die Trefferquote um den Faktor 3 gegenüber dem Basismodell BGE-M3 (von 32,7 % auf 96,9 %). Die übrigen generischen Modelle, darunter drei BERT-Varianten (1–3 % Hit@1) und zwei BIM-spezifische Modelle (0,5–5,7 %), schneiden noch schwächer ab.

Der Grund für den grossen Unterschied: Die generischen Modelle kennen den Zusammenhang zwischen IFC-Queries und KBOB-Bezeichnungen nicht. Ein Beispiel: Die Query «IfcPile BORED Beton C25/30» enthält für einen Bauingenieur eindeutige Hinweise auf Bohrpfahlbeton. Für ein generisches Modell sind «IfcPile» und «BORED» unbekannte Fachwörter. Es erkennt allenfalls «Beton», kann aber nicht zwischen Tiefbaubeton, Bohrpfahlbeton oder Magerbeton unterscheiden. Das Fine-Tuning lehrt dem Modell diese Verknüpfungen.

Zusätzlich wurde GPT-5.5 Thinking von OpenAI mit dem vollständigen KBOB-Katalog als Kontext getestet. Es ordnet 337 von 389 Queries korrekt zu (86,6 % Hit@1), liegt also 10 Prozentpunkte unter dem feinabgestimmten Modell. Die 52 Fehler betreffen drei systematische Muster: Verwechslung von Stahlprofil und Stahlblech bei IfcPlate-Queries (24 Fehler), Zuordnung von Spannkabel-Hüllrohren zu Armierungsstahl statt zu Stahlblech/Stahlprofil oder Kunststoff (12 Fehler) sowie Verwechslung von Tiefbaubeton mit Hartbeton bei Fahrbahnbelägen (7 Fehler). GPT-5.5 versteht die allgemeine Bedeutung der Queries, unterscheidet aber nicht zuverlässig zwischen eng verwandten KBOB-Einträgen.

Für den Praxiseinsatz bleibt das feinabgestimmte Embedding-Modell überlegen: Es erreicht die höchste Trefferquote, arbeitet lokal ohne Internetverbindung und verarbeitet alle 389 Queries in wenigen Sekunden. GPT-5.5 benötigte für dieselbe Aufgabe über fünf Minuten (gemessen über Microsoft 365 Copilot) und verursacht laufende Kosten.

7.5 Anwendung auf reale Infrastrukturprojekte

Die bisherigen Ergebnisse basieren auf kontrollierten Testdaten. In der Praxis stammen die Queries jedoch aus realen IFC-Modellen, deren Modellierungstiefe und Attributbefüllung je nach Autorenwerkzeug und Projektbeteiligten variieren. Um die Praxistauglichkeit zu prüfen, wurde der Prototyp auf vier reale IFC-Modelle aus Infrastrukturprojekten (Brückenbauwerke) angewendet.

Die Modelle wurden direkt im Dashboard verarbeitet und die Zuordnung kontrolliert. Viele IFC-Elemente erzeugen identische Queries, weil sie dieselbe Kombination von Eigenschaften haben (vgl. Abschnitt 5.3.1, Abschnitt 5.3.2 und Abschnitt 6.2.1). Im Dashboard werden diese automatisch gruppiert (vgl. Abschnitt 5.4.3).

Tab. 7.5 zeigt die Ergebnisse. Die Genauigkeit liegt zwischen 60 % und 89 %, abhängig von der Modellierungstiefe des jeweiligen Modells. Bei Projekt 3 und 4 sind einzelne Queries nicht eindeutig einem KBOB-Eintrag zuzuordnen, weshalb die Genauigkeit als Bereich angegeben wird.

Projekt	Elemente	Queries	korrekt	falsch	Genauigkeit
1	1781	9	8	1	89%
2	680	9	7	2	78%
3	307	15	12–13	2–3	80–86 %
4	45	5	3–4	1–2	60–80 %

Tab. 7.5: Trefferquote des feinabgestimmten Modells auf vier realen Infrastrukturprojekten, eigene Tabelle.

Die Fehler lassen sich auf vier Muster zurückführen:

1. **Generische Entität** (3 Projekte): Elastomerlager sind nicht als `IfcBearing` modelliert, sondern als `IfcBuildingElementProxy`. Das trainierte Modell kennt `IfcBuildingElementProxy` nicht im Zusammenhang mit Lagermaterialien und schlägt Polystyrol statt Polyurethan vor (korrekter Eintrag auf Rang 10 bis 15).
2. **Unspezifische Materialbezeichnung** (2 Projekte): Eine Sauberkeitsschicht trägt nur «unbewehrter Beton» ohne Stärkeklasse. Das Modell wählt «Betonfertigteile normalfest» statt Magerbeton, weil das Unterscheidungsmerkmal fehlt (korrekter Eintrag auf Rang 16).
3. **Fehlende geometrische Angaben** (1 Projekt): Die Bohrpfähle enthalten keinen Durchmesser (er kann nicht berechnet werden, da die Länge fehlt), den der KBOB-Katalog zur Unterscheidung benötigt. Das Modell ordnet die Query dem allgemeineren «Bohrpfahlbeton» zu statt dem spezifischeren Tiefgründungseintrag (korrekter Eintrag auf Rang 4).
4. **Profil vs. Blech** (1 Projekt): Eine Aluminium-Sichtschutzwand wird als «Aluminiumprofil» statt «Aluminiumblech» klassifiziert, weil die Wand als `IfcRailing` modelliert ist und Geländer üblicherweise aus Profilen bestehen. Ohne Geometrieinformationen ist die Unterscheidung nicht möglich (korrekter Eintrag auf Rang 2).

Bei den meisten Fehlern steht der korrekte Eintrag auf Rang 2 bis 4, sodass der *Human-in-the-loop*-Ansatz diese auffängt.

Neben der Prüfung der Zuordnungsqualität wurde auch die nachgelagerte Berechnung validiert. Die Ökobilanzwerte einer Stichprobe aus der selbst erstellten IFC-Datei wurden in ECO₂nstruct (Infra Suisse, 2023) nachgerechnet. Die Gesamtabweichung beträgt rund 6 % und ist auf unterschiedliche Kennwerte pro Bezugseinheit zurückzuführen (siehe Anhang A.13).

7.6 Zusammenfassung der Ergebnisse

Das feinabgestimmte Modell ordnet auf kontrollierten Testdaten in 97 von 100 Fällen den fachlich korrekten KBOB-Eintrag als ersten Vorschlag zu. Schreibfehler in den Eingabedaten senken die Trefferquote um rund 9 Prozentpunkte, fehlende Angaben um rund 22 Prozentpunkte. Auf realen Infrastrukturprojekten liegt die Genauigkeit zwischen 60 % und 89 %, abhängig von der Modellierungstiefe und der Nutzung spezialisierter IFC-Entitäten. Das Fine-Tuning verdreifacht die

Trefferquote gegenüber dem Basismodell. Selbst unter ungünstigen Bedingungen erscheint der korrekte Eintrag in neun von zehn Fällen noch unter den ersten zehn Vorschlägen, was den *Human-in-the-loop*-Ansatz tragfähig macht. Die verbleibenden Fehler folgen systematischen Mustern und lassen sich durch gezielte Ergänzung der Trainingsdaten oder durch Regelanpassungen beheben.

Das folgende Kapitel diskutiert diese Fehlermuster im Detail, ordnet die Ergebnisse in den Stand der Technik ein und leitet daraus Empfehlungen für den Praxiseinsatz sowie weiteren Forschungsbedarf ab.

8 Diskussion und Ausblick

Dieses Kapitel ordnet die Ergebnisse ein und zeigt ihre Grenzen auf. Abschnitt 8.1 analysiert die aufgetretenen Fehler und identifiziert vier wiederkehrende Muster. Abschnitt 8.2 benennt die Limitationen der Arbeit, bevor Abschnitt 8.3 die Konsequenzen für Forschung und Praxis ableitet. Abschnitt 8.4 fasst die zentralen Erkenntnisse zusammen und Abschnitt 8.5 zeigt Ansatzpunkte für Weiterentwicklungen auf.

8.1 Fehleranalyse und Diskussion

Die in den Abschnitten 7.2.4, 7.3 und 7.5 dokumentierten Fehler treten in unterschiedlichen Szenarien auf, folgen aber übergreifend vier wiederkehrenden Mustern. Die folgende Analyse ordnet sämtliche Fehler diesen Kategorien zu und diskutiert ihre Ursachen.

- 1. Fehlende oder widersprüchliche Trainingsdaten:** Einige Fehler entstehen, weil das Modell für bestimmte Materialien keine oder falsche Beispiele im Trainingsdatensatz kennt. Der deutlichste Fall sind die drei PTFE-Gleitlager-Queries (Rang 21 bis 23). Der KBOB-Katalog enthält keinen eigenen Eintrag für PTFE. Als fachlich nächste Entsprechung wurde «Polyethylen PE» als Referenz definiert. Die Trainingsdaten enthalten für dieselben Queries jedoch «Kautschukdichtungsmasse» als positives Beispiel, also als Material, das das Modell als korrekt ansehen soll. Es handelt sich also um einen Widerspruch zwischen Trainingsdaten und Referenzdaten. Ebenso fehlt für «Flüssigkunststoff» jedes Trainingsbeispiel, weshalb der korrekte Eintrag erst auf Rang 42 erscheint. Solche Fehler lassen sich ohne grossen Aufwand beheben: Ein erneutes Training mit korrigierten oder ergänzten Beispielpaaren genügt, um die Zuordnung richtigzustellen.
- 2. Mehrdeutigkeit durch fehlende Kontextinformation:** Fehlt in der Query eine zentrale Angabe, kann das Modell nicht zwischen ähnlichen Materialfamilien unterscheiden. Die Robustheitstests zeigen diesen Effekt quantitativ: Queries mit einer Stärkeklasse (C30/37, S235, B500B) erreichen selbst im kombinierten Degradationsszenario noch 89,3 % Hit@1, Queries ohne Stärkeklasse nur 61,6 %. Wird das Materialwort vollständig entfernt, sinkt Hit@1 auf 38,2 %. Ein konkretes Beispiel aus den realen Projekten: Eine Sauberkeitsschicht trägt nur die Bezeichnung «unbewehrter Beton» ohne Stärkeklasse. Das Modell wählt «Betonfertigteil normalfest» statt «Magerbeton», weil das Unterscheidungsmerkmal fehlt. Der Prototyp kann fehlende Informationen grundsätzlich nicht kompensieren. Er nutzt die vorhandenen Angaben zuverlässig, aber bei zu wenig Kontext bleibt die Zuordnung zwangsläufig unsicher. Um das Problem zu beheben sind im Anhang A.10 Modellierungsrichtlinien definiert.
- 3. Abweichende IFC-Modellierung:** In drei von vier untersuchten realen Projekten sind Elastomerlager nicht als spezialisierte Entität modelliert, sondern als generisches Ersatzelement. Das Modell kennt dieses generische Element nicht im Zusammenhang mit Lagermaterialien, weil die

Trainingsdaten auf die fachlich korrekte Entität ausgerichtet sind. Es schlägt stattdessen Polystyrol statt Polyurethan vor (korrekter Eintrag auf Rang 10 bis 15). Der wirksamste Hebel liegt hier bei Modellierungsrichtlinien, etwa im BIM-Abwicklungsplan eines Projekts, die den Einsatz spezialisierter Entitäten vorschreiben. Zusätzliche Trainingsdaten mit generischen Ersatzelementen können das Problem teilweise abfangen, lösen es aber nicht grundsätzlich, weil die Modellierungspraxis zwischen Autorenwerkzeugen stark variiert.

4. **Sprachliche Nähe im KBOB-Katalog:** Bei drei der 12 Fehler aus den Test-Daten und einem Fehler aus den realen Projekten unterscheiden sich der vorgeschlagene und der korrekte Eintrag nur in einer Nuance. Beispielsweise verwechselt das Modell «Kautschukdichtungsmasse» mit «Dichtungsbahn bituminös», weil beide Einträge Abdichtungsmaterialien beschreiben. Ebenso wird ein als IfcRailing modelliertes Aluminium-Element als «Aluminiumprofil» statt «Aluminiumblech» klassifiziert, weil Geländer üblicherweise aus Profilen bestehen. Das Modell schliesst aus der Entität auf die wahrscheinlichste Produktform, ohne die tatsächliche Geometrie zu kennen. In diesen Fällen erkennt das Modell die richtige Materialfamilie, trifft aber nicht die exakte Variante. Der korrekte Eintrag steht jeweils auf Rang 2. Eine zusätzliche Re-Ranking-Stufe mit einem Cross-Encoder wäre eine mögliche Lösung.

Übergreifend zeigen die Ergebnisse, dass die Trefferquote mit der Vollständigkeit der Materialattribute im IFC-Modell steigt. Die Stärkekategorie ist der stärkste einzelne Vorhersagefaktor. Für den *Human-in-the-loop*-Ansatz bleibt der Prototyp auch bei Fehlern praxistauglich, weil Hit@10 selbst im härtesten Testszenario bei 87,7% liegt und acht der zwölf Fehler den korrekten Eintrag auf Rang 2 oder 3 aufweisen.

8.2 Limitationen der Arbeit

Die vorangehende Fehleranalyse zeigt, wo der Ansatz an seine Grenzen stösst. Die folgenden fünf Limitationen gehen über einzelne Fehlermuster hinaus und schränken die Generalisierbarkeit der Ergebnisse ein.

1. **Abhängigkeit von der IFC-Modellqualität:** Der Prototyp kann nur verarbeiten, was im Modell an Attributen vorhanden ist. Die Robustheitstests belegen eine starke Abhängigkeit von der Stärkekategorie und dem Materialwort (vgl. Abschnitt 7.3). Für Modelle mit systematisch unvollständiger Attributbefüllung bietet der Ansatz keine Kompensationsmöglichkeit.
2. **Begrenzter und länderspezifischer Zielkatalog:** Die KBOB-Ökobilanzdaten decken ausschliesslich den Schweizer Markt ab und umfassen für den Infrastrukturkontext rund 110 Einträge. Materialien ohne eigenen Katalogeintrag, etwa PTFE oder Flüssigkunststoff, können nur über einen fachlich ähnlichen Stellvertreter zugeordnet werden (vgl. Abschnitt 7.2.4). Die Ergebnisse sind nicht direkt auf andere Datenbanken wie die deutsche ÖKOBAUDAT oder EPD-Kataloge übertragbar, da sich Bezeichnungen, Granularität und Bezugseinheiten unterscheiden.

