

Generierung von Feedback-Regeln für Modelle aus annotierten Musterlösungen

Institute for Computer Science and Business Information Systems (ICB) at
University Duisburg-Essen, Campus Essen

vorgelegt von

Michael Vivian Mboni Saha

michael.mboni-saha@stud.uni-due.de

<https://michael-mboni.dev/>

Matrikelnummer: 3154361

Masterarbeit

zur Erlangung des akademischen Grades

Master of Science (M.Sc.)

Betreuer:	Dr. Michael Striewe
Hochschullehrer:	Prof. Dr. Michael Goedicke
Studiengang:	M.Sc. Software and Network Engineering
Studiensemester:	Wintersemester 2023/2024
Datum:	30.03.2024

Inhaltsverzeichnis

1. Einleitung	10
1.1. Motivation	10
1.2. Zielsetzung und Abgrenzung	12
1.3. Aufbau der Arbeit	12
2. Hintergrund	14
2.1. E-Assessment-Systeme	14
2.1.1. Definition und Vorteile	15
2.1.2. Mögliche Aufgabentypen	17
2.2. Konzeptionelle Modelle	20
2.3. Automatisierte Bewertung	22
2.3.1. Definition und Abgrenzung	22
2.3.2. Automatisierte Bewertungsmethoden für UML	23
2.3.3. Herausforderung der automatisierten Bewertung	25
2.4. State of the Art	26
2.4.1. Machine learning zur Bewertung von UML	26
2.4.2. Bewertung auf Basis von Ähnlichkeitsmaßen	28
2.4.3. Graph matching zur Bewertung von UML	29
2.4.4. Automatisierte Bewertung mit GReQL	31
2.5. JACK	33
3. Problemanalyse	36
3.1. Beschreibung des Bewertungsprozesses	36
3.2. Untersuchung des Problems	37
3.3. Ableitung und Abgrenzung der Anforderungen	39
4. Entwicklung des Konzepts	41
4.1. Verschiedene Herangehensweisen	41
4.1.1. YAML-basierten Annotationen	41
4.1.2. Verwendung von Natural Language Processing Tools	43
4.2. Konzept	44

5. Implementierung	47
6. Evaluation	48
7. Diskussion	49
8. Zusammenfassung und Ausblick	50
Literatur	51
A. Textbausteine	56
A.1. Erster Baustein	56

Abbildungsverzeichnis

2.1. Ansatz der Studie [10].	27
2.2. Beispiel für einen Graphenabgleich [4].	30
4.1. Beispiel einer Annotation von UML mit YAML	42
4.2. Repräsentatives schema des konzepts.	46

Tabellenverzeichnis

Glossar

E-Assessment-Systemen auch als elektronische Bewertungssysteme bezeichnet, sind computergestützte Technologien und Plattformen, die entwickelt wurden, um den Prozess der Bewertung und Beurteilung von akademischen Leistungen, Prüfungen und Aufgaben zu unterstützen oder zu automatisieren. 10, 13

formative Bewertung ist ein kontinuierlicher Prozess, der während des Lernens stattfindet. Ihr Hauptziel ist es, den Fortschritt der Schüler zu verfolgen, ihre Bedürfnisse zu identifizieren und ihnen bei der Verbesserung ihrer Leistung zu helfen.. 15

Geschlossene Fragen Diese Fragen zeichnen sich durch ihre Fähigkeit aus, eine begrenzte Auswahl vorab definierter Antwortmöglichkeiten anzubieten, aus denen die Lernenden die adäquate Antwort auswählen müssen. 17

JACK ist ein e-Assessment System, das von der Universität Duisburg-Essen entwickelt wurde. Es handelt sich um eine webbasierte Plattform, die es Lehrkräften ermöglicht, Online-Bewertungen wie Quiz, Prüfungen und Umfragen zu erstellen und durchzuführen. Jack bietet eine Vielzahl von Funktionen, die den Lehrkräften bei der Verwaltung ihrer Prüfungen helfen, wie z. B. Benotung, Berichterstattung und Verfolgung des Lernfortschritts. [29]. 12, 13

konzeptuellen Modellen sind abstrakte Darstellungen oder konzeptionelle Strukturen, die darauf abzielen, Ideen, Beziehungen, Konzepte oder Entitäten eines bestimmten Bereichs zu beschreiben, ohne in konkrete Details oder Implementierungsdetails einzugehen.. 12

Offene Fragen Im Kontrast zu geschlossenen Fragen präsentieren sie keine vordefinierten Antworten und gestatten den Lernenden, ihre Gedanken selbstständig zu artikulieren. 18

summative Bewertung ist eine Art von Bewertung, die am Ende eines definierten Lernzeitraums durchgeführt wird, wie z.B. am Ende eines Kurses, eines Moduls oder eines Schuljahres.. 15

Abkürzungsverzeichnis

OPAL Open Academic Learning Platform

UML Unified Modeling Language

UDE Universität Duisburg Essen

GReQL Graph Repository Query Language

XMI XML Metadata Interchange

UXF UML eXchange Format

NLP Natural Language Processing

Abstract

Englishce Zusammenfassung hier. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquid ex ea commodo consequat. Quis aute iure reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint obcaecat cupiditat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1. Einleitung

Das erste Kapitel dieser Arbeit befasst sich mit der grundlegenden Motivation und definiert dann das allgemeine Problem, das gelöst werden soll. Es folgt eine Erläuterung der Zielsetzung und zum Schluss wird der Aufbau der Arbeit im Detail beschrieben.

1.1 Motivation

Nach den Prüfungsphasen in Bildungseinrichtungen auf Sekundarstufen und Universitätsniveau manifestiert sich regelmäßig die Herausforderung, der zahlreichen Korrekturaufgaben, welche Lehrer und Dozenten zu bewältigen haben. Diese Aufgabe kann mitunter mühsam und zeitaufwendig sein, insbesondere im Falle einer hohen Anzahl an Prüfungen [22]. Die Evaluation von Prüfungsleistungen von 20 Prüflingen mag noch als vertretbar erscheinen, doch wie verhält es sich in Lehrveranstaltungen, in denen sich 400 oder gar mehr Studierende beteiligen? Bereits in den 1960er Jahren wurden Bestrebungen unternommen, solche Probleme durch die Entwicklung von E-Assessment-Systemen zu lösen [20].

Heute ist die automatisierte und computerbasierte Bewertung von akademischen Arbeiten weit verbreitet im Hochschulbereich. Diverse Plattformen wie Blackboard [8], Open Academic Learning Platform (OPAL) [15], und Moodle [27] wurden konzipiert, um den Prüfungsprozess zu rationalisieren und vollständig zu automatisieren. Diese Plattformen finden insbesondere in der Beurteilung geschlossener Fragen wie Multiple-Choice-Aufgaben Anwendung, bei denen die Antworten klar und die Bewertung einfach automatisiert werden kann. Die Frage, die sich jedoch stellt, ist: Wie gestaltet sich die Evaluation offener Fragenstellungen in Übungen, wie beispielsweise die Erstellung konzeptioneller Modelle, bei denen Lösungsansätze variieren können, jedoch dennoch korrekt sind?

