

# **Erstellen einer standardisierten Material-Kostengliederung für Projekte einer Bausoftware mittels Natural Language Processing**

Florian Weidner

Fakultät für Informatik

WS 2022/23

IFC ist ein öffentlicher Standard im Bauwesen zur digitalen Beschreibung von Gebäudemodellen. In diesen Modellen können auch die Materialein der einzelnen Bauteile spezifiziert werden. Ziel ist es eine Kostengliederungsstruktur in der Bausoftware ORCA AVA aus den Materialieninformationen einer IFC-Datei zu generieren. Materialien können an verschiedenen Stellen im Modell angegeben werden. Ein Algorithmus soll diese Möglichkeiten in ein Interface zusammenführen. Außerdem handelt es sich bei der Materialangabe um ein Freitextfeld. Hier soll mit Hilfe von Natural Language Processing und Artificial Intelligence eine Lösung entwickelt werden, um eine standartisierte Liste der Materialien zu erschaffen.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Ausgangssituation . . . . .	4
1.2	Motivation . . . . .	5
1.3	Methodik . . . . .	5
1.3.1	Wissenschaftliche Vorgehensweise . . . . .	5
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Projektmanagement . . . . .	6
2.1.1	Vorgehensmodell . . . . .	6
2.1.2	DevOps . . . . .	6
2.2	IFC Format . . . . .	7
2.2.1	Verwendung . . . . .	7
2.2.2	Geschichte . . . . .	8
2.2.3	Dateiformat . . . . .	8
2.2.4	Baustoffe in IFC-Dateien . . . . .	8
2.3	Möglichkeiten für die Materialangabe eines Bauteils . . . . .	8
2.4	Das Format Kostengliederung in der ORCA AVA . . . . .	8
<b>3</b>	<b>Problemstellung und Anforderungen</b>	<b>10</b>
3.1	Problemstellung . . . . .	10
3.2	Anforderungen . . . . .	10
3.2.1	Funktionale Anforderungen . . . . .	10
3.2.2	Weitere Anforderungen . . . . .	10
3.2.3	Ziele . . . . .	10
<b>4</b>	<b>Theoretische Konzeption für die Erstellung der Material-Kostengliederung</b>	<b>10</b>
4.1	Textklassifizierungsalgorithmus ... . . . .	10
4.2	WordNet ... . . . .	10
4.3	... . . . .	10
<b>5</b>	<b>Gegenüberstellung der möglichen Konzepte</b>	<b>10</b>
5.1	Messkriterien . . . . .	10
5.2	Vergleich der Konzepte . . . . .	10
5.3	Festsetzen eines Algorithmus . . . . .	10

<b>6</b>	<b>Praktische Umsetzung</b>	<b>10</b>
6.1	Zusammenfassen der Materialschnittstelle einer IFC Datei . . . . .	10
6.1.1	Entwurf des Algorithmus . . . . .	10
6.1.2	Implementieren des Algorithmus . . . . .	10
6.2	Standardisierung der Materialnamen . . . . .	10
6.2.1	Nutzen von Artificial Intelligence . . . . .	10
6.2.2	Erstellen einer Datengrundlage . . . . .	10
6.2.3	Implementierung der Standardisierung . . . . .	10
6.3	Erstellen einer Kostengliederung . . . . .	10
6.3.1	Implementieren . . . . .	10
<b>7</b>	<b>Maßnahmen zur Qualitätssicherung</b>	<b>10</b>
7.1	Clean Code . . . . .	10
7.2	Technische Hilfsmittel . . . . .	10
7.3	Tests und Abnahme . . . . .	10
<b>8</b>	<b>Abschluss</b>	<b>10</b>
8.1	Bewertung der praktischen Umsetzung . . . . .	10
8.2	Fazit . . . . .	10
8.3	Ausblick . . . . .	10
<b>A</b>	<b>Erster Abschnitt des Anhangs</b>	<b>12</b>