3. **Domänen- und Werkzeugspezifität:** Trainings- und Testdaten stammen ausschliesslich aus dem Brückenbau. Die 389 Test-Queries wurden grösstenteils mit Tekla Structures erzeugt. Bei den vier realen Projekten handelt es sich ebenfalls um Brückenbauwerke. Eine Aussage zur Trefferquote bei anderen Infrastrukturtypen lässt sich daher nicht treffen. Beim Einsatz anderer Autorenwerkzeuge bleibt zu überprüfen, welche Eigenschaften in die IFC-Datei exportiert werden.
4. **Rein textbasiertes Matching ohne Geometrie:** Der Bi-Encoder (das verwendete Sprachmodell für die Kandidatensuche) vergleicht textuell repräsentierte Informationen. Dazu gehört auch der für `IfcPile` aus Geometriedaten abgeleitete Durchmesser. Andere geometrische Merkmale wie die Produktform (Profil, Blech, Rohr) oder weitere Dimensionen fliessen bisher nicht in die Kandidatensuche ein. Dadurch bleibt der Ansatz fehleranfällig, wenn die KBOB-Zuordnung eines Materials von seiner Form abhängt. Das zeigt sich im Beispiel Aluminiumprofil statt Aluminiumblech (vgl. Abschnitt 7.5).
5. **Kein paarweises Reranking im Standardbetrieb:** Die Matching-Pipeline verwendet einen Bi-Encoder, der Query und Katalogeinträge getrennt codiert. Ein Cross-Encoder, der Query und Kandidat gemeinsam bewertet, ist konzeptionell vorgesehen, aber nicht umgesetzt (vgl. Abschnitt 5.4.2). Da viele Fehler den korrekten Eintrag knapp hinter Rang 1 aufweisen, könnte ein trainierter Cross-Encoder diese Fälle korrigieren.

Alle fünf Limitationen lassen sich durch gezielte Erweiterungen adressieren, wie Abschnitt 8.5 zeigt.

8.3 Implikationen für Forschung und Praxis

Aus den Ergebnissen werden zunächst Handlungsempfehlungen für die Planungspraxis abgeleitet, anschliessend offene Forschungsfragen formuliert.

8.3.1 Implikationen für die Praxis

Der grösste Hebel für die Zuordnungsqualität liegt nicht im Algorithmus, sondern in der Modellqualität. Modellierungsrichtlinien sollten vollständige Materialbezeichnungen vorschreiben, also mit Stärkeklasse (z. B. C30/37), spezialisierter Entität (z. B. `IfcBearing` statt `IfcBuildingElementProxy`) und einem eindeutigen Materialwort. Eine Modellierungsrichtlinie ist im Anhang A.10 aufgeführt. Trotz der Abhängigkeit von der Vollständigkeit der Informationen bleibt der Ansatz praxistauglich: Die Kandidatensuche läuft lokal in wenigen Sekunden, ohne Cloud-Anbindung oder laufende Kosten, und vertiefte KBOB-Kenntnisse sind nicht nötig, weil der korrekte Eintrag fast immer unter den ersten zehn Vorschlägen erscheint.

Neben der Modellqualität bestehen zwei infrastrukturelle Hürden auf Seiten der KBOB-Daten. Erstens stehen diese nur als Excel-Tabelle bereit. Eine maschinenlesbare Bereitstellung würde die Integration in automatisierte Workflows erleichtern. Zweitens verwenden die Katalogeinträge heterogene

Bezugseinheiten (m, m², kg). Für eine maschinelle Bilanzierung wäre eine volumenbezogene Einheit vorteilhaft, da sich diese Grösse direkt aus dem IFC-Modell auslesen lässt.

8.3.2 Implikationen für die Forschung

Der gewählte Ansatz, ein feinabgestimmtes Sprachmodell für die semantische Kandidatensuche, ist nicht an die KBOB-Datenbank gebunden. Die Architektur lässt sich auf andere Ökobilanzdatenbanken übertragen, etwa die deutsche ÖKOBAUDAT oder projektspezifische EPD-Kataloge. Voraussetzung ist ein angepasster Zielkatalog und ein entsprechender Trainingsdatensatz. Die Forschungsfrage dabei ist, ob die entwickelten Trainingsmethoden, insbesondere die harten Negativbeispiele und die batchweise Kollisionsvermeidung, auch bei anderen Katalogstrukturen wirksam bleiben.

Der Vergleich mit GPT-5.5 zeigt komplementäre Stärken: Das feinabgestimmte Modell erreicht 96,9 % durch Domänenspezialisierung, das grosse Sprachmodell ohne Anpassung 86,6 %. Für seltene Materialien ohne Trainingsbeispiele könnte ein grosses Sprachmodell als Rückfallebene dienen, das aus allgemeinem Weltwissen eine plausible Zuordnung ableitet. Die Kombination beider Verfahren stellt eine vielversprechende Forschungsrichtung dar.

8.4 Zusammenfassung der wichtigsten Erkenntnisse

Der entwickelte Prototyp beantwortet die übergeordnete Fragestellung dieser Arbeit: Ein feinabgestimmtes Sprachmodell ordnet auf kontrollierten Testdaten in 96,9 % der Fälle den korrekten KBOB-Eintrag auf Rang 1 zu. Auf vier realen Brückenbauprojekten bestätigt sich die Praxistauglichkeit, wenngleich die Trefferquote dort erwartungsgemäss tiefer liegt (60 bis 89 %).

Die KBOB-Ökobilanzdaten erweisen sich als geeigneter Schweizer Zielkatalog, weisen aber Lücken bei Spezialmaterialien auf (vgl. Abschnitt 8.1). Die Teilfragen zur Künstlichen Intelligenz werden durch drei Befunde beantwortet: (1) Die Stärkeklasse ist der wirksamste Kontextfaktor und schützt die Zuordnung selbst unter gleichzeitigen Schreibfehlern. (2) Domänenspezifisches Fine-Tuning verdreifacht die Trefferquote gegenüber generischen Modellen und übertrifft GPT-5.5 um 10 Prozentpunkte. (3) Ohne Materialwort sinkt die Trefferquote auf 38,2 %.

Die übergreifende Erkenntnis: Der limitierende Faktor ist nicht der Algorithmus, sondern die Vollständigkeit der Attribute im IFC-Modell.

8.5 Ausblick auf zukünftige Entwicklungen

Die folgenden Abschnitte gliedern mögliche Weiterentwicklungen in drei Zeithorizonte.

8.5.1 Kurzfristige Verbesserungen

Die wirksamste Massnahme ist die Korrektur und Ergänzung der Trainingsdaten: Widersprüchliche Beispielpaare (PTFE-Gleitlager) bereinigen, fehlende Materialien (Flüssigkunststoff) ergänzen und Trainingsbeispiele mit generischen Ersatzelementen (`IfcBuildingElementProxy`) hinzufügen. Zusätzlich entsteht im laufenden Betrieb ein natürlicher Verbesserungszyklus, weil jede bestätigte oder korrigierte Zuordnung als neues Trainingspaar dient.

8.5.2 Mittelfristige Erweiterungen

Zwei architektonische Erweiterungen adressieren die verbleibenden strukturellen Schwächen. Erstens könnte ein trainierter Cross-Encoder als zusätzliche Bewertungsstufe die Kandidatenliste nachsortieren und so die knappen Rang-2/3-Fehler korrigieren (vgl. Abschnitt 5.4.2). Zweitens könnte die Integration geometrischer Informationen (Produktform, Dimensionen) aus dem IFC-Modell die Unterscheidung zwischen ähnlichen Katalogeinträgen wie Aluminiumprofil und Aluminiumblech ermöglichen.

Auf der Datenebene wäre ein Wechsel vom JSONL-Format zu einer Datenbank sinnvoll, wenn der Prototyp im Projektalltag eingesetzt wird. Damit liessen sich Versionierungen konsistenter verwalten.

8.5.3 Langfristige Forschungsrichtungen

Der Transfer auf andere Infrastrukturdomänen (Tunnel-, Strassenbau) erfordert jeweils angepasste Trainingsdatensätze. Ob die entwickelten Trainingsmethoden (harte Negativbeispiele, batchweise Kollisionsvermeidung) dabei gleich wirksam bleiben, muss geprüft werden. Denkbar ist auch ein modularer Ansatz mit domänen- oder sprachspezifischen Modellen, da das multilinguale Basismodell bereits gemischte deutsch-englische Queries verarbeitet.

Eine weitere vielversprechende Richtung ist die Kombination mit LLMs als Rückfallebene für seltene Materialien ohne Trainingsbeispiele. Ebenso erfordert der Transfer auf andere Ökobilanzdatenbanken (ÖKOBAUDAT, EPDs) entsprechend angepasste Trainingsdaten.

Literaturverzeichnis

- Ajallouda, L., Saissi, M. H. und Zellou, A., 2025. «Embedding Models: A Comprehensive Review with Task-Oriented Assessment». In: *International Journal of Advanced Computer Science and Applications* 16.10.
- BAFU, 2021. *Ökofaktoren Schweiz 2021 gemäss der Methode der ökologischen Knappheit*. Umwelt-Wissen Nr. 2121. Bern: BAFU. URL: <http://www.bafu.admin.ch/uw-2121-d>.
- BAFU, 2025. *Langfristige Klimastrategie 2050*. URL: <https://www.bafu.admin.ch/bafu/de/home/themen/thema-klima/klimawandel-stoppen-und-folgen-meistern/klimapolitik/ziele-und-strategien/langfristige-klimastrategie-2050.html> (Stand 20.09.2025).
- Belsky, M., Sacks, R. und Brilakis, I., 2013. «A Framework for Semantic Enrichment of IFC Building Models». In: *Proceedings of the 30th CIB W78 International Conference*. Beijing, S. 514–523. ISBN: 2706-6568. URL: <https://itc.scix.net/pdfs/w78-2013-paper-60.pdf>.
- Bharti, P., Hippolyte, J.-L., Chrubasik, M., Isci, S., Butean, A. und Worall, G., 2025. «Semi-automatic mapping with SBERT for ontology-based interoperability in construction data systems». In: *MATEC Web of Conferences* 413. Hrsg. von Yang, Q., Wang, Q., Charki, A., Livina, V. und Sun, X., S. 03001. ISSN: 2261-236X. DOI: [10.1051/mateconf/202541303001](https://doi.org/10.1051/mateconf/202541303001).
- Bleiker, S., Pohl, T., Stürzinger, T. und Osterwalder, D., 2025. *Grundlagen für Referenz- und Zielwerte für graue Emissionen von Infrastrukturbauten*. Hombrechtikon: Umtec Technologie AG. URL: https://www.utechag.ch/fileadmin/user_upload/Bericht_Referenz-_und_Zielwerte_Infrastrukturbau.pdf.
- Borrmann, A., König, M., Koch, C. und Beetz, J., Hrsg. 2021. *Building Information Modeling: Technologische Grundlagen und industrielle Praxis*. VDI-Buch. Wiesbaden: Springer Fachmedien Wiesbaden. ISBN: 978-3-658-33361-4. DOI: [10.1007/978-3-658-33361-4](https://doi.org/10.1007/978-3-658-33361-4).
- Buchheister, D. J., Demory, C.-H., Pohl, T. und Osterwalder, D., 2025. *Anwendung der Ökobilanzmethode für die Infrastruktur von Nationalstrassen*. 8B001. ASTRA.
- buildingSMART International, 2023. *IFC 4.3 Aggregation*. URL: https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/HTML/concepts/Object_Composition/Aggregation/content.html (Stand 25.04.2026).
- buildingSMART International, 2026. *IFC 4.3.2 Documentation*. URL: <https://ifc43-docs.standards.buildingsmart.org/> (Stand 10.05.2026).
- Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I. und Specia, L., 2017. «SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation». In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. SemEval 2017. Hrsg. von Bethard, S., Carpuat, M., Apidianaki, M., Mohammad, S. M., Cer, D. und Jurgens, D. Vancouver, Canada: Association for Computational Linguistics, S. 1–14. DOI: [10.18653/v1/S17-2001](https://doi.org/10.18653/v1/S17-2001).
- Chen, H., Benfer, R., Zong, C., Reitberger, R., Präger, L. und Lang, W., 2025. «A RAG-based framework for material matching in LCA: integrating semantic and character similarity with AI-driven explanations». In: 2025 Building Simulation Conference. DOI: [10.26868/25222708.2025.1449](https://doi.org/10.26868/25222708.2025.1449).
- Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D. und Liu, Z., 2025. *M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation*. DOI:

- [10.48550/arXiv.2402.03216](https://arxiv.org/abs/2402.03216). arXiv: [2402.03216 \[cs\]](https://arxiv.org/abs/2402.03216). URL: <http://arxiv.org/abs/2402.03216> (Stand 30.04.2026). Vorveröffentlichung.
- Devlin, J., Chang, M.-W., Lee, K. und Toutanova, K., 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. DOI: [10.48550/arXiv.1810.04805](https://arxiv.org/abs/1810.04805). arXiv: [1810.04805 \[cs\]](https://arxiv.org/abs/1810.04805). URL: <http://arxiv.org/abs/1810.04805> (Stand 05.10.2025). Vorveröffentlichung.
- Forth, K., Kaltenegger, J., Petrova, E. und De Wolf, C., 2025. «BIM-based material property enrichment using text-based and ontology-based semantic similarity matching». In: *IOP Conference Series: Earth and Environmental Science* 1554.1, S. 012043. ISSN: 1755-1315. DOI: [10.1088/1755-1315/1554/1/012043](https://doi.org/10.1088/1755-1315/1554/1/012043).
- Forth, K., Abualdenien, J. und Borrmann, A., 2023. «Calculation of embodied GHG emissions in early building design stages using BIM and NLP-based semantic model healing». In: *Energy and Buildings* 284, S. 112837. ISSN: 03787788. DOI: [10.1016/j.enbuild.2023.112837](https://doi.org/10.1016/j.enbuild.2023.112837).
- Forth, K., Berggold, P. und Borrmann, A., 2025. «Domain-Specific Fine-Tuning of LLM for Material Matching of BIM Elements and Material Passports». In: *Computing in Civil Engineering 2024. Proceedings*, S. 574–582. DOI: [10.1061/9780784486122.064](https://doi.org/10.1061/9780784486122.064).
- Hermann, F., 2023. «Methodik zur CO2-Bilanzierung im Infrastrukturbau auf Grundlage von BIM-Modellen und der visuellen Aufbereitung von Ergebnissen in einer digitalen Umgebung». Magisterarb. Leipzig: Hochschule für Technik, Wirtschaft und Kultur Leipzig.
- Hermann, F., Max, P. C. und Kunz, M., 2024. «Automatisierte, BIM-basierte Ökobilanzierung am Beispiel des Infrastrukturbaus». In: *Bautechnik* 101.2, S. 128–133. ISSN: 0932-8351, 1437-0999. DOI: [10.1002/bate.202300115](https://doi.org/10.1002/bate.202300115).
- Hugging Face, 2025. *WordPiece tokenization*. URL: <https://huggingface.co/learn/llm-course/chapter6/6> (Stand 25.04.2026).
- IfcOpenShell Contributors, 2026. *IfcOpenShell*. URL: <https://github.com/IfcOpenShell/IfcOpenShell>.
- Infra Suisse, 2023. *ECO2nstruct – der Ökobilanzrechner für Infrastrukturbau-Projekte*. URL: <https://infra-suisse.ch/nachhaltigkeit/eco2nstruct/> (Stand 22.05.2026).
- Jurafsky, D. und Martin, J. H., 2026. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models*. 3rd. URL: <https://web.stanford.edu/~jurafsky/slp3/>.
- KBOB und ecobau, 2026. *Ökobilanzdaten im Baubereich*. Version 8.02. URL: <https://www.kbob.admin.ch/de/oekobilanzdaten-im-baubereich>.
- KI.NRW, 2025. *Glossar: KI-Schlüsselbegriffe*. URL: <https://www.ki.nrw/ki-schluesselfbegriffe/> (Stand 02.10.2025).
- Kudo, T., Edmond, Hayashi, K., Nacht, P. K. K., Gowda, T. und Hiroyuki, T., 2026. *google/sentencepiece*. URL: <https://github.com/google/sentencepiece>.
- Macdonald, C., 2026. *Hygros/ifc-kbob-ai-matcher*. URL: <https://github.com/Hygros/ifc-kbob-ai-matcher>.
- NNBS, 2025. *Warum nachhaltig Bauen?* URL: <https://nnbs.ch/warum-nachhaltig-bauen> (Stand 25.09.2025).