In der allgemeinen Informatikbildung ist es üblich, Modelle zur Beschreibung der Architektur eines Systems oder abstrakterer Konzepte heranzu-

ziehen. Diese Modelle können in Form von Diagrammen wie Unified Modeling Language (UML)-Klassendiagrammen, Sequenzdiagrammen oder Zustandsdiagrammen präsentiert werden. Bei einer Bewertungsaufgabe, bei der ein Studierender ein Modell auf Grundlage eines gegebenen Textes erstellen soll, sieht sich der Evaluierende oft mit der Herausforderung konfrontiert, das Ergebnis des Studierenden mit der erwarteten Lösung abzugleichen. Diese Form der Beurteilung kann jedoch diverse Schwierigkeiten aufwerfen:

1. Erstens, in der Modellierung existiert keine eindeutige Lösungsstrategie, was die Aufgabe komplexer gestaltet. Die Resultate jedes Studierenden müssen wiederholt mit der vorgeschlagenen Lösung verglichen werden, um sicherzustellen, dass die Arbeit alle wesentlichen Komponenten der Lösung berücksichtigt.
2. Das zweite Hindernis in diesem Evaluationsmodell ist die inhärente Subjektivität. Einige Studierende könnten vollständig funktionierende Modelle entwickeln, die jedoch von der erwarteten Lösung abweichen, und könnten dadurch benachteiligt werden. Dies kann zu einer gewissen Starre im Unterricht führen und verhindern, dass Studierende originelle Ansätze verfolgen.
3. Eine weitere bedeutende Herausforderung dieses Evaluationsmodells ist die potenzielle Subjektivität der Evaluierenden. Jeder Korrektor könnte eine individuelle Auslegung dessen haben, was als korrekte Antwort betrachtet wird, was zu Unstimmigkeiten in der Benotung und einer ungleichen Verteilung von Noten zwischen den Arbeiten der Studierenden führen kann. Dies könnte besonders problematisch sein, wenn mehrere Evaluierende dieselben Arbeiten bewerten, da dies zu signifikanten Unterschieden in den vergebenen Noten führen kann.
4. Zuletzt kann dieses Evaluationsmodell auch beträchtlich zeitaufwendig sein, insbesondere in Lehrveranstaltungen mit einer hohen Anzahl an Studierenden. Ein ausführlicher Abgleich der Studierendenarbeiten mit der Referenzlösung erfordert Zeit und könnte zu Verzögerungen bei der Mitteilung der Ergebnisse an die Studierenden und einer erhöhten Arbeitsbelastung für die Evaluierenden führen.

In diesem Zusammenhang könnte die Implementierung einer automatisierten Bewertung auf Basis klar definierter und spezifischer Kriterien, die in jeder Lösung berücksichtigt sein müssen, eine optimale Lösung darstellen. Eine automatisierte Evaluierung durch eine computerbasierte Anwendung könnte sämtliche dieser Probleme lösen, indem sie eine rasche, objektive

und vorurteilsfreie Bewertung ermöglicht. Die automatische Evaluierung von konzeptuellen Modellen ist kein neues Forschungsthema. Umfangreiche Untersuchungen in dieser Richtung wurden bereits durchgeführt, und verschiedene Ansätze wurden entwickelt, um dieses Ziel zu realisieren.

1.2 Zielsetzung und Abgrenzung

Im Rahmen dieser Masterarbeit wird das Ziel verfolgt, ein Verfahren zur (halb-)automatischen Erstellung von Bewertungsregeln auf Basis annotierter Musterlösungen zu entwickeln und in Form einer Softwareanwendung zu prototypisieren. Dieses Verfahren soll dazu beitragen, den Prozess der Regeldefinition zu vereinfachen und zu optimieren.

In der Fakultät für Informatik an der Universität Duisburg Essen (UDE) wird ein E-Assessment-System namens "JACK" [29] verwendet, um Studierende automatisch bei bestimmten Prüfungen und Übungen zu bewerten. JACK ist in der Lage, verschiedene Arten von Aufgaben, sowohl geschlossene als auch offene Fragen, zu bewerten. Unter den verschiedenen Aufgabentypen fallen auch Aufgaben zur Erstellung von UML-Diagrammen. JACK wurde entwickelt, um die Lösungen der Studierenden zu bewerten und Noten zu vergeben. Bei der Bewertung von UML-Diagrammen verwendet JACK regelbasierte Ansätze. Die Lehrenden müssen die Regeln definieren, die das Diagramm des Studierenden validieren und ihm eine Note zuweisen. Dieser Prozess der Regeldefinition ist jedoch zeitaufwändig und birgt das Risiko, dass kleine, aber bedeutende Unaufmerksamkeitsfehler und Flüchtigkeitsfehler auftreten.

Das angestrebte Verfahren soll daher Lehrkräften dabei unterstützen, solche Bewertungsregeln effizienter und fehlerminimiert zu erstellen, indem es auf vorhandene annotierte Musterlösungen zurückgreift. Durch die Entwicklung einer Softwareanwendung, die diesen Prozess (halb-)automatisch durchführt, wird das Ziel verfolgt, eine zeit- und ressourcensparende Lösung bereitzustellen. Damit kann die Qualität der automatischen Bewertung von UML-Diagrammen in JACK verbessert und der Aufwand für die Lehrkräfte reduziert werden.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sechs Hauptkapitel, die jeweils einen spezifischen Aspekt des Forschungsgebiets beleuchten.

Im zweiten Kapitel werden die grundlegenden Konzepte und Informationen eingeführt, die für das Verständnis der Arbeit erforderlich sind. Dies umfasst eine Erklärung der E-Assessment-Systemen, konzeptioneller Modelle sowie der automatisierten Bewertung. Darüber hinaus werden verschiedene Ansätze zur automatisierten Bewertung im Detail betrachtet. Ein Überblick über den aktuellen Stand der Forschung auf diesem Gebiet wird ebenfalls gegeben, gefolgt von einer Einführung in das E-Assessment-System JACK.

Das dritte Kapitel widmet sich der detaillierten Analyse des zugrunde liegenden Problems. Hier wird der Prozess der Erstellung und Einreichung von Übungen sowie deren Bewertung durch die JACK-Plattform erläutert. Die spezifischen Herausforderungen und Probleme, die in dieser Arbeit adressiert werden, werden identifiziert und diskutiert.

Anschließend folgt das Kapitel vier, welches hauptsächlich das Konzept zur Bewältigung der zuvor im vorhergehenden Kapitel definierten Problematik behandelt. In diesem Kapitel wird eine theoretische Herangehensweise zur Problemlösung vorgestellt, und es wird ein methodischer Ansatz zur Problembewältigung skizziert. Die praktische Umsetzbarkeit dieser Lösungsstrategie wird im darauf folgenden Abschnitt ausführlich erörtert.

Das Kapitel fünf beschäftigt sich mit der praktischen Umsetzung der entwickelten Lösung. Hier werden die verwendeten Technologien und der Entwicklungsprozess vorgestellt. Die Gründe für die Auswahl bestimmter Technologien, Herangehensweisen und Implementierungsmethoden werden erläutert. Die Gesamtarchitektur des entwickelten Prototyps wird präsentiert und begründet.