# 1 Einleitung

Inhalt dieser wissenschaftlicher Arbeit ist das automatische Erstellen einer Material-Kostengliederung aus einer Industry Foundation Classes (IFC)-Datei mit Hilfe von Natural Language Processing (NLP) für eine Projekt einer Bausoftware. Die Bausoftware ist die ORCA AVA aus dem mittelständigen Softwarehaus „ORCA Software GmbH“ aus Neubeuern. In diesem Kapitel soll eine kurze Einführung über die „ORCA Software GmbH“ und das Produkt ORCA AVA gegeben werden. Außerdem wird die Motivation für das Feature und die wissenschaftliche Vorgehensweise dieser Arbeit beschrieben.

## 1.1 Ausgangssituation

Die im Titel beschriebene Bausoftware ist die ORCA AVA aus dem Softwarehaus „ORCA Software GmbH“. Dieses wurde im Jahr 1990 von Dipl.-Ing. Siegfried Tille und Dipl.-Ing. Heinz Nießen gegründet. Der Hauptsitz des Unternehmens ist in Neubeuern, bei Rosenheim. Das Unternehmen ist auf die Produktentwicklung von Software für die Baubranche spezialisiert. Im Vordergrund stehen die Ausschreibungssoftware ORCA AVA und die Ausschreibungstext-Plattform AUSSCHREIBEN.DE. Ziel der Entwicklung ist es die Ausschreibung, Vergabe und Abrechnung (AVA) eines Bauvorhabens für Planer, Architekten und Bauingenieure zu vereinfachen. Der Leitfaden ist, dass die Software soll für jeden verständlich und intuitiv zu bedienen sein. Diese Arbeit fokussiert sich auf eine Erweiterung der ORCA AVA. Sie ist für alle Architektur- und Ingenieurbüros, Wohnungsbaugesellschaften, Unternehmen und Behörden zur einfachen Abwicklung von Bauprojekten mit Ausschreibung, Vergabe und Abrechnung. Zusätzlich bildet sie das Kostenmanagement von solchen Projekten ab. Die Software ist außerdem Building Information Modeling (BIM) fähig und bietet Deutsches Institut für Normung e. V. (DIN) zertifizierte Schnittstellen für den Datenaustausch an. Es stehen drei verschiedene Editionen zur Verfügung. Die ORCA AVA Starter Edition (SE), die ORCA AVA Professional Edition (PE) und die ORCA AVA Enterprise Edition (EE). Die aktuellste Version ist die 25.0.

Technisch wird die ORCA AVA in .NET entwickelt. Ein Großteil der Anwendung besteht noch aus Visual Basic (VB) Code. Alle neuen Komponenten und Erweiterungen werden in C# implementiert. Neue Graphical User Interface (GUI)-Komponenten werden dementsprechend mit Windows Presentation Foundation (WPF) entwickelt. WPF ist ein .NET Framework für das Erstellen von Windows Applikationen mit graphischer Benutzeroberfläche von Microsoft. (Microsoft 2022) Die ORCA AVA und der IFC Manager (siehe Abschnitt 2.2.1) laufen in eigenen Prozessen und kommunizieren auf der

Prozessebene auf der lokalen Umgebung.