- Petrosa, D., Haverkamp, P., Backes, J. G., Crampen, D., Blankenbach, J. und Traverso, M., 2025. «Development of a BIM-based AI-driven matching tool for LCA datasets». In: *Discover Sustainability* 6.1, S. 1237. ISSN: 2662-9984. DOI: [10.1007/s43621-025-02203-8](https://doi.org/10.1007/s43621-025-02203-8).
- Reimers, N. und Gurevych, I., 2019. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. DOI: [10.48550/arXiv.1908.10084](https://doi.org/10.48550/arXiv.1908.10084). arXiv: [1908.10084 \[cs\]](https://arxiv.org/abs/1908.10084). URL: <http://arxiv.org/abs/1908.10084> (Stand 26. 10. 2025). Vorveröffentlichung.
- Reimers, N. und Gurevych, I., 2025. *SBERT.net*. SentenceTransformers Documentation. URL: <https://sbert.net/> (Stand 09. 11. 2025).
- SN EN 15643, 2021. *Nachhaltigkeit von Bauwerken - Allgemeine Rahmenbedingungen zur Bewertung von Gebäuden und Ingenieurbauwerken*. SIA.
- SN EN 15804, 2022. *Nachhaltigkeit von Bauwerken - Umweltproduktdeklarationen - Grundregeln für die Produktkategorie Bauprodukte*. SIA.
- SN EN 15941, 2025. *Nachhaltigkeit von Bauwerken - Datenqualität für die Erfassung der Umweltqualität von Produkten und Bauwerken - Auswahl und Anwendung von Daten*. SIA.
- SN EN 17472, 2022. *Nachhaltigkeit von Bauwerken - Bewertung der Nachhaltigkeit von Ingenieurbauwerken - Rechenverfahren*. SIA.
- SN EN ISO 22057, 2022. *Nachhaltigkeit von Gebäuden und Ingenieurbauwerken - Datenvorlagen für die Verwendung von Umweltproduktdeklarationen (EPDs) für Bauprodukte in der Bauwerksinformationsmodellierung (BIM)*. SIA.
- SNBS und NNBS, 2026. *Kriterienbeschrieb SNBS-Infrastruktur*. 1.1. URL: https://nnbs.ch/wp-content/uploads/2026/03/2026-03-01_Kriterienbeschrieb_SNBS-Infrastruktur_V1.1.pdf.
- Stettler, C. und Vorhoff, G., 2023. *Ökoinventare für Verkehrsinfrastruktur*. VSS. URL: https://carbotech.ch/cms/wp-content/uploads/VSS_Oekoinventare-Infrastruktur_v04-VSS_fin-Kopie.pdf.
- Trümpler, L., 2026. *louistrue/ifc-lite*. URL: <https://github.com/louistrue/ifc-lite>.
- Tschümperlin, L., Ramseier, L. und Frischknecht, R., 2020. *Ökobilanz ausgewählter Betonsorten*. 4.1. Zürich: Stadt Zürich, Amt für Hochbauten. URL: https://treeze.ch/fileadmin/user_upload/downloads/Publications/Case_Studies/Building_and_Construction/551_%C3%96kobilanz_ausgew%C3%A4hlter_Betonsorten_v4.1_TitelblattAHB.pdf.
- Ward, I., 2026. *wardi/jsonlines*. URL: <https://github.com/wardi/jsonlines>.

Abbildungsverzeichnis

2.1	Mögliche Struktur und Beziehungen von IFC-Elementen, eigene Grafik nach buildingSMART International (2023).	10
2.2	Schematische Einordnung von NLP in die Teilgebiete von KI, eigene Grafik in Anlehnung an Jurafsky und Martin (2026).	11
2.3	Beispielhafte Tokenisierung eines Queries mit dem Modell BGE-M3 (J. Chen et al., 2025), eigene Darstellung.	12
4.1	Methodisches iteratives Vorgehen, eigene Darstellung.	18
4.2	Von der Forschungsfrage und den Teilfragestellungen zum Suchstring, eigene Darstellung.	19
5.1	Zentrale Funktionen des Prototyps, eigene Grafik.	22
5.2	Konzeptioneller Systemablauf des Prototyps, eigene Grafik.	23
5.3	Relevante Informationen für die IFC-Elemente, eigene Grafik.	24
5.4	Der Mapping-Datensatz als zentrales Objekt für den Prototyp, eigene Grafik.	27
5.5	Der Human-in-the-loop Ansatz für den Prototyp, eigene Grafik.	29
5.6	Genereller Ablauf für das Fine-Tuning, eigene Grafik.	30
6.1	Tab «Uploads» des Prototyps, Screenshot Benutzeroberfläche des Prototyps.	37
6.2	Tab «AI-Mapping» des Prototyps mit selektiertem Stahlträger, Screenshot Benutzeroberfläche des Prototyps.	37
6.3	Tab «Charts» des Prototyps mit Säulen-Diagramm der berechneten UBP'21, Screenshot Benutzeroberfläche des Prototyps.	38
6.4	Eintrag eines IfcBeam in der JSONL-Datei (nicht alle <code>top_k_matches</code> dargestellt), eigene Darstellung.	39
6.5	Aggregationslogik des Prototyps, eigene Darstellung.	41
6.6	Ablauf vom KBOB-Eintrag zur Ausgabe im Dashboard, eigene Darstellung.	42
6.7	Beispiel-Queries aus der selbst erstellten IFC-Datei, eigene Darstellung.	44

Tabellenverzeichnis

2.1	Überblick der massgebenden Normen in Bezug auf Nachhaltigkeit von Bauwerken, eigene Grafik in Anlehnung an SN EN 17472 (2022, 8)	4
2.2	Phasen der Nachhaltigkeitsbewertung, eigene Grafik in Anlehnung an SN EN 17472 (2022, 32)	5
2.3	Bereiche und Themen der Nachhaltigkeitsbewertung gemäss SNBS, eigene Grafik in Anlehnung an SNBS und NNBS (2026, 10)	6
2.4	Auszug der Ökobilanzdaten der KBOB, Screenshot KBOB und ecobau (2026).	7
5.1	Beispiele möglicher Kontextstrings aus den IFC-Informationen, jeweils ein Kontextstring pro Zeile, eigene Tabelle.	26
5.2	Relevante Informationen aus der KBOB-Liste am Beispiel Tiefbaubeton. Die vollständige Tabelle befindet sich im Anhang A.6, eigene Tabelle.	26
6.1	Verwendete Trainings-Hyperparameter des Fine-Tunings auf «Calculon», eigene Tabelle. .	46
7.1	Hit@1 nach Materialfamilie auf den 389 evaluierten Queries, eigene Tabelle.	50
7.2	Die 12 fehlerhaften Queries des feinabgestimmten Modells auf kontrollierten Testdaten, gruppiert nach Fehlermuster, eigene Tabelle.	51
7.3	Vergleich der vier Evaluationsszenarien auf 389 Test-Queries, eigene Tabelle.	53
7.4	Modellvergleich auf 389 Test-Queries. Die obere Gruppe zeigt aufgabenspezifische Ansätze, die untere die besten generischen Embedding-Modelle. Mit GPT-5.5 wurde nur die beste Zuordnung bestimmt, weshalb Hit@10 fehlt, eigene Tabelle.	54
7.5	Trefferquote des feinabgestimmten Modells auf vier realen Infrastrukturprojekten, eigene Tabelle.	56

Verwendete Hilfsmittel

Hilfsmittel	Verwendungszweck	Einsatz im Projekt
GitHub Copilot	Unterstützung bei Code-Vorschlägen, Refactoring, Debugging-Hinweisen, Eval-/Testskripten sowie bei README- und Architekturfragen	Implementierung und Weiterentwicklung des Prototyps
M365 Copilot	Unterstützung bei sprachlicher Überarbeitung, Zusammenfassungen, Strukturierung einzelner Kapitel sowie methodischen Sparringsfragen	Dokumentation und Ausarbeitung der Masterarbeit
VS Code + lokale LaTeX-Toolchain	Schreiben, Bauen und Prüfen der Arbeit mit LaTeX Workshop, LuaLaTeX, biber, latexmk und SumatraPDF	Erstellung, Kompilierung und Korrekturlesen der wissenschaftlichen Arbeit
Git / GitHub	Versionsverwaltung, Nachvollziehbarkeit der Entwicklung sowie Pflege von README und Repository-Dokumentation	Entwicklung, Dokumentation und Veröffentlichung des Prototyps
Hugging Face Spaces	Bereitstellung einer lauffähigen Demo des Prototyps für Tests und Demonstrationszwecke	Demo- und Bereitstellungsplattform
Zotero + Better BibTeX	Literaturverwaltung, PDF-Ablage, Annotationen sowie Export und Nachpflege von BibLaTeX-Einträgen	Literaturrecherche und Quellenmanagement

Anhang

A.1 Auszug KBOB-Ökobilanzdaten

BAUMATERIALIEN	Rohdichte/ Flächen- masse	Bezug	UBP'21			Primärenergie										Treibhausgasemissionen		
						erneuerbar					nicht erneuerbar (Graue Energie)							
			Total UBP	Herstellung UBP	Entsorgung UBP	Total kWh oil-eq	Herstellung total kWh oil-eq	Herstellung energetisch genutzt kWh oil-eq	Herstellung stofflich genutzt kWh oil-eq	Entsorgung kWh oil-eq	Total kWh oil-eq	Herstellung total kWh oil-eq	Herstellung energetisch genutzt kWh oil-eq	Herstellung stofflich genutzt kWh oil-eq	Entsorgung kWh oil-eq	Total kg CO ₂ -eq	Herstellung kg CO ₂ -eq	Entsorgung kg CO ₂ -eq
Vorbereitungsarbeiten	-																	
Baugrubensicherung, Bohrpfahlwand, gespriesst	4'884	m ²	1'300'000	1'140'000	167'000	132	123	123	0	9.21	2'120	1'870	1'870	0	248	861	801	60.5
Baugrubensicherung, Spundwand, auskragend	153	m ²	287'000	286'000	1'690	45.7	45.6	45.6	0	0.089	732	728	728	0	4.16	169	168	1.06
Beton	kg/m ³																	
Magerbeton (ohne Bewehrung)	2'150	kg	104	68.8	35.4	0.010	0.008	0.008	0	0.002	0.128	0.077	0.077	0	0.052	0.063	0.050	0.013
Tiefbaubeton (ohne Bewehrung)	2'350	kg	163	128	35.1	0.016	0.014	0.014	0	0.002	0.189	0.138	0.138	0	0.051	0.109	0.096	0.013
Bohrpfahlbeton (ohne Bewehrung)	2'325	kg	174	139	35.1	0.017	0.015	0.015	0	0.002	0.197	0.146	0.146	0	0.051	0.119	0.107	0.013
Betonfertigteil, hochfester Beton, ab Werk	2'770	kg	440	409	31.0	0.076	0.074	0.074	0	0.002	0.697	0.652	0.652	0	0.045	0.276	0.265	0.011
Betonfertigteil, Normalbeton, ab Werk	2'500	kg	282	248	33.9	0.095	0.093	0.093	0	0.002	0.432	0.383	0.383	0	0.050	0.177	0.165	0.012
Mauersteine	kg/m ³																	
Backstein	900	kg	371	328	42.5	0.081	0.078	0.078	0	0.002	0.787	0.734	0.734	0	0.053	0.266	0.254	0.013
Kalksandstein	1'400	kg	258	216	42.5	0.039	0.037	0.037	0	0.002	0.389	0.337	0.337	0	0.053	0.162	0.149	0.013
Zementstein	1'700	kg	214	172	42.5	0.026	0.024	0.024	0	0.002	0.225	0.173	0.173	0	0.053	0.126	0.114	0.013
Andere Massivbaustoffe	kg/m ³																	
Kies gebrochen ¹	1'400 - 1'500	kg	64.8	24.6	40.2	0.008	0.006	0.006	0	0.002	0.085	0.032	0.032	0	0.053	0.018	0.005	0.013
Rundkies ¹	1'600 - 1'700	kg	59.4	19.2	40.2	0.004	0.002	0.002	0	0.002	0.068	0.016	0.016	0	0.053	0.016	0.003	0.013
Sand ¹	1'500	kg	59.4	19.2	40.2	0.004	0.002	0.002	0	0.002	0.068	0.016	0.016	0	0.053	0.016	0.003	0.013