Das sechste Kapitel widmet sich der Bewertung der entwickelten Lösung. Es werden Kriterien definiert, anhand derer der Prototyp bewertet wird. Der Evaluationsprozess wird erläutert, einschließlich der durchgeführten User Studies. Im Kapitel sieben werden die Ergebnisse der Arbeit kritisch diskutiert und interpretiert. Hierbei werden die Implikationen der Ergebnisse auf das Gesamtforschungsgebiet erörtert, eventuelle Limitationen der Arbeit aufgezeigt und mögliche Ansätze für zukünftige Forschung identifiziert.

Abschließend bietet Kapitel sechs eine Zusammenfassung der wichtigsten Erkenntnisse und Ergebnisse dieser Masterarbeit. Es werden auch Perspektiven für die Weiterentwicklung der vorgestellten Lösung sowie die Bereitstellung der vollständigen Dokumentation des entwickelten Tools dargelegt.

2. Hintergrund

Im vorliegenden Kapitel wird eine grundlegende Einführung in die zentralen Konzepte, Methoden und Ansätze gegeben, die für das Verständnis und die weiterführende Behandlung dieser Masterarbeit von Bedeutung sind. Zunächst erfolgt eine Erläuterung und Definition des Begriffs “E-Assessment-Systeme”, um eine klare Basis für die nachfolgende Diskussion zu schaffen. Im Anschluss wird die Thematik der “Konzeptionellen Modelle” ausführlich behandelt, wobei der Fokus auf den Modellen liegt, die in der Informatik allgemein angewandt werden. Darüber hinaus erfolgt eine umfassende Einführung in das Konzept der “Automatisierten Bewertung”, wobei insbesondere die verschiedenen Methoden und Ansätze, die in diesem Bereich von Relevanz sind, beleuchtet werden.

Des Weiteren wird ein Überblick über den aktuellen Stand der Technik (“State of the Art”) hinsichtlich der im Bereich der automatisierten Bewertung verwendeten Techniken und Tools gegeben. Schließlich erfolgt eine generelle Vorstellung des Tools “JACK”, welches im Rahmen dieser Arbeit eine herausragende Rolle spielt und in späteren Abschnitten ausführlicher behandelt wird.

Dieses Kapitel dient somit als Grundlage und Orientierungshilfe, um den Leser in die Thematik einzuführen und die notwendigen Begrifflichkeiten und Zusammenhänge zu vermitteln, die im weiteren Verlauf der Masterarbeit von zentraler Bedeutung sein werden.

2.1 E-Assessment-Systeme

Nach Eilers et al. sind E-Assessment-Systeme computergestützte Bildungstechnologien, die entwickelt wurden, um den Prozess der Bewertung von Lernleistungen in Bildungseinrichtungen zu automatisieren, zu verbessern und zu erweitern [12]. Diese Systeme ermöglichen die Erfassung, Bewertung und Analyse von Schüler- oder Studentenleistungen in einem digitalen Umfeld. E-Assessment-Systeme verwenden verschiedene Arten von

Aufgaben und Prüfungen, darunter Multiple-Choice-Fragen, Essays, Simulationen, interaktive Aufgaben und mehr, um das Wissen, die Fähigkeiten und die Kompetenzen der Lernenden zu bewerten.

2.1.1 Definition und Vorteile

Die Nutzung von E-Assessment-Systemen bietet eine Reihe von Vorteilen, die sich aus ihrer Fähigkeit zur Automatisierung ergeben. Diese Systeme verwenden vordefinierte Algorithmen und Kriterien, um die Leistung der Lernenden objektiv zu bewerten, was eine schnellere und effizientere Verarbeitung der Ergebnisse ermöglicht. Dies ermöglicht es Lehrern, mehr Zeit für pädagogische Aktivitäten aufzuwenden, anstatt sich mit manuellen Prüfungen und Aufgabenbewertungen zu befassen. Die Art der Verwendung eines E-Assessment-Systems hängt von den angestrebten Zielen ab [37]. Es handelt sich nicht nur um ein Werkzeug zur Bewertung von Schülern am Ende eines Semesters (summative), sondern sie können auch am Lernprozess der Schüler teilnehmen (formative).

Die summative Bewertung ist eine Art von Bewertung, die am Ende eines definierten Lernzeitraums durchgeführt wird, wie z.B. am Ende eines Kurses, eines Moduls oder eines Schuljahres. Ihr Hauptziel besteht darin, die Gesamtleistung der Schüler zu messen und festzustellen, inwieweit sie die zuvor festgelegten Lernziele erreicht haben [37]. Diese Art der Bewertung wird in der Regel nach Abschluss des Lernens durchgeführt, oft in Form einer Abschlussprüfung oder eines Abschlussprojekts. Die summative Bewertung wird hauptsächlich verwendet, um eine Note zu vergeben oder Entscheidungen wie die Promotion oder den Erwerb eines Abschlusses zu treffen. Ihr Fokus liegt weniger auf detailliertem Feedback an die Schüler als vielmehr auf der Bewertung ihres Kompetenzniveaus [37].

Im Gegensatz zur summative Bewertung handelt es sich bei der formativen Bewertung um einen kontinuierlichen Prozess, der während des Lernens stattfindet. Das Hauptziel ist es, den Fortschritt der Schüler zu verfolgen, ihre Bedürfnisse zu identifizieren und ihnen bei der Verbesserung ihrer Leistung zu helfen [18]. Formative Bewertungen werden regelmäßig während des Lernprozesses durchgeführt, oft in Form von Quiz, Übungen oder Klassendiskussionen. Sie bieten den Schülern konstruktives Feedback, das sie dazu ermutigt, ihr eigenes Lernen zu verstehen, ihre Schwächen zu identifizieren und sich entsprechend zu verbessern. Darüber hinaus leitet die formative Bewertung die Lehrer an und zeigt ihnen notwendige Anpassungen in ihrer Unterrichtsgestaltung auf, um den spezifischen Bedürfnissen der Schüler gerecht zu werden. Dies fördert effektiveres und

zielgerichtetes Lernen [18].