## 1.2 Motivation

BIM bedeutet, dass die Planung von Bauprojekten vollständig auf digitaler Basis durchgeführt wird. Für jeden Projektbeteiligten besteht somit Zugriff auf alle projektrelevanten Daten über Kosten, Mengen und Zeitabläufen. Somit können Baukosten einfacher ermittelt und der Bauprozess besser überwacht werden. In Abbildung 1 ist zu sehen, dass BIM ein Relevanter Begriff in der Baubranche ist. Die Verwendung der Praktik ist allerdings noch nicht sehr verbreitet (Thomas Baumanns, Dr. Philipp-Stephan Freber, Dr. Kai-Stefan Schober, Dr. Florian Kirchner 2016, p. 20). Mit dem „IFC First“ Ansatz ist das langfristige Ziel der ORCA AVA das Thema BIM noch mehr abzudecken. Die Ausschreibung soll automatisch aus dem 3D-Gebäudemodell entstehen. Bauteile aus dem Gebäudemodell sollen mit Kurztext, Langtext, Menge, Preis und vordefinierten Kostengliederungen in den Programmteil Bauelemente der ORCA AVA übernommen werden. Das ganze ist ein gutes Werbemittel für den Verkauf der Software, da die Relevanz des Themas so hoch ist und somit viele Kunden anspricht. Ein erster Teil des langfristigen Ziels, ist die Übernahme der Baustoffe aus der IFC-Datei. Die Übernahme von weiteren Daten aus dem Gebäudemodell können darauf aufgebaut werden. //todo: Kanomodell

Ein weiterer Trendbegriff, der mit dem Feature abgedeckt werden soll ist „künstliche Intelligenz“ und „Maschinelles Lernen“. Das Suchinteresse der beiden Themen sind in Abbildung ?? zu sehen. Die Werte geben das Suchinteresse relativ zum höchsten Punkt im Diagramm an. Der Begriff „Maschinelles Lernen“ hat seit Anfang 2015 eine steigendes Suchinteresse. Bei „künstliche Intelligenz“ ist das Suchinteresse von Anfang 2017 bis Ende 2021 konstant hoch. Der starke Anstieg ab November 2022 hängt wahrscheinlich mit der Veröffentlichung der Software ChatGPT zusammen. Man erkennt das das Interesse über die letzten Jahre hoch ist.

## 1.3 Methodik

### 1.3.1 Wissenschaftliche Vorgehensweise

## 2 Grundlagen

In diesem Kapitel werden die Grundlagen im Bereich der Fachspezifischen Themen, Technik und Projektmanagement vermittelt. Dies sollte Verständnis für die folgenden