A.2 Übersicht LCA-Tools

Tool/Plattform	Fokus	Datenbasis/Normen	Schweiz-Kompatibilität	IFC-Kompatibilität	automatisierte Bauteilzuordnung	Kosten	Eignung Infrastruktur	Begründung
IfcLCA	BIM-basierte Ökobilanzierung	KBOB, SIA 2032	Hoch	Ja	Mittel	Niedrig	Mittel	Open-Source-Tool mit direkter IFC-Auswertung und KBOB-Integration. Automatisches Material Matching ist vorhanden, wirkt aber prototypisch und nicht KI-basiert.
Vyzn	BIM-basierte Nachhaltigkeitsanalyse	KBOB, SIA, SNBS, Minergie	Hoch	Ja	Hoch	Mittel	Gering	Schweizer Plattform mit KI-gestützter Modellanalyse und normkonformen Ergebnissen aus digitalen Gebäudemodellen. Fokus liegt klar auf Hochbau.
viride	Nachhaltigkeitsbewertung von Gebäuden	KBOB, SIA 2032, SIA 2040, SNBS, Minergie	Hoch	Ja	Niedrig	Mittel	Gering	Unterstützt manuelle Eingabe und BIM-Workflows mit IFC-Modellen. Eine KI-basierte Bauteilzuordnung ist öffentlich nicht ausgewiesen.
One Click LCA	Kommerzielle LCA-Plattform für Bauwerke	Internationale Datenbanken, EPDs, viele Standards	Hoch	Ja	Hoch	Hoch	Mittel	Breite BIM-Integrationen und KI-gestütztes Material Mapping, unter anderem für Synonyme, Schreibvarianten und Mehrsprachigkeit.
Eco2nstruct	Ökobilanzierung im Infrastrukturbau	Schweizer Ökobilanzdaten, Infrastrukturmaterialien und Prozesse	Hoch	Nein	Niedrig	Niedrig	Hoch	Speziell für Infrastrukturbauprojekte in der Schweiz entwickelt. Erfassung von Materialien, Maschinen und Prozessen, aber keine erkennbare BIM/IFC-basierte automatische Zuordnung.
ORIS	Infrastruktur-LCA und Variantenvergleich	Regionale Materialdaten, LCA-Daten, Lieferketten	Gering	Ja	Hoch	Mittel	Hoch	Infrastrukturplattform mit BIM-Integration, automatisierter Mengenermittlung und AI-powered material mapping für BIM-Elemente.
SimaPro	Allgemeine LCA-Modellierung	ecoinvent, internationale LCA-Datenbanken, ISO/EN-Methoden	Gering	Nein	Niedrig	Hoch	Mittel	Fachlich starkes LCA-Werkzeug, aber nicht BIM-nativ. Automatische BIM-Bauteilzuordnung ist nur über externe Workflows oder Forschungsansätze möglich.
GaBi (Sphera)	Industrielle und produktbezogene LCA	Sphera/GaBi-Datenbanken, EN/ISO, EPD	Gering	Nein	Niedrig	Hoch	Mittel	LCA-Automation ist für Produktportfolios und Stücklisten verfügbar, aber keine spezifische BIM-native Bauteilzuordnung für Bauwerksmodelle erkennbar.
openLCA	Open-Source-LCA-Modellierung	frei wählbare Datenbanken, z. B. ecoinvent	Mittel	Nein	Niedrig	Niedrig	Gering	Sehr flexibel und offen, aber keine eigene BIM-native Zuordnung von Bauteilen zu Ökobilanzdaten. Integration wäre über API oder eigene Skripte möglich.
eTool (cerclos)	Gebäude- und Infrastruktur-LCA	Internationale LCA-Daten, EN 15978, ISO 14040/14044	Mittel	Nein	Mittel	Mittel	Mittel	Revit-Plugin und CSV/BIM-Import verfügbar. Familien oder Mengen werden mit Vorlagen abgeglichen, aber nicht KI-basiert automatisiert.
EC3	Embodied Carbon und EPD-Vergleich	EPD-Datenbank, produktspezifische Emissionsdaten	Gering	Nein	Mittel	Niedrig	Mittel	BIM- und Autodesk-Integrationen für Mengenübernahme vorhanden. Die ökologische Zuordnung bleibt stark EPD- und prüfungsbasiert.
Tally	Revit-basierte Gebäude-LCA	Tally/GaBi-Datenbank, LCA-Daten	Gering	Nein	Niedrig	Mittel	Gering	Revit-Plugin mit Materialzuweisung innerhalb des Modells. Beziehungen zwischen BIM-Elementen und LCA-Materialien müssen durch Nutzende definiert werden.
EcoTool (CH)	Frühe Ökobilanzierung im Hochbau	KBOB, SIA-Logik	Hoch	Nein	Niedrig	Niedrig	Gering	Browserbasiertes Tool für frühe Planungsphasen mit Bauteilgenerator und KBOB-Daten, aber ohne erkennbare BIM/IFC-basierte Bauteilzuordnung.
madaster	Materialpass und Kreislaufwirtschaft	EPDs, Ökobaumat, ecobau/KBOB je nach Markt	Hoch	Ja	Mittel	Mittel	Mittel	Erstellt Materialpässe aus IFC und reichert Modellinformationen mit Umwelt- und Kreislaufdaten an. Die automatische Zuordnung hängt stark von gepflegten Materialinformationen im Modell ab.

Prüfkriterien	Hoch			Mittel			Niedrig	
	Schweiz-Kompatibilität	KBOB-Daten direkt im Tool nutzbar		KBOB-Daten nur über Import, eigene Datensätze oder externe Workflows nutzbar			Keine erkennbare Nutzung von KBOB-Daten	
	IFC-Kompatibilität	IFC-Modelle können eingelesen und ausgewertet werden		BIM-Daten können über Schnittstellen, Plugins oder Exporte genutzt werden, aber nicht vollständig IFC-nativ			Keine direkte IFC- oder BIM-Verarbeitung erkennbar	
	automatisierte Bauteilzuordnung	Bauteile oder Materialien werden automatisiert und KI-gestützt Ökobilanzdaten zugeordnet		Teilautomatische Zuordnung möglich, z.B. über Regeln, Vorlagen, Datenbankabgleich oder manuelle Bestätigung			Keine automatisierte Zuordnung zu Ökobilanzdaten erkennbar	
	Kosten	Kommerzielle Lizenz mit hohen Kosten oder nur auf Anfrage verfügbar		Kostenpflichtiges Tool mit moderaten, projekt- oder nutzerbasierten Gebühren			Kostenlos, Open Source oder mit kostenfrei nutzbarer Basisversion	
	Eignung Infrastruktur	Tool ist explizit für Infrastrukturprojekte oder Tiefbau ausgelegt		Tool kann für Infrastruktur verwendet werden, ist aber nicht primär darauf ausgerichtet			Tool ist primär auf Hochbau, Produkte oder allgemeine Ökobilanzen ausgerichtet	

A.2.1 Erläuterungen zu den LCA-Tools

IfcLCA wertet IFC-Modelle direkt aus und verknüpft sie mit Schweizer KBOB-Daten. Es unterstützt IFC 2x3 und IFC 4, verarbeitet Modellinformationen browserbasiert und stellt Kennwerte wie Treibhausgasemissionen, Primärenergie und Umweltbelastungspunkte dar, ist aber momentan eher ein experimentelles Werkzeug als eine etablierte Lösung.

Vyzn ist eine Schweizer Plattform für BIM-gestützte Gebäudeanalysen. Sie kombiniert digitale Gebäudemodelle mit Auswertungen zu CO₂, grauer Energie, Energiebedarf, Komfort und Kosten. Für die Ökobilanzierung verweist Vyzn auf Grundlagen wie SIA 2032, SIA 390/1 und KBOB sowie auf die Zulassung für Minergie und SNBS. Der Schwerpunkt liegt damit auf modellbasierten Nachhaltigkeitsanalysen im Hochbau.

viride verfolgt ebenfalls einen gebäudebezogenen Ansatz im Hochbau. Die Plattform unterscheidet zwischen manueller Dateneingabe mit greenDESIGN und BIM-basierter Auswertung mit greenBIM. Es werden Schweizer Standards und Daten wie SIA-Normen, Minergie, SNBS und KBOB berücksichtigt.

One Click LCA ist ein international verbreitetes LCA-Werkzeug mit breiter Datenbasis, zahlreichen Standards und BIM-Schnittstellen. Es eignet sich für unterschiedliche Gebäudetypen und Zertifizierungssysteme, ist aber nicht spezifisch auf den Schweizer Infrastrukturbau und die direkte Zuordnung von IFC-Infrastrukturmodellen zu KBOB-Daten ausgerichtet.

Eco2nstruct wurde speziell für Infrastrukturbauprojekte entwickelt. Das Werkzeug erlaubt die Berechnung von CO₂-Belastung und Umweltbelastungspunkten und berücksichtigt neben Baustoffen auch Betriebsstoffe, Baumaschinen und Bauprozesse. Damit adressiert es den Infrastrukturbau direkter als viele gebäudebezogene LCA-Werkzeuge. Eco2nstruct hat jedoch keine Möglichkeit um ein BIM-Modell auszuwerten und verfolgt einen rein manuellen Ansatz.

ORIS fokussiert auf Infrastrukturprojekte. Die Plattform unterstützt die Auswertung von IFC-Modellen und ermöglicht Variantenvergleiche, Materialzuordnungen, automatisierte Mengenauswertungen und CO₂-Bewertungen. Für den Schweizer Markt ist ORIS jedoch nur eingeschränkt anwendbar, da weder eine Auswertung nach UBP vorgesehen ist noch KBOB-Ökobilanzdaten integriert sind.

Die Bewertung basiert teils auf eigener Anwendung und teils auf öffentlich verfügbaren Produktinformationen der Anbieter. Da nicht für alle Werkzeuge ein Testzugang verfügbar war, konnten mehrere Plattformen nicht praktisch geprüft werden.

A.3 Von der Forschungsfrage zum Suchstring

Forschungsfrage

Wie kann ein KI-gestütztes Verfahren prototypisch umgesetzt werden, das Materialnamen aus IFC-Modellen semantisch den passenden Ökobilanzdaten zuordnet und dabei kontextabhängige Informationen berücksichtigt, um eine praxisnahe Anwendung im Schweizer Infrastrukturbau zu ermöglichen?

Teilfragestellungen

Welche Besonderheiten ergeben sich bei der Anwendung im Schweizer Markt bezüglich Nachhaltigkeitsbewertung?

Wie können Schweizer Ökobilanzdaten, Normen und Richtlinien in die Ökobilanzierung integriert werden?

Welche Anforderungen stellen SNBS und KBOB an die Granularität und Genauigkeit der Materialklassifikation?

Wie und welche Kontextinformationen (z.B. Bewehrungsgehalt, Beziehungen zu anderen Bauteilen, Lage im Modell) aus dem IFC-Modell können für das semantische Mapping verwendet werden um die Ergebnisse zu verbessern?

Welche NLP-Methoden bzw. -Modelle eignen sich zur automatisierten Verknüpfung von Materialbezeichnungen aus IFC-Modellen mit EPD- bzw. Ökobilanzdaten, insbesondere unter Berücksichtigung von Datenlücken und sprachlicher Uneinheitlichkeit?

Welche Mindestanforderungen muss ein IFC-Modell für die semantische Anreicherung und der daraus folgenden Ökobilanzierung erfüllen?

Schlüsselbegriff	Synonyme/Varianten
Materialnamen	Material name, Materialbezeichnung, material classification, material mapping, material identification
IFC-Modell	IFC, Building Information Modeling, BIM, IFC file, IFC schema
Ökobilanzdaten	Life Cycle Assessment, LCA, Environmental Product Declaration, EPD, Ökobilanz, KBOB
semantisches Mapping	semantic mapping, semantic matching, semantic linking, semantische Zuordnung
NLP-Methoden	Natural Language Processing, NLP, Large Language Model, LLM, BERT, embedding, entity linking

Suchstring

("IFC" OR "Building Information Modeling" OR "BIM") AND ("semantic" OR "semantisch" OR "semantic mapping" OR "semantic similarity") AND ("Natural Language Processing" OR "NLP" OR "Large Language Model" OR "LLM" OR "BERT" OR "embedding")

Scopus

Search Sources SciVal ⓘ ⓘ ⓘ ⓘ ⓘ CM

Advanced query ⓘ

Search within

Article title, Abstract, Keywords

Search documents *

"IFC" OR "Building Information Modeling" OR "BIM"

AND

Search within

Article title, Abstract, Keywords

Search documents

"semantic" OR "semantisch" OR "semantic mapping" OR "semanti

AND

Search within

Article title, Abstract, Keywords

Search documents

"Natural Language Processing" OR "NLP" OR "BERT"

Save search

Set search alert

Add search field

Reset

Search

Documents

Beta

Preprints

Secondary documents

72 documents found

Analyze results

Refine search

Search within results

Filters

Year

Range

Individual

Document title

Authors

Source

Year

Citations

1

Framework for automated building code compliance checking to improve transparency, trust, validation, and design interpretation

Purushotham, N., Kailashnath, C., Mutis, I., 181, 106598

Automation in Construction

2026

0

Show abstract

Check Full Text

View at Publisher

Related documents

2

Semiautomatic IFC-Based BIM Reconstruction for As-Designed Bridges from 2D Plans Leveraging Semantic Segmentation and Enrichment

Li, H., Zhang, J., Journal of Computing in Civil Engineering

2025

0

Article

Help

Neue Suche

Zeitschriftensuche

Fachdatenbanken

Macdonald, Colin

DE

FR

IT

EN

SUCHKRITERIEN

Einfache Suche

Überall enthält "IFC" OR "Building Information Modeling" OR "BIM"

UND Überall enthält "semantic" OR "semantisch" OR "semantic mapping" OR "semantic similarity"

Suche

Clarivate

English

Products

Web of Science

Smart Search

Advanced Search

Research Assistant

Sign In

Register

Search

Refine results for "IFC" OR "Building Information Modeling" OR "BIM" (Topic)

83 results from All Databases for:

"IFC" OR "Building Information Modeling" OR "BIM" (Topic) and "semantic" OR "semantisch" OR "sem..."

Copy query link

Did you mean? "IFC" OR "Building Information Modeling" OR "BIM" (Topic) and "semantic" OR "semantische" OR "semantic mapping" OR "semantic similarity" (Topic) and "Natural Language Processing" OR "NLP" OR "BERT" (Topic) and Preprint Citation Index (Exclude - Database) | 83 results

Add Keywords

Quick add keywords:

automated compliance checking acc

automated compliance checking

industry foundation

Refined By:

NOT Database: Preprint Citation Index

Clear all

Analyze Results

Citation Report

Create Alert

83 Documents

You may also like...

Refine results

Export Refine

Search within topic...