Angeichts dieser Unterschiede kann man je nach Art der Übung verschiedene Vorteile der Verwendung von E-Assessment-Systemen nutzen:

1. **Effizienz und Zeitersparnis:** Online-Bewertungssysteme ermöglichen eine automatische Bewertung von Prüfungen und Aufgaben, was den Zeitaufwand für Lehrer erheblich reduziert. Lehrer können so mehr Zeit für die Entwicklung von Lehrinhalten und die Unterstützung der Lernenden aufwenden [3].
2. **Skalierbarkeit:** Diese Systeme sind äußerst skalierbar und können gleichzeitig eine große Anzahl von Prüfungen und Aufgaben verwalten. Dies ist besonders nützlich für Bildungseinrichtungen, die viele Schüler oder Studenten betreuen [18] [3].
3. **Schnelles Feedback:** Online-Bewertungssysteme bieten den Lernenden sofortiges Feedback. Dies trägt zur Beschleunigung des Lernprozesses bei und ermöglicht es den Studenten, ihre Leistung sofort zu überprüfen und zu verbessern [18] [3].
4. **Individualisierung:** Einige E-Assessment-Systeme ermöglichen es, Prüfungen und Aufgaben an die individuellen Bedürfnisse und Ziele der Lernenden anzupassen. Dies fördert personalisiertes Lernen und ermöglicht es den Studenten, in ihrem eigenen Tempo zu arbeiten [3].
5. **Umfassende Datenanalyse:** Diese Systeme sammeln umfangreiche Daten zur Leistung der Lernenden. Durch die Analyse dieser Daten können Bildungseinrichtungen Trends identifizieren, Schwächen und Stärken erkennen und das Lehrplanangebot entsprechend anpassen [3].
6. **Reduzierung von Betrug und Plagiat:** Online-Bewertungssysteme verfügen über Sicherheitsmechanismen, die Betrug und Plagiat bei Prüfungen und Aufgaben minimieren. Dies trägt zur Integrität des Bewertungsprozesses bei [3].
7. **Flexibilität und Zugänglichkeit:** Online-Bewertungssysteme ermöglichen die Durchführung von Prüfungen und Aufgaben in verschiedenen Umgebungen, einschließlich Online- und Hybrid-Lernumgebungen. Dies bietet den Studenten Flexibilität und Zugänglichkeit zu Bildungsbewertungen [18] [3].

Zusammenfassend kann festgehalten werden, dass E-Assessment-Systeme als eine effiziente und vielseitige Methode zur Bewertung von Lernenden betrachtet werden können, die sowohl formative als auch summativ Bewertungen ermöglicht. Zahlreiche Vorteile werden durch sie ge-

boten. Mit diesem Verständnis werden im nächsten Kapitel die Mögliche Aufgabentypen in E-Assessment-Systemen nähere Einblicke gewonnen.

2.1.2 Mögliche Aufgabentypen

In der Erstellung von Prüfungen zur Bewertung von Schülern und Studierenden lassen sich im Allgemeinen zwei primäre Aufgabentypen identifizieren, die konsequent zu zwei verschiedenen Arten von Übungsszenarien führen: **Geschlossene Fragen** und **Offene Fragen** [37]. Diese Klassifikationen sind jeweils durch spezifische Charakteristika gekennzeichnet und bieten Vorzüge, die auf präzise Beurteilungsziele ausgerichtet sind [25].

Geschlossene Fragen (Closed questions)

Die Kategorie der Geschlossenen Fragen, auch als “closed questions” bezeichnet, repräsentiert eine geläufige Form von Aufgaben in E-Assessment-Systemen. Diese Fragen zeichnen sich durch ihre Fähigkeit aus, eine begrenzte Auswahl vorab definierter Antwortmöglichkeiten anzubieten, aus denen die Lernenden die adäquate Antwort auswählen müssen [18]. In der Regel umfasst diese Kategorie Multiple-Choice-Fragen, Wahr-Falsch-Fragen und ähnliche Formate. Die Vorzüge dieses Aufgabentyps sind vielschichtig. Primär sind sie als effektive Instrumente zur Beurteilung des Verständnisses spezifischer Konzepte und der Retention von Wissen zu betrachten. Darüber hinaus ermöglichen sie eine prompte und automatisierte Evaluierung, was unzweifelhaft im Interesse der Lehrenden und Prüfenden liegt. Allerdings sollte in Betracht gezogen werden, dass geschlossene Fragen in ihrer Fähigkeit, kritisches Denken, Kreativität und eigenständige Problemlösungsfähigkeiten zu bewerten, eingeschränkt sein können. Infolgedessen eignen sie sich vorwiegend für Beurteilungsziele, die auf die Prüfung grundlegender Kenntnisse und konzeptioneller Fertigkeiten abzielen [18]. Einige Beispiele für eine typische Übung mit geschlossener Frage:

1. Multiple-Choice-Fragen: Bei Multiple-Choice-Fragen wird den Lernenden eine Frage gestellt, und sie müssen aus einer Liste von vorgegebenen Antwortmöglichkeiten die richtige auswählen. Diese Art von Frage eignet sich gut, um das Verständnis von Fakten, Konzepten und Definitionen zu überprüfen [6].
2. Wahr/Falsch-Fragen: Wahr/Falsch-Fragen erfordern, dass die Lernenden entscheiden, ob eine gegebene Aussage wahr oder falsch ist. Diese Fragen sind besonders nützlich, um das Verständnis von Sachverhalten zu überprüfen [24].

3. Zuordnungsaufgaben (Matching Questions): Bei Zuordnungsaufgaben müssen die Lernenden Elemente aus zwei verschiedenen Listen miteinander in Beziehung setzen. Dies kann verwendet werden, um das Verständnis von Zusammenhängen und Beziehungen zwischen Konzepten zu prüfen [18].
4. Lückentexte (Fill-in-the-Blanks): Bei Lückentexten müssen die Lernenden fehlende Wörter oder Phrasen in einem Satz oder Text ergänzen. Diese Art von Aufgabe kann verwendet werden, um das Verständnis von Kontext und Details zu überprüfen [18].

Der Vorteil von geschlossenen Fragen in E-Assessment-Systemen liegt in ihrer klaren Struktur und ihrer objektiven Bewertbarkeit. Sie ermöglichen eine schnelle Auswertung und sind besonders geeignet, um das Wissen über Fakten und Grundlagen zu überprüfen. Darüber hinaus können sie automatisch bewertet werden, was die Effizienz bei der Beurteilung von Lernenden in großen Gruppen erhöht.

Offene Fragen (Open-Ended questions)

Dem gegenüber bieten Offene Fragen, auch als “open-ended questions” bezeichnet, einen anpassungsfähigeren und nuancierteren Ansatz zur Evaluation. Im Kontrast zu geschlossenen Fragen präsentieren sie keine vordefinierten Antworten und gestatten den Lernenden, ihre Gedanken selbstständig zu artikulieren [37]. Offene Fragen können in verschiedenen Formen gestellt werden, wie schriftliche Antworten, Lösungen von Problemen, argumentative Erläuterungen oder ähnliche Formate. Einer der Hauptvorteile dieser Fragestellungen liegt darin, dass sie die Beurteilung von kritischem Denken, Kreativität, Synthese- sowie schriftlichen oder mündlichen Ausdrucksfähigkeiten ermöglichen [37]. Sie bieten zudem detailliertere Einblicke in das Verständnis und die Fertigkeiten der Lernenden. Es ist jedoch bedeutend zu betonen, dass die Bewertung der Antworten auf diese Fragen häufig komplexer und subjektiver ist, eine menschliche Bewertung erfordert und länger dauern kann als die automatisierte Auswertung [18]. Des Weiteren könnten offene Fragen aufgrund ihres ressourcenintensiven Charakters unter Umständen weniger geeignet sein für umfangreiche Beurteilungen. Einige Beispiele für eine typische Übung mit offener Frage:

1. Essay-Fragen: Essay-Fragen sind offene Fragen, bei denen die Lernenden in ausführlichen schriftlichen Antworten ihr Wissen, ihre Analysefähigkeiten und ihre Argumentationsfähigkeiten darlegen müssen. Diese Art von Frage eignet sich gut, um komplexe Konzepte zu vertiefen und kritisches Denken zu fördern [18].