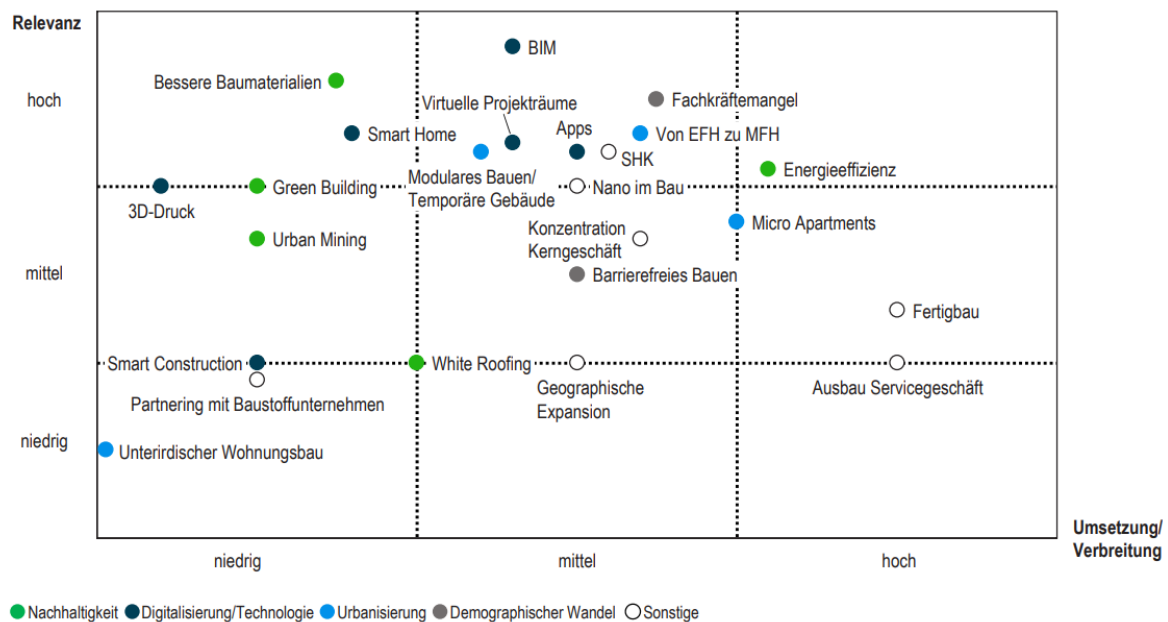


Abbildung 1: Megatrends Nachhaltigkeit und Digitalisierung in der Bauwirtschaft. Bilder aus (Thomas Baumanns, Dr. Philipp-Stephan Freber, Dr. Kai-Stefan Schober, Dr. Florian Kirchner 2016, p. 20)

Kapitel schaffen. Zuerst geht es um das Projektmanagement. Anschließend geht es um das Format IFC, dessen Geschichte und dessen Nutzungsmöglichkeiten für diese Arbeit. Außerdem wird die Struktur und der Nutzen einer Kostengliederung in der ORCA AVA veranschaulicht.

## 2.1 Projektmanagement

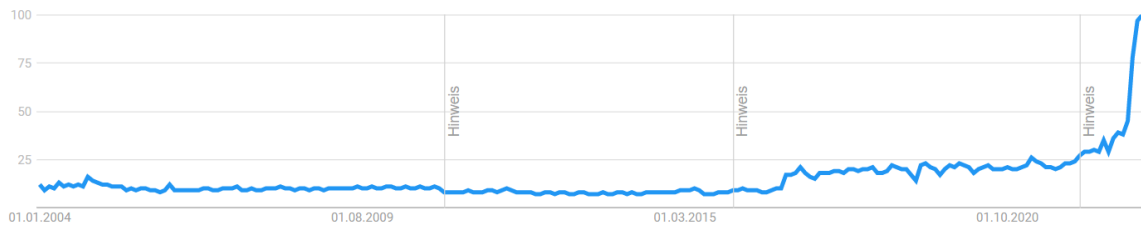
Bevor auf technische und fachliche Aspekte von IFC und Kostengliederungen eingegangen wird, folgt die Einführung in das Projektmanagements mit SCRUM und Development Operations (DevOps).

### 2.1.1 Vorgehensmodell

Das Projekt wurde mit Hilfe agiler Softwareentwicklung durchgeführt. Die durchlaufenden Sprints sind zwei Wochen lang. Funktionale Anforderungen werden vom Projektmanagement gestellt.

### 2.1.2 DevOps

Die Bezeichnung DevOps vereint die beiden Praktiken „Development“ (Entwicklung) und „Operations“ (Vorgänge). Die traditionelle Trennung von Entwicklung und Soft-



(a) Suchinteresse: Künstliche Intelligenz



(b) Suchinteresse: Machine Learning

Abbildung 2: Google Suchinteresse der beiden Begriffe „künstliche Intelligenz“ und „Maschinelles Lernen“ seit 2004

warebetrieb führt oft zu Interessenskonflikten. Entwickler wollen stetig die Software verbessern, der Betrieb will Änderungen vermeiden um die Stabilität des Systems zu gewährleisten. Durch DevOps entsteht ein Softwareentwicklungsprozess, den man durch Praktiken wie Continuous Integration, Continuous Delivery, Continuous Deployment, automatisiertes Testen, Infrastructure-as-Code und automatische Veröffentlichungen beschleunigt. DevOps steht auch für eine Entwicklungskultur mit offener Zusammenarbeit, Kommunikation, Transparenz und Eingestehen von Fehlern, um Konflikte im Team zu vermeiden. Im Entwicklungsteam der ORCA AVA wird diese Praktik umgesetzt. Die Technischen Hilfsmittel, die für den DevOps Prozess verwendet werden, sind in Abschnitt 7.2 (Beetz & Harrer 2021)

## 2.2 IFC Format

Die Daten für die Material-Kostengliederung werden aus einem digitalem Gebäudemodell entnommen. Der öffentliche internationale Standard für Gebäudemodelle ist IFC. (buildingSMART International Ltd. 2017) Dieser wird auch in der bestehenden ORCA AVA benutzt um den Ausschreibungsprozess zu unterstützen. IFC Dateien können geöffnet, angeschaut und Informationen über das Modell in die Hauptsoftware übernommen werden.