Quick Filters

Review Article

Open Access

Database

Web of Science Core Collection

Derwent Innovations Index

KCI-Korean Journal Database

0/83

Add To Marked List

Export

Sort by Relevance

1 of 2

1

Method for automatic compliance examination of construction engineering based on artificial intelligence (AI) and building information modeling (BIM), involves using rule engine to automatically check logic rule and logic fact by rule reasoning algorithm and ontology semantic reasoning algorithm

CN119271823-A

Inventors : LI Q; YUAN S; (...); SHI J

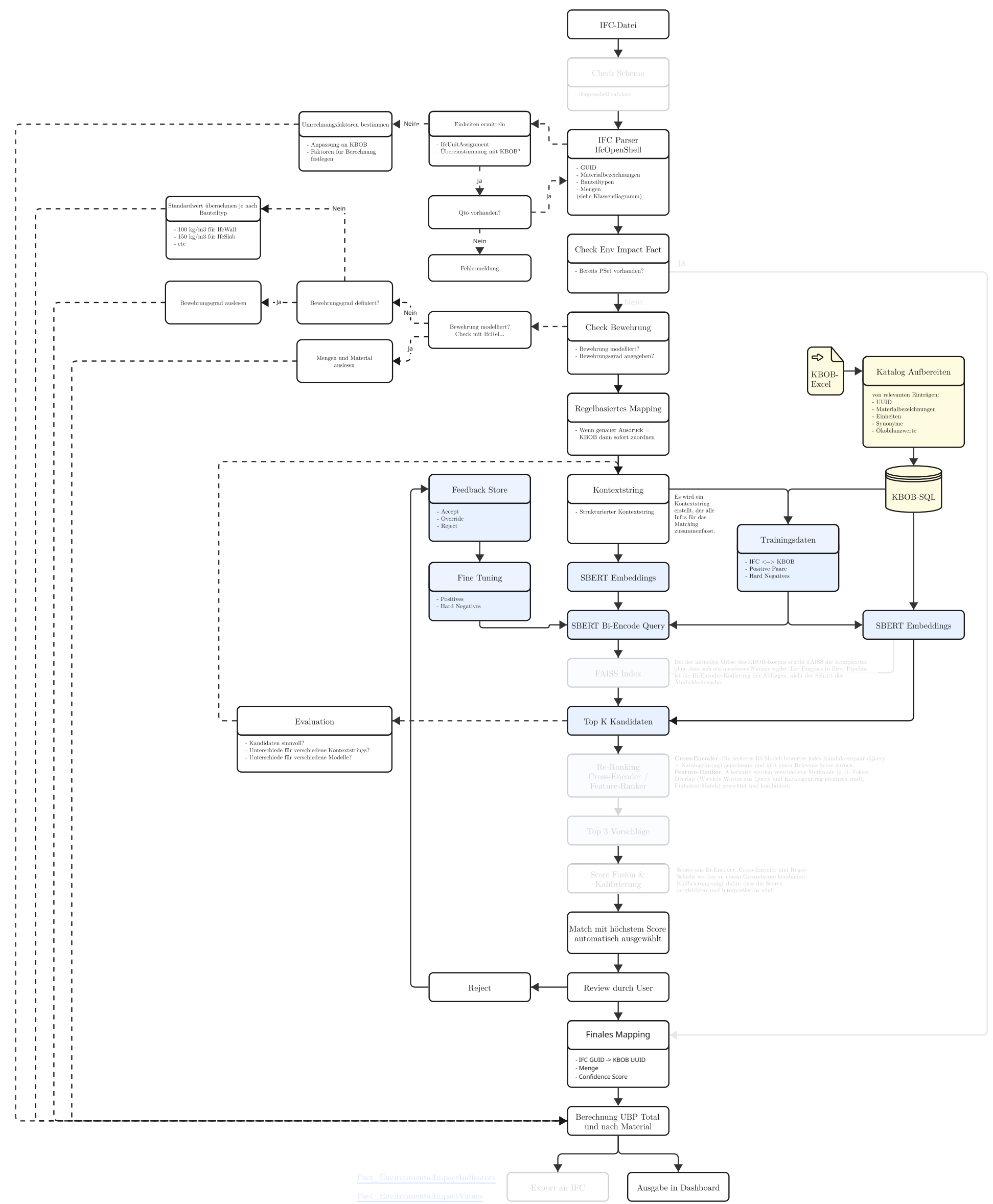
Assignee : UNIV SHANGHAI JIAOTONG CHONGQING RES INS

Derwent Primary Accession Number : 2025-06497W

Master-Thesis MSc FHNW VDC | 2026 | Colin Macdonald

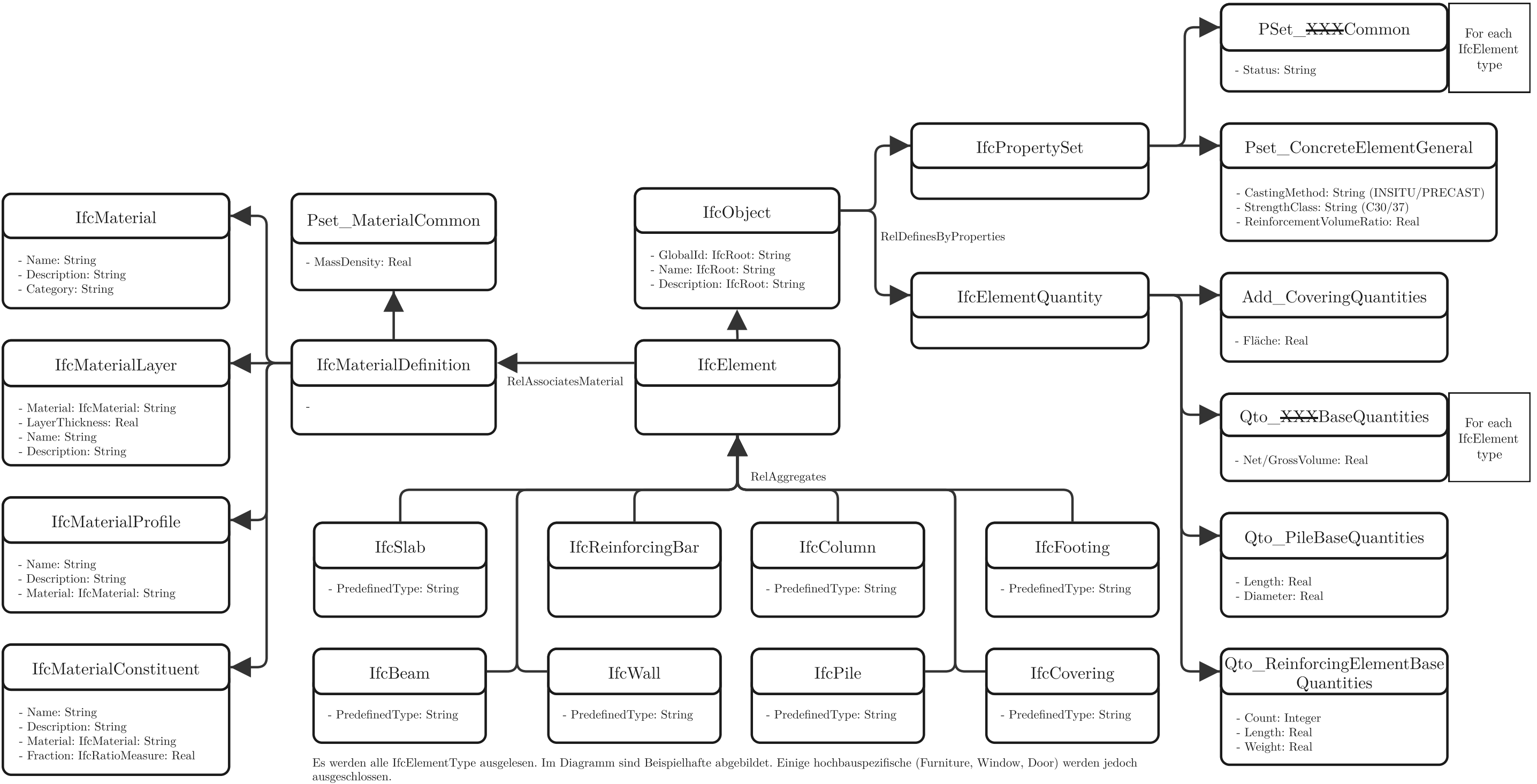
73

A.4 Detaillierter Systemablauf



Die ausgegrauten Schritte wurden nicht implementiert, da sie für den Prototyp keinen Mehrwert bringen. Im Konzept wurden sie jedoch belassen, da sie eine sinnvolle Erweiterung für eine zukünftige Weiterentwicklung darstellen.

A.5 Klassendiagramm IFC 4.3

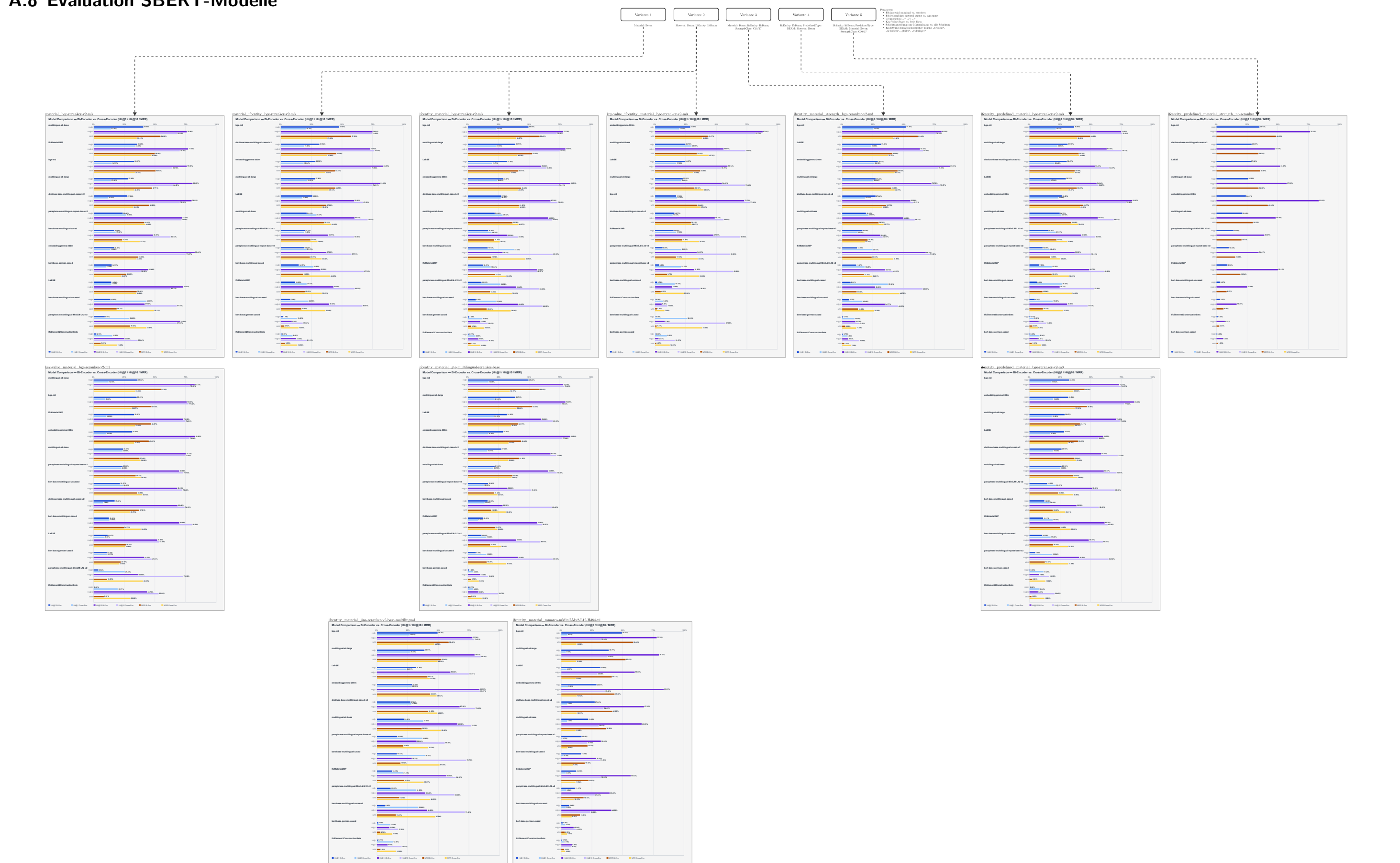


[illegible]