2. Fallstudien und Szenario-basierte Fragen: Bei diesen Fragen werden den Lernenden reale oder fiktive Szenarien oder Fallstudien vorgelegt, die sie analysieren und Lösungen oder Empfehlungen entwickeln müssen. Dies fördert die Anwendung von Wissen auf komplexe Probleme [18].
3. Reflexionsfragen: Reflexionsfragen ermutigen die Lernenden dazu, über ihr eigenes Lernen, ihre Erfahrungen und ihre Entwicklung nachzudenken. Diese Art von Frage ist besonders nützlich, um metakognitive Fähigkeiten zu fördern und das Bewusstsein für den Lernprozess zu schärfen [18].
4. Problemstellungen und Aufgaben mit freier Lösung: Bei dieser Art von Fragen werden den Lernenden komplexe Probleme oder Aufgaben gestellt, für die es keine festen Lösungen gibt. Die Lernenden müssen ihre eigenen Lösungen entwickeln und ihre Entscheidungen begründen [18].

Offene Fragen bieten den Lernenden die Möglichkeit, ihr Verständnis und ihre Fähigkeiten auf eine tiefere Weise zu zeigen, die über reine Faktenkenntnisse hinausgeht. Sie fördern kritisches Denken, Problemlösungsfähigkeiten und die Fähigkeit zur Kommunikation komplexer Ideen. Allerdings erfordert die Bewertung von offenen Fragen in der Regel mehr Zeit und Aufwand von Lehrkräften oder Experten, da die Antworten vielfältig und subjektiver Natur sein können.

Übungen, bei denen aus einem Text ein UML-Diagramm erstellt werden soll, gehören in der Regel zu den offenen Fragen in E-Assessment-Systemen. Dies liegt daran, dass sie von den Lernenden verlangen, nicht nur Faktenwissen anzuwenden, sondern auch kreativ denken und die Informationen aus dem Text analysieren müssen, um ein geeignetes UML-Diagramm zu erstellen. Diese Übungen erfordern von den Lernenden, dass sie ein tieferes Verständnis für das gegebene Thema entwickeln und die Informationen aus dem Text in einen visuellen Kontext übertragen können [42].

Zusammenfassend sind geschlossene Fragen und offene Fragen zwei unterschiedliche Ansätze zur Bewertung in E-Assessment-Systemen. Geschlossene Fragen eignen sich gut zur Bewertung von Grundkenntnissen und zur automatischen Bewertung, während offene Fragen eine erhöhte Flexibilität bieten, um komplexe Fähigkeiten zu bewerten, obwohl sie möglicherweise eine intensivere Bewertung erfordern. Die Auswahl zwischen diesen beiden Arten von Aufgaben hängt von den spezifischen Bewertungszielen, der Art der zu bewertenden Fähigkeiten sowie den verfügbaren Ressourcen für die Bewertung und Auswertung ab. In Kom-

bination tragen diese beiden Fragekategorien wesentlich zu einer umfassenden und ausgewogenen Bewertung der Lernenden im Kontext des E-Assessment bei.

2.2 Konzeptionelle Modelle

Ausgehend von der Definition ist ein konzeptionelles Modell eine abstrakte Darstellung oder eine konzeptionelle Struktur, die darauf abzielt, Ideen, Beziehungen, Konzepte oder Entitäten eines bestimmten Bereichs zu beschreiben, ohne in konkrete Details oder Implementierungsdetails einzugehen. Solche Modelle werden häufig in verschiedenen Bereichen wie Informatik, Wissenschaft, Ingenieurwissenschaften, Management, Philosophie usw. verwendet, um das Verständnis eines komplexen Themas zu klären, die Kommunikation und Diskussion zu erleichtern und als Grundlage für die Gestaltung oder Analyse konkreter Systeme zu dienen [2].

Konzeptionelle Modelle können verschiedene Formen annehmen, darunter Diagramme, Schemata, grafische Darstellungen, textuelle Beschreibungen sowie mathematische Repräsentationen. Diese dienen häufig als initialer Schritt innerhalb des Modellierungs- oder Problemlösungsprozesses, um die grundlegenden Konzepte und Zusammenhänge zu erfassen, bevor man sich in die detaillierten Facetten vertieft [2]. In der Fachdisziplin der Informatik, zum Beispiel, wird ein konzeptionelles Modell genutzt, um Schlüsselparameter und Verknüpfungen innerhalb einer Datenbank zu definieren, ohne Einzelheiten zur Datenspeicherung oder -abfrage preiszugeben [2]. Es kann auch verwendet werden, um die logische Architektur eines Systems zu beschreiben, wobei die Beziehung zwischen Objekten, Klassen und verschiedenen Entitäten hervorgehoben wird.

Im Allgemeinen nutzen Modellierungssprachen in den betreffenden Fachgebieten Notationen, die graphische Symbole einschließen und in zweidimensionalen visuellen Darstellungen resultieren [28]. Solche Darstellungen sind gebräuchlicherweise als Diagramme bekannt, und dies findet oft seinen Ausdruck in den Namen spezifischer Modelltypen wie "Entity-Relationship-Diagramm" oder "UML-Klassendiagramm". Falls grafische Symbole innerhalb eines Modells eingesetzt werden, erfolgt ihre Annotation üblicherweise durch textliche Kennzeichnungen, um ihre Relevanz im Kontext des modellierten Objekts zu präzisieren. Die Modellierungssprachen, die im Rahmen der initialen Forschung herausragen (wie ERD, UML, EPC, BPMN und Petri-Netze), repräsentieren beispielhafte Instanzen solcher graphischen (oder diagrammatischen) Modellierungssprachen [42].

Im Fachgebiet der Informatik gibt es verschiedene Arten von konzep-

tionellen Modellen, die je nach ihrem Anwendungsbereich und Ziel unterschiedliche Formen und Eigenschaften aufweisen. Hier sind einige häufig vorkommende Arten von konzeptionellen Modellen in der Informatik:

1. **Entity-Relationship-Modelle (ER-Modelle):** Diese Modelle werden verwendet, um die Struktur von Datenbanken zu beschreiben, indem sie Entitäten (Objekte oder Konzepte) und deren Beziehungen zueinander darstellen. ER-Modelle verwenden typischerweise Diagramme, um Entitäten, Attribute und Beziehungen grafisch darzustellen [16].
2. **UML-Diagramme (Unified Modeling Language):** UML ist eine weit verbreitete Modellierungssprache in der Softwareentwicklung. Sie umfasst verschiedene Diagrammtypen, die zur Modellierung von Softwarearchitekturen, Prozessen und Verhaltensweisen verwendet werden [33].
3. **Data Flow Diagrams (DFD):** DFDs werden verwendet, um den Datenfluss und die Datenverarbeitung in Informationssystemen darzustellen. Sie zeigen, wie Daten zwischen Prozessen, Datenlagern und externen Entitäten fließen [26].
4. **BPMN-Diagramme (Business Process Model and Notation):** BPMN ist eine Modellierungssprache, die sich auf die Darstellung von Geschäftsprozessen konzentriert. Mit BPMN-Diagrammen können Abläufe, Aktivitäten und Entscheidungen innerhalb eines Unternehmensmodells visualisiert werden [45].
5. **Petri-Netze:** Petri-Netze sind mathematische Modelle, die zur Modellierung und Analyse von parallelen und verteilten Systemen verwendet werden. Sie sind besonders nützlich bei der Modellierung von Prozessen in der Softwareentwicklung und der Kommunikation zwischen Komponenten [32].
6. **Systemarchitekturmodelle:** Diese Modelle bieten eine Übersicht über die Architektur eines Software- oder Informationssystems und zeigen die Hauptkomponenten und ihre Interaktionen.