### **2.2.1 Verwendung**

Die ORCA AVA kann IFC-Dateien einlesen. Im IFC Manager wird das 3D-Modell dann angezeigt. In dem geöffneten Fenster gibt es dann einige fachliche Ansichten, die die IFC-Daten nochmal abstrahieren. Es können bestimmte Mengen oder die Anzahl verschiedener Bauteile, die aus dem IFC-Modell berechnet werden, in die ORCA AVA übernommen werden. Es werden auch alle Eigenschaften eines Bauteils in der Eigenschaften-Ansicht angezeigt. Hier kann man auch die Materialbezeichnung des Bauteils finden, die im Modell hinterlegt ist.

Für das Arbeiten mit IFC-Dateien wird die open-source Bibliothek `xbim-toolkit` verwendet. Die .NET Bibliothek kann IFC-Dateien lesen, schreiben und anzeigen. Außerdem unterstützt es bei der Berechnung von komplexer Geometrie, um die Modelldaten für Analysen nutzbar zu machen. Seit 2009 wird das Projekt in Zusammenarbeit mit der *Norhumbria University* weiterentwickelt. Mittlerweile bildet es die Standards IFC Version 2.3 (IFC2x3) und IFC Version 4 (IFC4) zu 100% ab. Außerdem bietet es an, auch IFC2x3 Modelle über das IFC4 Interface anzuprogrammieren. Somit können mit einer Codebasis beide Formate abgebildet und unterstützt werden. (?)

### **2.2.2 Geschichte**

1994 startete die Entwicklung an dem offenen Datenmodellstandard IFC. Dieser sollte die Anforderungen der Industrie an Interoperabilität gerecht werden und eine gemeinsame Basis zum Austausch von Informationen durch verschiedenen Anwendungen geschaffen werden. Mit BIM sollten Daten lesbar, editierbar für verschiedene Systeme durch den Bauprozess und kompletten Lebenszyklus eines Gebäudes geteilt werden. (Laakso & Kiviniemi 2012)

### **2.2.3 Dateiformat**

IFC ist ein Implementierungs-Unabhängiges Datenmodell, welches in verschiedenen Umgebungen benutzt werden kann. Es kann beispielsweise in eine relationales Datenbankschema gegossen werden oder auch als Dateiformat implementiert werden. (Laakso & Kiviniemi 2012)

### **2.2.4 Baustoffe in IFC-Dateien**

Das IFC Modell bietet auch Materialangaben zu verschiedenen Bauteilen an. (?)



## 2.3 Möglichkeiten für die Materialangabe eines Bauteils

## 2.4 Das Format Kostengliederung in der ORCA AVA

Die Kostengliederung bietet eine Struktur, um Gesamtkosten einer Baumaßnahme in Kostengruppen unterteilt auswerten zu können. Logisch zusammengehörende Kosten können so in eine Kostengruppe (KG) zusammengerechnet werden. Außerdem ist der Aufbau eine Baumstruktur, wodurch Kostengruppen hierarchisch addiert werden können. Bei einem Bauprojekt kann man so Kosten über alle Projektphasen vergleichen. Von der Kostenschätzung über die Ausschreibung bis zur Rechnungsfreigabe. So können Kostenauswertungen nach den verschiedenen Kostengruppen durchgeführt werden. In einem ORCA AVA Projekt kann man verschiedene Kostengliederungen definieren. Es existieren bereits Standardkostengliederungen beim Erstellen eines neuen Projektes. Zusätzlich können neue Kostengliederungen erstellt oder importiert werden. (*Auswählen eines Algorithmus für maschinelles Lernen - Azure Machine Learning* n.d.)