A.7 Bewertung offizielle SBERT-Modelle

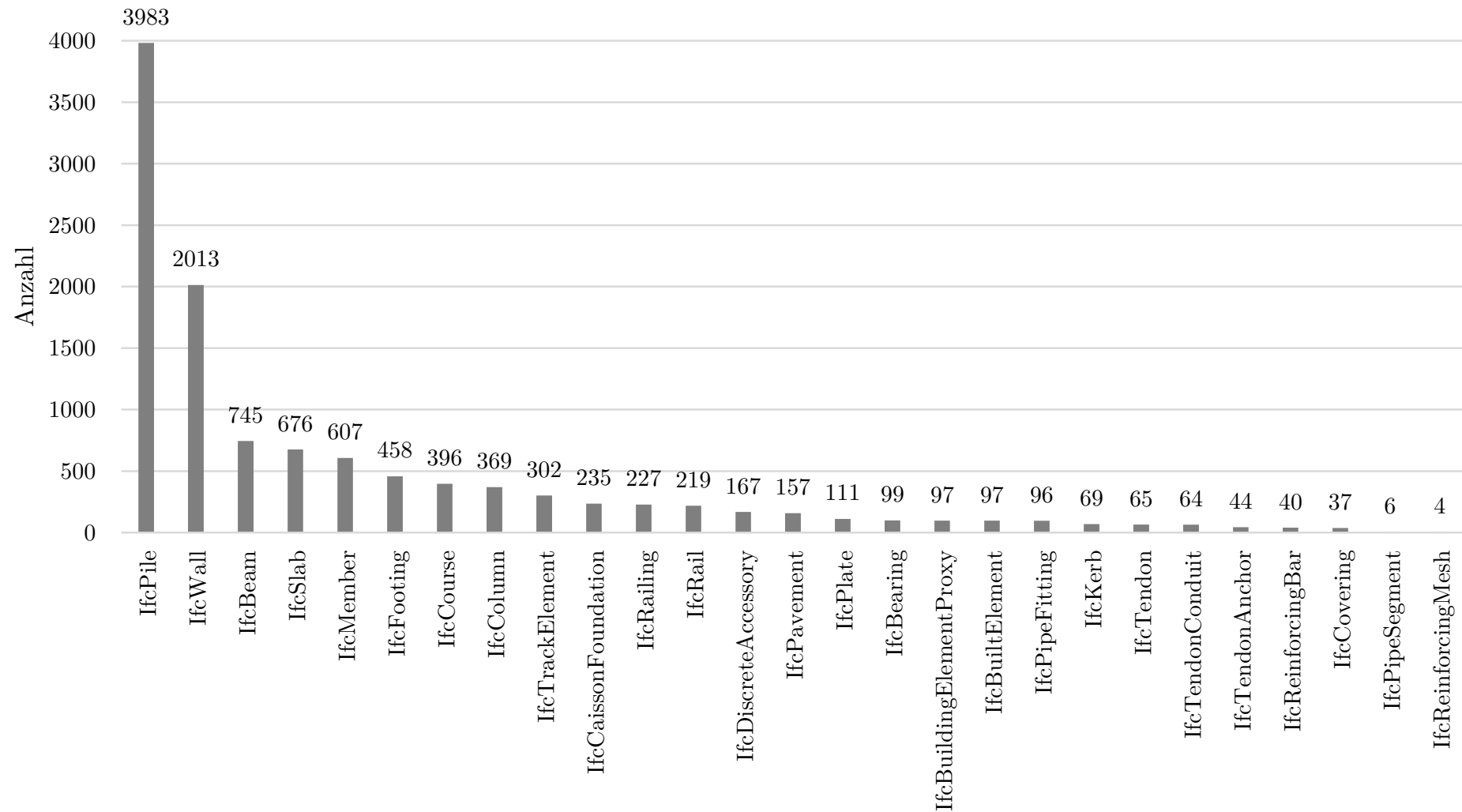
Model	AvgPerf	SemSearch	Speed	Kurztext	DE/FR	Feintuning	TotalScore	Rank	Model
all-MiniLM-L12-v1	59.80	50.78	22.06	80	60	90	66.39	11	all-MiniLM-L12-v1
all-distilroberta-v1	59.84	50.94	11.76	75	60	85	63.90	18	all-distilroberta-v1
all-MiniLM-L12-v2	59.76	50.82	22.06	80	60	90	66.39	12	all-MiniLM-L12-v2
all-MiniLM-L6-v1	58.05	48.07	41.76	80	60	90	68.20	6	all-MiniLM-L6-v1
all-MiniLM-L6-v2	58.80	49.54	41.76	80	60	90	68.75	1	all-MiniLM-L6-v2
all-mpnet-base-v1	62.34	54.69	8.24	85	65	90	67.74	7	all-mpnet-base-v1
all-mpnet-base-v2	63.30	57.02	8.24	85	65	90	68.54	4	all-mpnet-base-v2
all-roberta-large-v1	61.64	53.05	2.35	80	60	80	63.54	20	all-roberta-large-v1
average_word_embeddings_glove.6B.300d	36.25	22.71	100.00	40	40	25	48.73	37	average_word_embeddings_glove.6B.300d
average_word_embeddings_komninos	36.39	21.64	64.71	40	40	25	43.27	38	average_word_embeddings_komninos
distiluse-base-multilingual-cased-v1	45.59	29.87	11.76	75	90	80	61.70	21	distiluse-base-multilingual-cased-v1
distiluse-base-multilingual-cased-v2	43.77	27.35	11.76	75	80	75	57.55	34	distiluse-base-multilingual-cased-v2
gtr-t5-base	59.40	51.15	7.35	80	60	60	61.12	25	gtr-t5-base
gtr-t5-large	62.38	54.85	2.35	80	60	55	61.66	22	gtr-t5-large
gtr-t5-xl	62.88	55.88	0.68	80	60	50	61.29	24	gtr-t5-xl
gtr-t5-xxl	63.25	55.76	0.15	80	60	45	60.81	26	gtr-t5-xxl
msmarco-bert-base-dot-v5	57.39	52.11	8.24	80	55	80	61.49	23	msmarco-bert-base-dot-v5
msmarco-distilbert-base-tas-b	55.91	49.25	11.76	75	55	80	60.43	28	msmarco-distilbert-base-tas-b
msmarco-distilbert-dot-v5	55.66	49.47	11.76	75	55	80	60.39	29	msmarco-distilbert-dot-v5
multi-qa-distilbert-cos-v1	59.41	52.83	11.76	80	60	90	65.12	15	multi-qa-distilbert-cos-v1
multi-qa-distilbert-dot-v1	59.59	52.51	11.76	80	60	90	65.12	16	multi-qa-distilbert-dot-v1
multi-qa-MiniLM-L6-cos-v1	58.08	51.83	41.76	85	55	95	68.71	3	multi-qa-MiniLM-L6-cos-v1
multi-qa-MiniLM-L6-dot-v1	56.55	49.19	41.76	85	55	95	67.65	8	multi-qa-MiniLM-L6-dot-v1
multi-qa-mpnet-base-cos-v1	61.88	57.46	8.24	90	60	90	67.39	10	multi-qa-mpnet-base-cos-v1
multi-qa-mpnet-base-dot-v1	62.18	57.60	8.24	90	60	90	67.52	9	multi-qa-mpnet-base-dot-v1
paraphrase-albert-small-v2	52.25	40.04	14.71	60	55	70	55.25	36	paraphrase-albert-small-v2
paraphrase-distilroberta-base-v2	54.69	43.10	11.76	75	60	85	60.53	27	paraphrase-distilroberta-base-v2
paraphrase-MiniLM-L12-v2	54.51	43.01	22.06	75	65	90	63.74	19	paraphrase-MiniLM-L12-v2
paraphrase-MiniLM-L3-v2	50.74	39.19	55.88	70	60	90	64.98	17	paraphrase-MiniLM-L3-v2
paraphrase-MiniLM-L6-v2	52.56	40.31	41.76	75	65	90	65.47	14	paraphrase-MiniLM-L6-v2
paraphrase-mpnet-base-v2	57.70	47.43	8.24	85	70	90	65.92	13	paraphrase-mpnet-base-v2
paraphrase-multilingual-MiniLM-L12-v2	51.72	39.19	22.06	80	90	90	68.75	2	paraphrase-multilingual-MiniLM-L12-v2
paraphrase-multilingual-mpnet-base-v2	53.75	41.68	7.35	85	90	90	68.25	5	paraphrase-multilingual-mpnet-base-v2
paraphrase-TinyBERT-L6-v2	53.63	41.07	13.24	65	55	80	57.22	35	paraphrase-TinyBERT-L6-v2
sentence-t5-base	56.23	44.63	7.35	80	60	60	58.71	33	sentence-t5-base
sentence-t5-large	58.89	49.05	2.35	80	60	55	59.27	32	sentence-t5-large
sentence-t5-xl	60.21	51.19	0.68	80	60	50	59.41	31	sentence-t5-xl
sentence-t5-xxl	62.64	54.40	0.15	80	60	45	60.33	30	sentence-t5-xxl

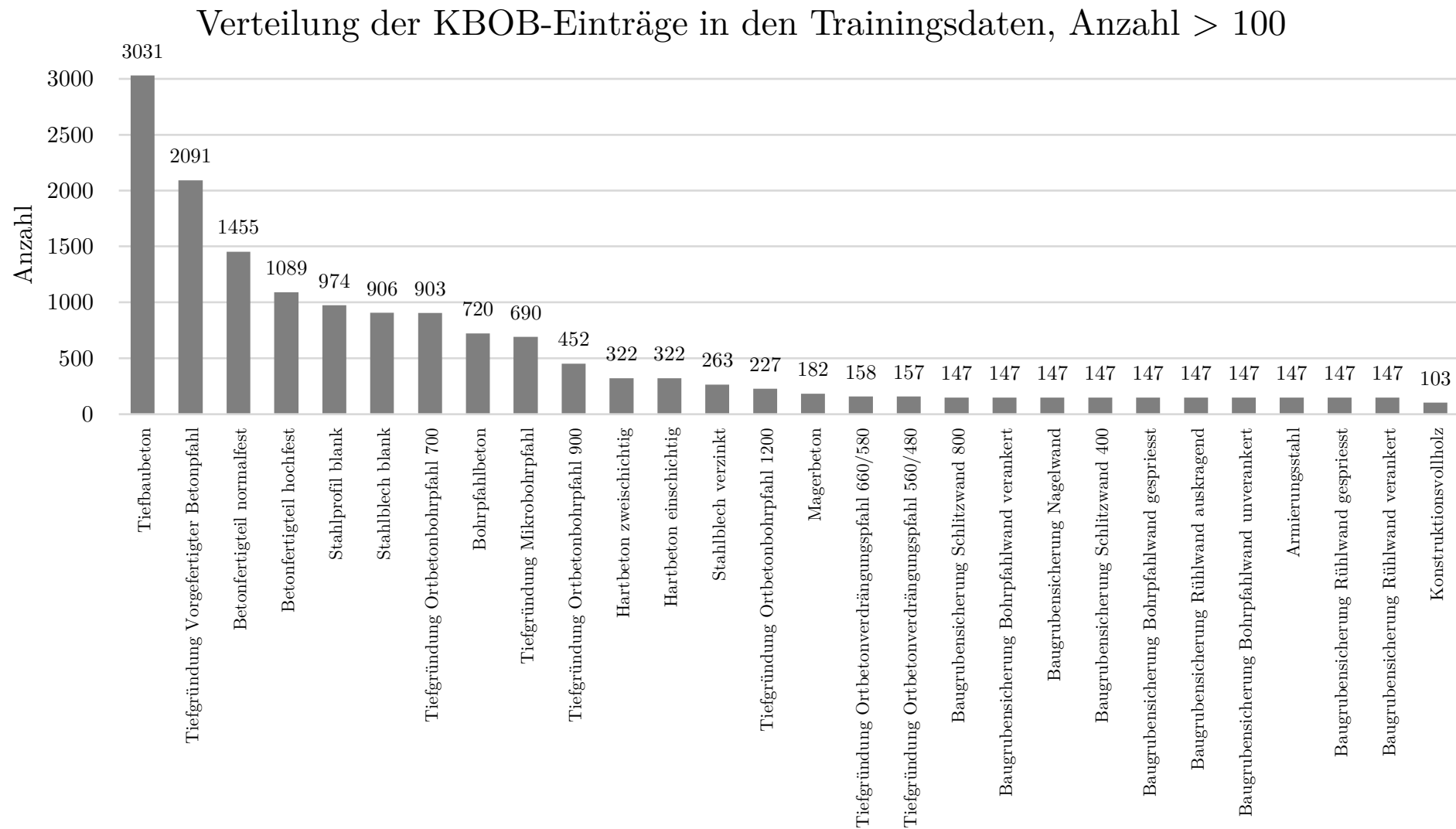
Kriterium	Gewichtung	
A1 AvgPerformance	35	gemäss Website
A2 SemanticSearch	20	gemäss Website
A3 Speed	15	gemäss Website
B1 Kurztext-Robustheit	10	Kurztext-Robustheit: Wie gut das Modell kurze, domänenspezifische Strings (Materialnamen, Abkürzungen) abbildet.
B2 DE/FR-Eignung	25	DE/FR-Eignung: Wie gut das Modell für deutsch/französische Daten geeignet ist (Multilingualität, Trainingsdaten).
B3 Feintuning-Eignung	10	Feintuning-Eignung: Wie einfach und effektiv das Modell für deinen Task nachtrainiert werden kann (SBERT-Architektur, Community-Erfal