Diese Arten von konzeptionellen Modellen dienen dazu, komplexe Systeme, Prozesse und Datenstrukturen in der Informatik zu erfassen, zu analysieren und zu kommunizieren. Die Wahl des geeigneten Modells hängt von den spezifischen Anforderungen und Zielen eines Projekts ab.

Das Thema dieser Masterarbeit hat als Anwendungsfall Aufgabe mit offenen Fragen, die sich mit der Modellierung mit UML-Diagrammen befassen und bewertet werden sollen. Die Unified Modeling Language (UML) ist eine standardisierte und visuelle Modellierungssprache, die in der Softwareentwicklung und Systemmodellierung weit verbreitet ist. UML dient dazu,

komplexe Systeme, insbesondere Softwareanwendungen, zu beschreiben, zu analysieren, zu entwerfen und zu dokumentieren. Sie wurde erstmals in den 1990er Jahren von Grady Booch, James Rumbaugh und Ivar Jacobson entwickelt und hat sich seitdem zu einem Industriestandard für die Modellierung von Software- und Systemarchitekturen entwickelt [9].

UML bietet eine breite Palette von Diagrammtypen, darunter Klassendiagramme, Aktivitätsdiagramme, Sequenzdiagramme, Zustandsdiagramme und viele mehr. Jeder Diagrammtyp konzentriert sich auf bestimmte Aspekte eines Systems und ermöglicht es den Entwicklern und Ingenieuren, die verschiedenen Elemente und deren Beziehungen in einer klaren und leicht verständlichen visuellen Darstellung festzuhalten [9]. Dies fördert eine bessere Kommunikation und Zusammenarbeit zwischen den Mitgliedern eines Entwicklungsteams sowie zwischen den verschiedenen Interessengruppen eines Projekts.

Die Verwendung von UML in der Softwareentwicklung bietet eine Reihe von Vorteilen, darunter die Möglichkeit, Systeme zu abstrahieren, zu modularisieren und zu dokumentieren, was die Entwicklung, Wartung und Erweiterung von Software erleichtert [9]. Darüber hinaus unterstützt UML die frühzeitige Fehlererkennung und das systematische Design von Softwarelösungen, was zu einer höheren Qualität und Zuverlässigkeit von Anwendungen führt. Aufgrund seiner weitverbreiteten Akzeptanz und seiner Fähigkeit, komplexe Ideen in leicht verständlichen Diagrammen darzustellen, spielt UML eine entscheidende Rolle in der modernen Softwareentwicklung und Systemmodellierung und bildet die Grundlage für die Erstellung und den Austausch von Modellen und Entwurfsmustern in der Industrie [9].

2.3 Automatisierte Bewertung

2.3.1 Definition und Abgrenzung

Die automatisierte, computergestützte Bewertung von akademischen Einreichungen, wird in der Hochschulbildung häufig verwendet. Informatikbasierte automatisierte Bewertungssysteme existieren bereits seit den 1960er Jahren [42], während automatisierte Bewertung von Diagrammen seit etwa zwei Jahrzehnten bekannt ist. Diese Systeme werden eingesetzt, um sowohl offene Aufgaben zu bewältigen, als auch geschlossene Fragen. Bei einer Bewertung, in der ein Student ein Modell aus einem gegebenen Text erstellen soll, ist der Korrektor oft dazu verpflichtet, einen Vergleich zwischen dem vom Studenten vorgelegten Ergebnis und der erwarteten Lösung durchzuführen. Allerdings kann dieses Korrekturmodell mehrere

Probleme aufwerfen:

1. Zunächst einmal, da es im Bereich der Modellierung keine eindeutige Lösung gibt, wird diese Aufgabe weniger offensichtlich, da die Ergebnisse jedes Studenten mehrmals mit der vorgeschlagenen Lösung verglichen werden müssen, um sicherzustellen, dass die Arbeit des Studenten alle wesentlichen Komponenten der Lösung berücksichtigt.
2. Das zweite Hindernis, das aus diesem Korrekturmodell resultiert, liegt in der inhärenten Subjektivität dieser Methode. Einige Studenten könnten Modelle entwickeln, die vollkommen funktionsfähig sind, aber von der erwarteten Lösung abweichen, und für diese Abweichung benachteiligt werden. Darüber hinaus kann dies zu einer gewissen Starrheit in der Lehre führen und die Studenten daran hindern, originale Ansätze zu erkunden.
3. Eine weitere wichtige Herausforderung dieses Korrekturmodells besteht in der potenziellen Subjektivität der Korrektoren. Jeder Korrektor kann seine eigene Interpretation dessen haben, was eine richtige Antwort ausmacht, was zu Inkonsistenzen bei den Bewertungen und einer ungleichmäßigen Verteilung der Noten zwischen den Studentearbeiten führen kann. Dies kann besonders problematisch werden, wenn mehrere Korrektoren dieselben Arbeiten bewerten, was zu erheblichen Unterschieden in den vergebenen Noten führen kann.
4. Schließlich kann dieses Korrekturmodell auch zeitaufwändig sein, insbesondere in Kursen mit einer großen Anzahl von Studenten. Der detaillierte Vergleich zwischen den Studentearbeiten und der Referenzlösung erfordert Zeit und kann zu Verzögerungen bei der Mitteilung der Ergebnisse an die Studenten führen und die Arbeitsbelastung der Korrektoren erhöhen.

In diesem Zusammenhang könnte die Einführung einer automatisierten Korrektur auf der Grundlage klar definierter spezifischer Kriterien, die in jeder Lösung zwingend enthalten sein müssen, eine optimale Lösung darstellen. Eine automatisierte Korrektur durch ein Computerprogramm könnte jedes dieser Probleme lösen, indem sie eine schnelle, faire und vorurteilsfreie Bewertung ermöglicht.