Technisch ist eine Kostengliederung als Modell im C# Code definiert. Der Aufbau bildet die Baumstruktur über eine Referenz zur ParenNode und mehreren ChildrenNodes ab. Das Model kann dann über die ORCA AVA interne Middleware in der Datenbank persistiert werden. Die persistierten Kostengliederungen werden dann über VB-Teil der Anwendung im entsprechenden Programmteil angezeigt. So müssen nach der generierten Kostengliederungs-Struktur die strukturierten Materialbezeichnungen in das C# Modell abgebildet werden. So wird nach dem Import die erstellte Kostengliederung automatisch in der Oberfläche angezeigt und kann für Kostenauswertungen benutzt werden.

## **3 Problemstellung und Anforderungen**

### **3.1 Problemstellung**

### **3.2 Anforderungen**

#### **3.2.1 Funktionale Anforderungen**

#### **3.2.2 Weitere Anforderungen**

#### **3.2.3 Ziele**

## **4 Theoretische Konzeption für die Erstellung der Material-Kostengliederung**

### **4.1 Textklassifizierungsalgorithmus ...**

### **4.2 WordNet ...**

### **4.3 ...**

## **5 Gegenüberstellung der möglichen Konzepte**

### **5.1 Messkriterien**

### **5.2 Vergleich der Konzepte**

### **5.3 Festsetzen eines Algorithmus**

## **6 Praktische Umsetzung**

### **6.1 Zusammenfassen der Materialschnittstelle einer IFC Datei**

#### **6.1.1 Entwurf des Algorithmus**

#### **6.1.2 Implementieren des Algorithmus**

### **6.2 Standardisierung der Materialnamen**

#### **6.2.1 Nutzen von Artificial Intelligence**

#### **6.2.2 Erstellen einer Datengrundlage**

#### **6.2.3 Implementierung der Standardisierung**

### **6.3 Erstellen einer Kostengliederung**

#### **6.3.1 Implementieren**

**IFC** Industry Foundation Classes

**IFC2x3** IFC Version 2.3

**IFC4** IFC Version 4

**BIM** Building Information Modeling

**KG** Kostengruppe

**DevOps** Development Operations

**NLP** Natural Language Processing

**AVA** Ausschreibung, Vergabe und Abrechnung

**BIM** Building Information Modeling

**DIN** Deutsches Institut für Normung e. V.

**SE** Starter Edition

**PE** Professional Edition

**EE** Enterprise Edition

**VB** Visual Basic

**GUI** Graphical User Interface

**WPF** Windows Presentation Foundation

## **A Erster Abschnitt des Anhangs**

In diesem Anhang wird ...

## Literatur

*Auswählen eines Algorithmus für maschinelles Lernen - Azure Machine Learning* (n.d.).

Accessed: 2023-1-18.

**URL:** <https://learn.microsoft.com/de-de/azure/machine-learning/how-to-select-algorithms>

Beetz, F. & Harrer, S. (2021), ‘Gitops: The evolution of devops?’, *IEEE Software* **39**.

buildingSMART International Ltd. (2017), ‘IFC4 documentation’. Accessed: 2023-1-31.

**URL:** [http://standards.buildingsmart.org/IFC/RELEASE/IFC4\\_1/FINAL/HTML/](http://standards.buildingsmart.org/IFC/RELEASE/IFC4_1/FINAL/HTML/)

Laakso, M. & Kiviniemi, A. O. (2012), ‘The IFC standard: A review of history, development, and standardization, information technology’, *J. Inf. Technol. Constr.* **17**(9), 134–161.

Microsoft (2022), ‘Windows presentation foundation (WPF)’. Accessed: 2023-2-10.

**URL:** <https://github.com/dotnet/wpf>

Thomas Baumanns, Dr. Philipp-Stephan Freber, Dr. Kai-Stefan Schober, Dr. Florian Kirchner (2016), *Bauwirtschaft im wandel - trends und potenziale bis 2020*, Technical report.