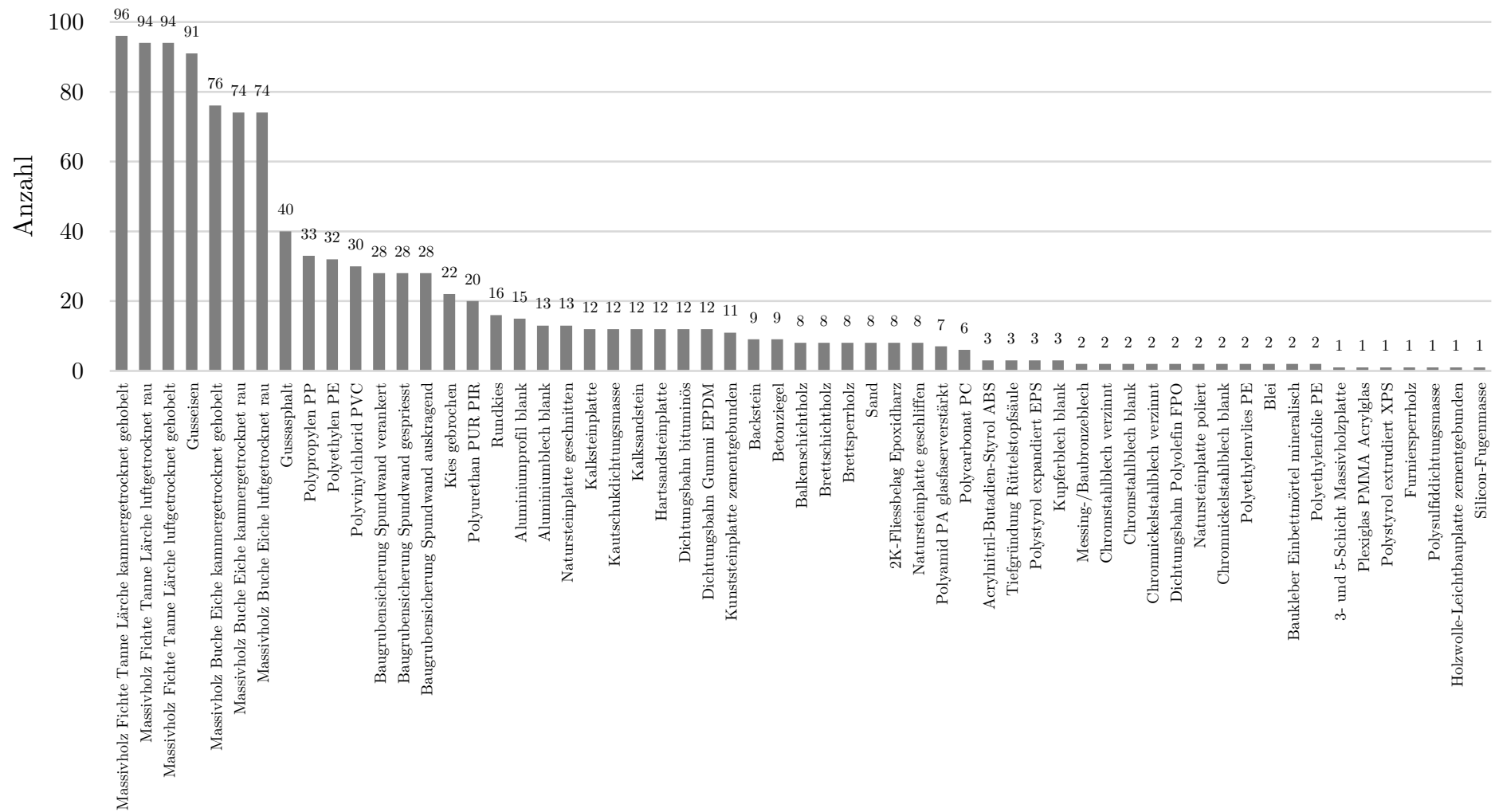
A.9 Verteilungen der Trainingsdaten

Verteilung der Ifc-Entities in den Trainingsdaten

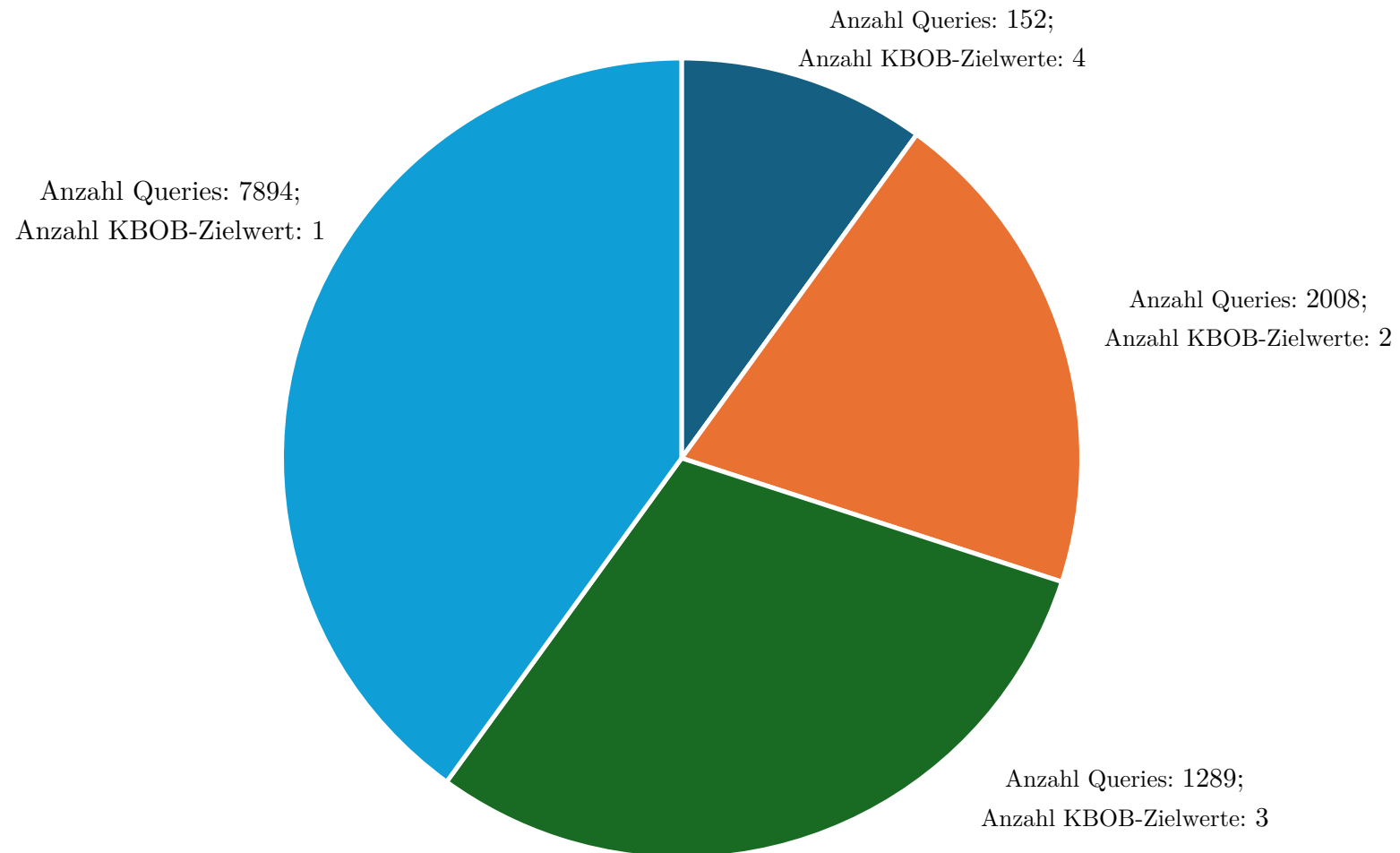




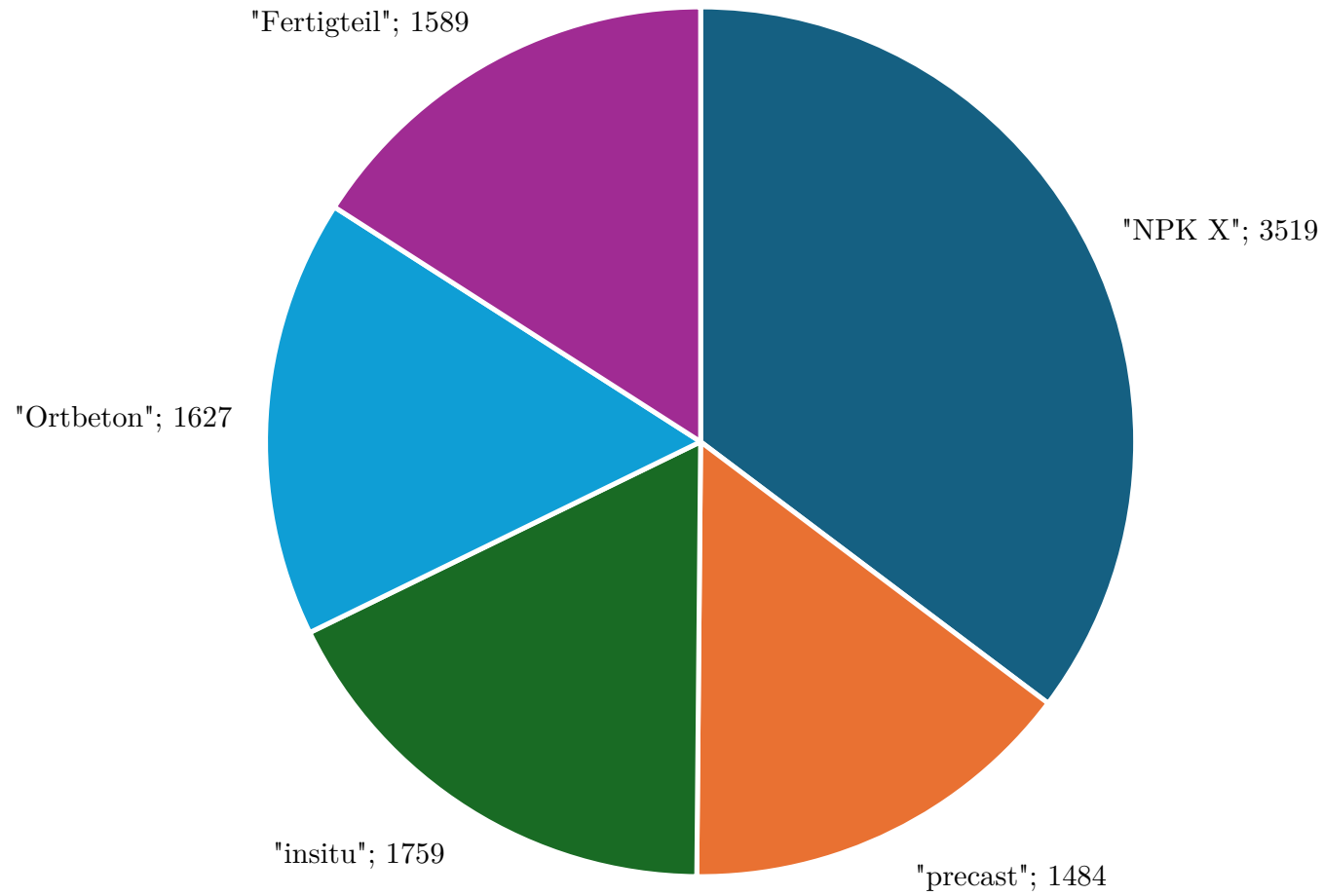
Verteilung der KBOB-Einträge in den Trainingsdaten, Anzahl < 100



Verteilung der Anzahl KBOB-Zielwerte der Trainingsdaten



Verteilung der Trainingsdaten nach Eigenschaften



A.10 Mindestanforderungen an das IFC-Modell

Die folgende Übersicht fasst zusammen, was im IFC-Modell enthalten sein sollte, damit der Prototyp eine belastbare Zuordnung zu KBOB-Einträgen (obere 6 Zeilen & Länge) und eine anschließende Umweltberechnung (untere 3 Zeilen) durchführen kann. Die Anforderungen leiten sich aus den Ergebnissen der Robustheitstests (Abschnitt 7.3) und der Anwendung auf reale Projekte (Abschnitt 7.5) ab.

Merkmal	Begründung
IfcEntity (spezifisch)	Ermöglicht die fachliche Einordnung (z. B. IfcBearing statt IfcBuildingElementProxy).
PredefinedType	Ermöglicht Unterscheidung innerhalb einer Entität (z. B. BORED vs. DRIVEN bei IfcPile).
ObjectType	Wird verwendet wenn PredefinedType als USERDEFINED definiert ist.
Name	Für die Unterscheidung von Baugrubensicherungen als IfcWall (Rühlwand, Schlitzwand, Spundwand, Nagelwand, Bohrpfahlwand).
Materialbezeichnung Material.Name	Ohne Materialwort sinkt Zuordnungsgenauigkeit. Entscheidend für die semantische Kandidatensuche.
Stärkeklasse (C30/37) StrengthClass	Stärkster Einzelprädiktor. Schützt die Zuordnung selbst bei gleichzeitigen Schreibfehlern.
Herstellungsart (insitu/precast) CastingMethod	Unterscheidung Ort beton vs. Fertigteil im KBOB-Katalog.
Volumen (Net-/GrossVolume) Qto_*BaseQuantities	Voraussetzung für die Umweltberechnung (Bezugseinheit kg via Rohdichte).
Länge (Length) Qto_PileBaseQuantities	Relevant für IfcPile zur Unterscheidung der Tiefgründungseinträge. Aus Volumen & Länge wird der Durchmesser berechnet.
Oberfläche (NetSurfaceArea) Qto_*BaseQuantities	Für die Berechnung einer möglichen Verzinkung.

Empfehlung: Die in der Tabelle aufgeführten Merkmale sollten im BIM-Abwicklungsplan als Mindestanforderung für die Modellierung und den IFC-Export definiert werden.

A.11 Dashboard Prototyp

IFC-material mapping with AI

Elements in current IFC	Total Uploaded IFC-Files	Nr of different Materials in the IFC
—	10	—

Uploads AI-Mapping Charts

SBERT-Modell ?

Hygros-LCA/bge-m3-ifc-kbob-finetuned ▼

Aktives SBERT-Modell: Hygros-LCA/bge-m3-ifc-kbob-finetuned

IFC-Datei hochladen

 Drag and drop file here
Limit 200MB per file • IFC, IFCZIP

Browse files

Mapping berechnen

Vorhandene JSONL-Datei laden

test-2026-4.3.jsonl ▼

JSONL laden

Breite linke Spalte (%)

40

IfcBeam | HEA200 | S235JR

Stahlprofil blank (Score: 0.915)

Oberfläche: 1.76 m²

☒ Verzinkung annehmen

IfcBeam | LINTEL | HEA200 | S235JR (2 Elemente)

Stahlprofil blank (Score: 0.921)

Oberfläche: 1.76 m²

☒ Verzinkung annehmen

IfcBeam | JOIST | HEA200 | S235JR

Stahlprofil blank (Score: 0.916)

Oberfläche: 1.76 m²

☐ Verzinkung annehmen

IfcBuildingElementProxy | RAIL | SCHIENE_S30 | Stahl (2 Elemente)

Stahlprofil blank (Score: 0.960)

⚠ Keine Oberflächenangabe (NetSurfaceArea / GrossSurfaceArea) im IFC vorhanden. Verzinkung kann nicht berechnet werden.

IfcColumn | COLUMN | 200*200 | Beton | Ortbeton

Tiefbaubeton (Score: 0.961)

⚠ Keine Bewehrung modelliert. Kontrolliere die Annahme des Bewehrungsgehalts:

☐ Bewehrung annehmen

Bewehrungsgehalt (kg/m³)

150.00

🗑

+

↓

✎

📄

🖱

🔄

✂

✂

📏

=

🔍

👁

↶

🕒

📶

🏠

🔄

🕒

📶

☀

🌙

?

📄

IfcBeam

✎

06g0K4kI9EGv3jqtF2...

📄

🔄

WORLD

1

🔍

Attributes

4

Name

Description

Tag

Predefined...

HEA200

S235JR

NOTDEFINI

📄

📄

🔍

Pset_BeamCom...

1

Status

NEU

Qto_BeamBaseQ...

1

NetVolume

0,00766

Quantities

4

Ready

679 elements

3.0M tris

59 FPS

112 MB

WebGPU

v1.16.0

📁 Auswahl speichern

Übersicht

IfcEntity	PredefinedType	Name	Beschrieb	Durchmesser	Material KBOB	AI Score
IfcReinforcingBar	SHEAR	Bügel	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	MAIN	1. Lage	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	unterer Stab	None	nan	Armierungsstahl	0.905
IfcBeam	EDGEBEAM			nan	Tiefbaubeton	0.962
IfcReinforcingBar	NOTDEFINED	Bügel	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	Bügel	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	Bügel	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	Längsstab	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	Längsstab	None	nan	Armierungsstahl	0.905
IfcReinforcingBar	NOTDEFINED	Längsstab	None	nan	Armierungsstahl	0.905

Chart Configuration

Grouping Mode

Group by Ifc Entity

Indikator-Familie

Umweltbelastungspunkte (UBP21)

Kennwerte

UBP21 Total (UBP) × UBP21 Herstellu... ×

Selection

IfcReinforcingBar × IfcBeam × IfcBuildingElem... × IfcColumn × IfcPile × IfcFooting × IfcRailing × IfcCovering × IfcSlab × IfcTendon × IfcWall × IfcBearing × IfcBuildingElem... × IfcReinforcingBa... × Verzinkung (ang... ×

Chart Type

Bar

Line

Pie

Bubble

Total Global Warming Potential

207'155 kg CO₂ eq

Total Environmental Impact Points

357'724'314 UBP

Total Primary Energy Non-Renewable

657'317 kWh oil...

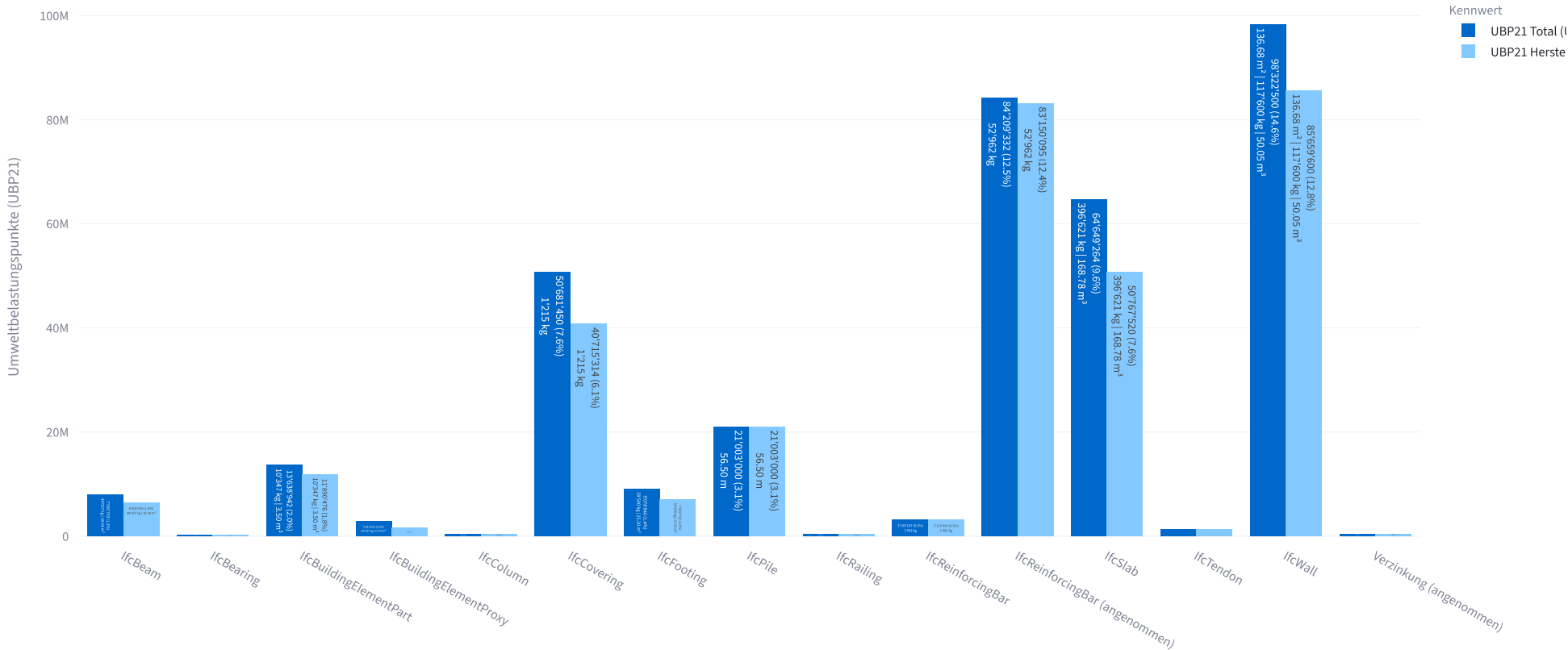
Total Volumen

286.78 m³

Total Masse

715'563 kg

Environmental Impact Visualization



Nicht berechenbare Einträge

Für die UBP-Berechnung braucht jedes Element eine passende Bezugsgrösse. Je nach KBOB-Eintrag wird ein Volumen [m³], eine Fläche [m²] oder eine Länge [m] benötigt. Fehlt diese Grundlage, kann kein Ergebnis berechnet werden.