2.3.2 Automatisierte Bewertungsmethoden für UML

Die Bewertung von UML-Diagrammen ist in verschiedenen Bildungskontexten von entscheidender Bedeutung, sei es in Informatiklehrgängen an Universitäten oder in der beruflichen Weiterbildung. Dabei kann die manuelle Bewertung von UML-Diagrammen zeitaufwändig und subjektiv sein,

insbesondere wenn es sich um eine große Anzahl von Diagrammen handelt. Um diese Herausforderungen zu bewältigen und eine effiziente und objektive Bewertung zu gewährleisten, werden verschiedene Methoden und Ansätze zur automatisierten Bewertung entwickelt und angewendet:

1. **Methoden zur Modellvergleich:** Diese Methoden beinhalten den Vergleich des bewerteten Modells mit einem oder mehreren Lösungsmodellen. Dieser Vergleich kann mithilfe von Ähnlichkeitsmaßen, wie Ähnlichkeitsmaßen oder Graphenabgleich, durchgeführt werden. Dabei werden Ähnlichkeiten und Unterschiede zwischen dem bewerteten Modell und den Lösungsmodellen ermittelt [42] [13].
2. **Regelbasierte Ansätze:** Regelbasierte Ansätze verwenden vordefinierte Regeln oder Kriterien zur Bewertung des bewerteten Modells. Diese Regeln können verschiedene Formen annehmen, darunter Graphabfragen, Eigenschaften, Metriken, Mängel oder Suchmuster innerhalb des bewerteten Modells. Die Bewertung erfolgt anhand der Einhaltung oder Nichteinhaltung dieser Regeln [42] [39].
3. **Constraints-basierte Ansätze:** Diese Ansätze sind eng mit regelbasierten Ansätzen verwandt und werden häufig in intelligenten Tutoresystemen eingesetzt. Sie beinhalten die Anwendung von Einschränkungen oder Regeln zur Bewertung des bewerteten Modells. Diese Einschränkungen können sich auf die Struktur, die Semantik oder andere Aspekte des Modells beziehen [42] [19].
4. **Methoden des maschinellen Lernens:** Einige aktuelle Forschungsartikel präsentieren Ansätze, die auf Methoden des maschinellen Lernens setzen, um die automatisierte Bewertung durchzuführen. Hierbei werden maschinelle Lernmodelle trainiert, um Modelle auf Grundlage vorheriger manueller Bewertungen zu bewerten. Dies beinhaltet das Training von Modellen zur Identifizierung von Mustern, Anomalien und potenziellen Problemen in UML-Diagrammen [42] [10] [23].
5. **Andere Techniken:** Neben den genannten Ansätzen gibt es verschiedene andere Techniken, wie die Simulation des bewerteten Modells, Teststrategien, die Gruppierung von Modellen und die Ausrichtung des bewerteten Modells mit einer annotierten textuellen Beschreibung [42].
 - (a) **Werkzeuge zur Modelltransformation und Codegenerierung:** Diese Werkzeuge ermöglichen die automatische Generierung von Code aus UML-Diagrammen. Sie helfen dabei, den manuellen Aufwand zur Übersetzung eines UML-Designs in ausführbaren Code zu reduzieren und die Korrektheit des Diagramms zu

überprüfen. Dies geschieht durch die automatisierte Erzeugung von ausführbarem Code aus den Elementen und Strukturen des UML-Diagramms und ermöglicht die anschließende Überprüfung der Korrektheit des generierten Codes [40].

Diese verschiedenen Ansätze und Techniken tragen dazu bei, die automatisierte Bewertung von UML-Diagrammen in der Bildung und anderen Anwendungsbereichen effektiver und vielfältiger zu gestalten. Sie ermöglichen eine präzise und umfassende Bewertung von Modellen und tragen zur Verbesserung der Qualität von UML-Diagrammen bei.

2.3.3 Herausforderung der automatisierten Bewertung

Das vorrangige Ziel der automatisierten Bewertung besteht in der Verleihung von Bewertungen, was als eine Klassifizierungsaufgabe innerhalb des Bewertungskontexts betrachtet werden kann. In Fällen von offenen Aufgaben erfolgt die Bewertung anhand der relativen Position der eingereichten Lösung innerhalb des umfassenden Lösungsraums. Dieser Lösungsraum beinhaltet eine Bandbreite von Antwortmöglichkeiten, die von vollständigen Lösungen über partielle Lösungen bis hin zu ungültigen oder nicht akzeptablen Lösungen reicht. Dieser methodische Ansatz zur automatisierten Bewertung, der auf der Unterscheidung und Beurteilung der Qualität von Antworten basiert, findet in verschiedenen Bereichen breite Anwendung. Beispiele hierfür sind die automatisierte Bewertung von Essays [43], die Beurteilung von Programmieraufgaben [17] sowie die Modellierung und Bewertung von Modellen in diversen Fachdisziplinen [36]. Dieser Ansatz ermöglicht eine effiziente, objektive und skalierbare Bewertung von eingereichten Arbeiten, und er trägt wesentlich dazu bei, den Bewertungsprozess zu rationalisieren und zu standardisieren.

Im Kontext der Erstellung eines Diagramms (UML, Sequenzdiagramm, Entity-Relationship-Diagramm oder andere) aus einem Text, der ein System beschreibt, wird diese Aufgabe in der Regel als offen angesehen. Die Studierenden müssen die im Text beschriebenen Konzepte, Entitäten und Beziehungen abstrakt interpretieren und darstellen, was bedeutet, dass es normalerweise keine eindeutige Antwort oder vordefinierte Lösung gibt. Verschiedene Studierende können leicht unterschiedliche Diagramme erstellen, um die gleiche textuelle Beschreibung darzustellen.

Da diese Aufgabe offen und subjektiv ist, müssen die Prüfer die Kreativität und das Verständnis jedes Studierenden bewerten. Ihr Ziel ist es zu bestimmen, ob das erstellte Diagramm die im Ausgangstext beschriebenen Konzepte und Beziehungen angemessen erfasst. Diese Komplexität macht

den Ansatz auf Grundlage von Regeln für die Bewertung dieser Aufgabe geeignet, da er es ermöglicht, vordefinierte Kriterien anzuwenden und spezifisches Feedback entsprechend dieser Kriterien bereitzustellen.

2.4 State of the Art

In diesem Kapitel wird die Diskussion auf bestimmte, mehr oder weniger aktuelle Technologien ausgedehnt, die im Kontext des E-Assessments verschiedene Methoden aus dem vorherigen Kapitel verwenden.

2.4.1 Machine learning zur Bewertung von UML

Der Artikel mit dem Titel “Automatic Assessment of Students’ Software Models Using a Simple Heuristic and Machine Learning” [10] stammt von den Autoren Younes Boubekur, Gunter Mussbacher und Shane McIntosh. In ihrem Artikel stellen sie einen innovativen Ansatz zur Bewertung von Studienarbeiten in Modellierungskursen vor. Ihr Hauptziel besteht darin, den zeit- und aufwändigen und subjektiven Bewertungsprozess von UML-Diagrammen zu bewältigen, der in der Softwaretechnikausbildung weit verbreitet ist. Die vorgeschlagene Methodik kombiniert einen einfachen heuristischen Algorithmus mit fortgeschrittenen maschinellen Lernverfahren, um nicht nur hochwertige Studienarbeiten zu identifizieren, sondern auch ungefähre Noten vorherzusagen.