Keine Einträge mit fehlender Berechnungsgrundlage gefunden.

Bewehrungsannahmen

Diese Beton-Bauteile haben kein modelliertes Bewehrungsseisen (IfcReinforcingBar/IfcTendon). Die Bewehrung wurde anhand des Bewehrungsgehalts (kg/m³) angenommen und als separate «Armierungsstahl»-Zeile in die UBP-Berechnung einbezogen.

GUID	Bauteilname	IfcEntity	Bewehrungsgehalt (kg/m³)	Quelle	Masse Stahl (kg)
3pb4qnJtnEThk7JQ8BrEmo		IfcColumn	150	default	105.6
1to9xqXY12rg_P\$tqDkoYb		IfcFooting	150	ifc	1080.0
0yGrfSVkr4aB04kO\$HM5ZC		IfcFooting	150	ifc	1080.0
2jKVau9jH9sfMEwXVySDFX		IfcFooting	150	ifc	1080.0
3bRM9PUSjD9vaQxTteRwCC		IfcSlab	150	default	116.2
3eDXlvU7zA5Be9Hd_Aiur6		IfcSlab	250	ifc	18750.0
1Yb32lwf9DAQpoDj8Qigno		IfcSlab	250	ifc	
0ngM4GUon1aw0cBfEb89eB	None	IfcWall	200	ifc	2500.0
3GdJhhVPv82hhvC\$4t9d4n	None	IfcWall	200	ifc	2500.0
11rE6Gr8D53uPLEpRCnJ8N	None	IfcWall	200	ifc	2500.0

Verzinkungsannahmen

Für diese Stahlbauteile wurde eine Verzinkung angenommen. Die Umweltbelastung wird anhand der Oberfläche (m²) und der Flächendichte (kg/m²) als separate «Verzinken»-Zeile in die Berechnung einbezogen.

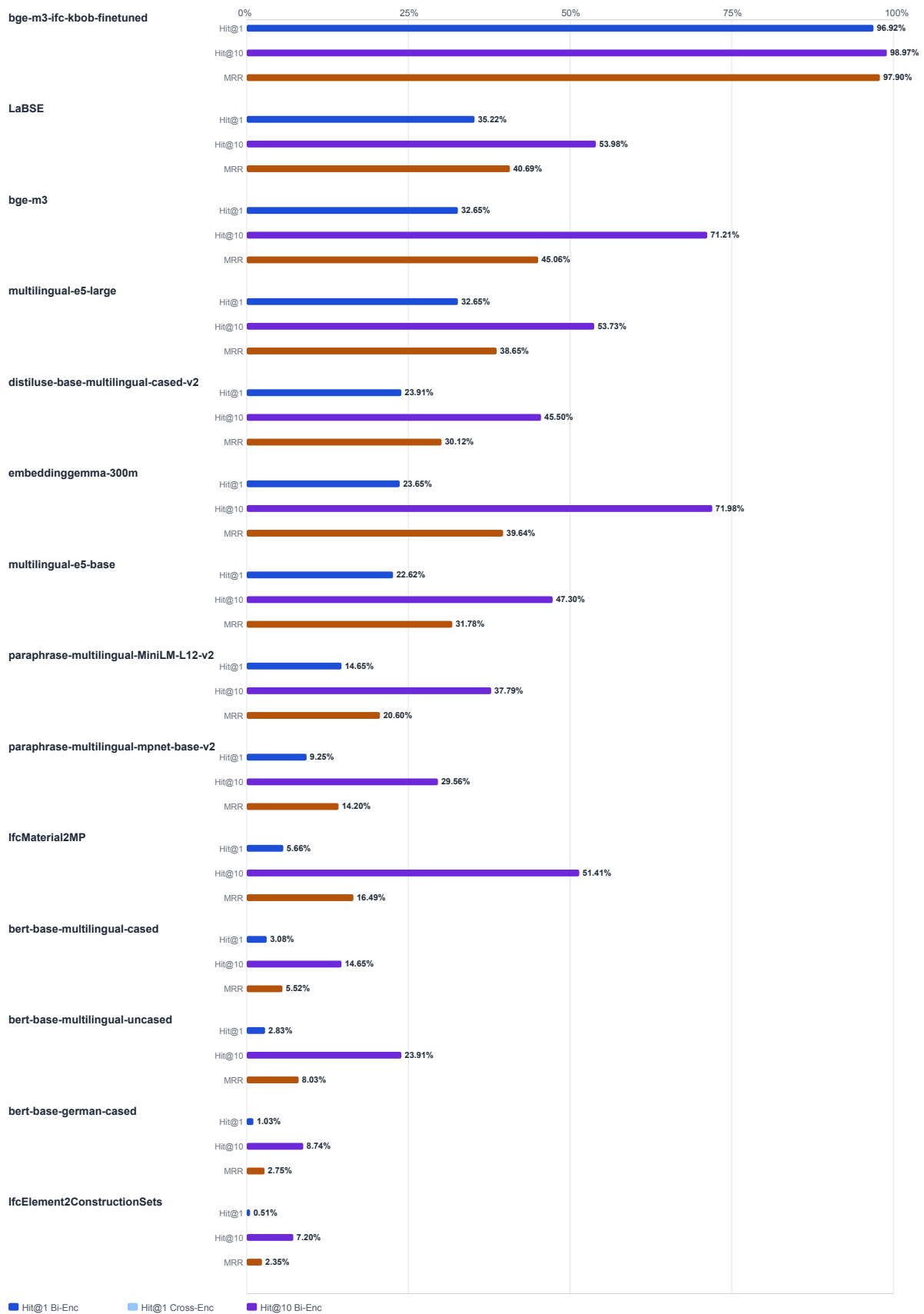
GUID	Bauteilname	IfcEntity	Oberfläche (m²)
06g0K4kl9EGv3jqtf2ipsO		IfcBeam	1.76
0SBObhWDTDK98JJHmBruUO		IfcBeam	1.76
1cyPtobFDChhb0VTizAecJ		IfcBeam	1.76
3j\$PL8Vij1ngjyU0rk7_gD		IfcRailing	2.14
0XrNmPcajBGgxgtsQvUC3f		IfcRailing	0.20
1f88r5io1EXf5wikw3\$dxh		IfcRailing	0.20
28imvfiLHEAfaQ0Lga3vln		IfcRailing	0.20
2CcrGgf0TDfx98\$2MXhnWF		IfcRailing	0.20
2btuyM0ND7tv0bbENxgQIA		IfcRailing	0.20
3SSADbaUPEEfZZ8SSItxuk		IfcRailing	0.20

Export Chart Data (csv)

Export All Data (csv)

A.12 Vergleich der Modelle mit den Test-Queries

Model Comparison — Bi-Encoder (Hit@1 / Hit@10 / MRR)



A.13 Validierung Prototyp

Berechnung von Ökobilanzwerten mit ECO₂nstruct:

Umweltbelastungsbericht für Kunstbauten



Total	Material Bauhauptgewerbe			230.12·10 ⁶ UBP
Abdichtungen				
	Dichtungsbahn bituminös	511.50 kg	4'354.40 UBP/kg	2.23·10 ⁶
Beton				
Magerbeton, NovoCon Sohlenbeton				
	Magerbeton, NovoCon Sohlenbeton, CEM 150 kg, CEM II/C-M (F-T)	3.60 m ³	167'923.00 UBP/m ³	604'522.80
Pumpbeton, NPK A - G				
	Pumpbeton, NPK G, C30/37, CEM II/B CH-Mix	259.69 m ³	382'343.00 UBP/m ³	99.29·10 ⁶
Gesteinskörnungen				
Grobe Gesteinskörnungen				
	Betonkies / Rundkies, 32/45 mm	28.54 t	59'400.00 UBP/t	1.70·10 ⁶
Metallbaustoffe				
Stabstähle				
	Bewehrungstahl B500B	16'310.00 kg	1'825.00 UBP/kg	29.77·10 ⁶
SBB				
SBB Ingenieurbau: Baugrubenabschlüsse				
	SBB Rühlwand mit Spritzbeton, verankert	103.20 m ²	347'500.00 UBP/m ²	35.86·10 ⁶
	SBB überschnittene Bohrpfehlwand, auskragend	33.48 m ²	1.06·10 ⁶ UBP/m ²	35.46·10 ⁶
SBB Ingenieurbau: Metalle				
	SBB Profilstahl, verzinkt	8.33 t	2.37·10 ⁶ UBP/t	19.74·10 ⁶
SBB: Schwellen				
	SBB Holzschwelle inkl. Befestigung	8.00 Stk	684'700.00 UBP/Stk	5.48·10 ⁶

Vergleich der Berechnung der Ökobilanzwerte zwischen Eco₂nstruct und dem Prototyp:

Werte Prototyp								
Eintrag KBOB	Masse	Volumen	Fläche	UBP/Masse	Rohdichte	UBP/m ³	UBP/m ²	UBP total
Dichtungsbahn bituminös	924	-	-	4350	1100	-	-	4'019'400
Magerbeton	-	3.6	-	104	2150	223'600	-	804'960
Tiefbaubeton	-	259.69	-	163	2350	383'050	-	99'474'255
Kies gebrochen	-	19.68	-	64.8	1450	93'960	-	1'849'419
Armierungsstahl	16310	-	-	1590	7850	-	-	25'932'547
Baugrubensicherung Rühlwand verankert	-	-	103.2	-	-	-	319'000	32'920'800
Baugrubensicherung Bohrpfahlwand unverankert	-	-	33.48	-	-	-	1'230'000	41'180'400
Stahlprofil blank	-	1.06	-	1290	7850	10'126'500	-	10'743'304
Verzinkung	21.0263	-	-	29500	-	-	-	620'276
Massivholz Buche Eiche luftgetrocknet rau	-	1.12	-	323	705	227'715	-	255'040

Werte Prototyp umgerechnet auf die gleiche Menge wie bei ECO ₂ nstruct								
Eintrag KBOB	Masse	Volumen	Fläche	UBP/Masse	Rohdichte	UBP/m ³	UBP/m ²	UBP total
Dichtungsbahn bituminös	511.5	-	-	4350	-	-	-	2'225'025
Magerbeton	-	3.6	-	-	-	223'600	-	804'960
Tiefbaubeton	-	259.69	-	-	-	383'050	-	99'474'255
Kies gebrochen	28.54	-	-	64800	-	-	-	1'849'419
Armierungsstahl	16310	-	-	1590	-	-	-	25'932'547
Baugrubensicherung Rühlwand verankert	-	-	103.2	-	-	-	319'000	32'920'800
Baugrubensicherung Bohrpfahlwand unverankert	-	-	33.48	-	-	-	1'230'000	41'180'400
Stahlprofil blank & Verzinkung	8.33	-	-	1'290'000	-	-	-	11'363'580
Massivholz Buche Eiche luftgetrocknet rau	-	1.12	-	323	-	-	-	255'040

Werte ECO ₂ nstruct							
Eintrag ECO ₂ nstruct	Masse	Volumen	Fläche	UBP/Masse	UBP/m ³	UBP/m ²	UBP total
Dichtungsbahn bituminös	511.5	-	-	4'354	-	-	2'230'000
Magerbeton, NovoCon Sohlenbeton, CEM 150 kg	-	3.6	-	-	167'923	-	604'523
Pumpbeton, NPK G, C30/37, CEM II/B CH-Mix	-	259.69	-	-	382'343	-	99'290'000
Betonkies / Rundkies, 32/45 mm	28.54	-	-	59'400	-	-	1'700'000
Bewehrungsstahl B500B	16310	-	-	1'825	-	-	29'770'000
SBB Rühlwand mit Spritzbeton, verankert	-	-	103.2	-	-	347'500	35'860'000
SBB überschnittene Bohrpfahlwand, auskragend	-	-	33.48	-	-	1'060'000	35'460'000
SBB Profilstahl, verzinkt	8.33	-	-	2'370'000	-	-	19'740'000
SBB Holzschwelle inkl. Befestigung	8	-	-	684'700	-	-	5'480'000

	Prototyp	ECO ₂ nstruct	
	UBP total	UBP total	Abweichung Begründung
Dichtungsbahn bituminös	2'225'025	2'230'000	0.22%
Magerbeton	804'960	604'523	-33.16% anderer Wert für UB P/m3
Tiefbaubeton	99'474'255	99'290'000	-0.19%
Kies gebrochen	1'849'419	1'700'000	-8.79% anderer Wert für UB P/kg
Armierungsstahl	25'932'547	29'770'000	12.89% anderer Wert für UB P/kg
Baugrubensicherung Rühlwand verankert	32'920'800	35'860'000	8.20% anderer Wert für UB P/m2
Baugrubensicherung Bohrpfahlwand unverankert	41'180'400	35'460'000	-16.13% anderer Wert für UB P/m2
Stahlprofil blank	11'363'580	19'740'000	42.43% anderer Wert für UB P/kg, keine Differenzierung der Verzinkung
Massivholz Buche Eiche luftgetrocknet rau	255'040	5'480'000	95.35% nicht vergleichbar da in ECO ₂ nstruct nur Stückzahl inkl. Befestigung wählbar
Gesamt	216'006'025	230'134'523	6.14%

Es zeigt sich, dass die Plattform ECO₂nstruct leicht andere Werte für die Berechnung verwendet als die der neusten KBOB-Liste. Insgesamt ist die Abweichung mit 6.1 % aber gering. Die Korrektheit der Berechnung mit dem Prototyp kann somit bestätigt werden.