Die Autoren beginnen ihre Arbeit damit, einen Überblick über bestehende Ansätze zur Bewertung von UML-Diagrammen zu geben. Diese Ansätze umfassen manuelle Bewertung, automatische Bewertung mithilfe regelbasierter Systeme und automatische Bewertung unter Verwendung von maschinellem Lernen. Anschließend stellen die Autoren ihren innovativen Ansatz vor, der einen einfachen heuristischen Algorithmus verwendet, um die Einreichungen der Studierenden mit einer Idealvorlage zu vergleichen [10]. Dieser heuristische Algorithmus basiert auf dem Prinzip, die Unterschiede zwischen der Einreichung des Studierenden und der Idealvorlage zu quantifizieren und anschließend eine Punktzahl aufgrund dieser Unterschiede zuzuweisen [21]. Darüber hinaus setzen die Autoren maschinelle Lernverfahren ein, um ungefähre Noten vorherzusagen, basierend auf den von dem heuristischen Algorithmus generierten Punktzahlen (siehe 2.1).

Um die Effektivität ihres vorgeschlagenen Ansatzes zu validieren, führten die Autoren eine empirische Studie mit 50 Studierenden durch, die an einem Softwaretechnikkurs teilnahmen. Die Studierenden wurden gebeten, UML-Diagramme für ein gegebenes Problem einzureichen, und die Autoren

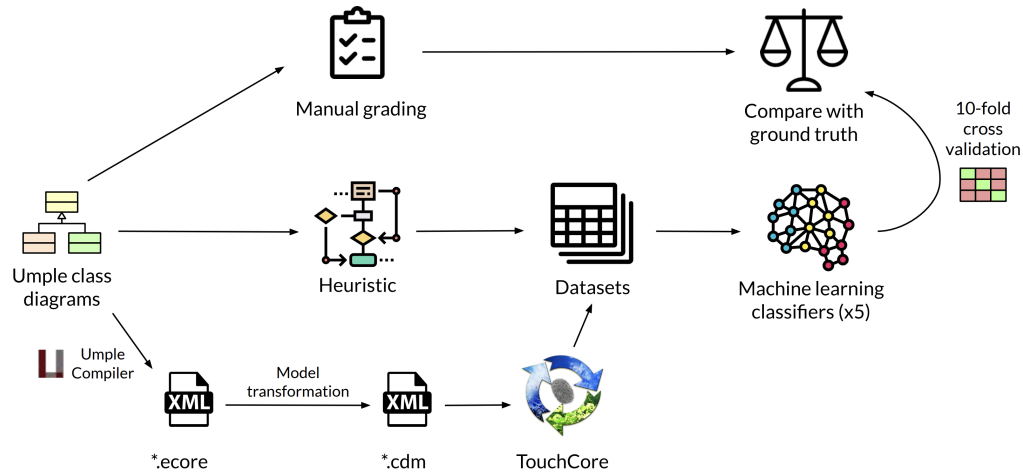


Abbildung 2.1: Ansatz der Studie [10].

wandten ihren innovativen Ansatz zur Bewertung dieser Einreichungen an. Um die Zuverlässigkeit der Bewertung sicherzustellen, wurden auch zwei menschliche Prüfer beauftragt, dieselben Einreichungen unabhängig voneinander zu bewerten.

Die Ergebnisse der Studie zeigen, dass der vorgeschlagene Ansatz mit der menschlichen Bewertung hinsichtlich der Identifizierung hochwertiger Einreichungen vergleichbar ist und erstaunlich präzise ungefähre Noten vorhersagen kann. Darüber hinaus verglichen die Autoren ihren Ansatz mit einer komplexen regelbasierten Technik und stellten fest, dass ihre Methode in Bezug auf Effizienz und Wartbarkeit überlegen ist [10].

Zusammenfassend bietet der innovative Ansatz eine vielversprechende Lösung zur Bewertung von UML-Diagrammen im Bereich der Software-technikausbildung. Durch die Kombination eines einfachen heuristischen Algorithmus mit fortgeschrittenen maschinellen Lernverfahren identifiziert ihr Ansatz nicht nur hochwertige Einreichungen, sondern bietet auch eine effiziente Möglichkeit zur Vorhersage ungefährer Noten. Die Autoren betonen das potenzielle Zeitersparnis und die Reduzierung der Subjektivität im Bewertungsprozess und schlagen vor, dass ihr Ansatz auf andere Bereiche erweitert werden kann, sofern es eine klare Definition einer Idealvorlage gibt und eine Methode existiert, um die Einreichungen der Studierenden mit dieser Idealvorlage zu vergleichen.

2.4.2 Bewertung auf Basis von Ähnlichkeitsmaßen

Die Autoren des Artikels “A Different Approach on Automated Use Case Diagram Semantic Assessment” [13] sind Daniel Siahaan, Siti Rochimah, Reza Fauzan und Evi Triandini. In diesem Artikel stellen die Autoren eine innovative Methode zur automatischen semantischen Bewertung von Anwendungsfall-Diagrammen im Rahmen von UML-Diagrammen vor. Das Hauptziel dieser Methode besteht darin, die Herausforderungen anzugehen, vor denen Pädagogen bei der Bewertung von Anwendungsfall-Diagrammen stehen, sowohl in zwischenmenschlicher als auch intrapersonaler Hinsicht.

Die Autoren verwenden einen semantischen Bewertungsansatz, der in zwei Hauptdimensionen unterteilt ist: Eigenschaft (property) und Beziehung (relationship). Alle in dieser Bewertung verwendeten Informationen stammen aus Beschriftungen, die aus einem XML Metadata Interchange (XMI)-Dokument [1] übersetzt wurden. Zur Messung der Ähnlichkeit zwischen den Beschriftungen verwenden die Autoren die Kosinus-Ähnlichkeit und nutzen WuPalmer zur Unterstützung von WordNet-Suchen [13].

Für ihre empirische Studie sammelten die Autoren einen Datensatz aus drei unterschiedlichen Projekten: Outlay, QuickBill und dem Restaurant Management System (RMS). Diese Projekte umfassen die Finanzaufzeichnung, den Point of Sale und das Restaurant-Bestellsystem. Der Quellcode für diese Projekte wurde von GitHub bezogen und folgt einem objektorientierten Ansatz. Den Studierenden wurden Demonstrationen zur Funktionsweise jedes Projekts gezeigt, und anschließend wurden sie aufgefordert, Anwendungsfall-Diagramme auf der Grundlage ihres Verständnisses der Projekte und der zugehörigen Terminologie zu erstellen. Insgesamt sammelten die Autoren 36 Anwendungsfall-Diagramme von den Studierenden, wobei jedes Projekt ein Antwortschlüssel-Diagramm hatte. Somit umfasst der Datensatz insgesamt 39 Anwendungsfall-Diagramme [13].

Bevor sie ihre vorgeschlagene Methode bewerteten, etablierten die Autoren einen Goldstandard als Referenz für die Bewertung. Dieser Goldstandard repräsentiert das durchschnittliche Ergebnis der Expertenbewertungen von Studentenantworten auf der Grundlage des Antwortschlüssels. Insgesamt beteiligten sich 21 Experten an dieser Bewertung, die von zwölf verschiedenen Universitäten stammten. Die Mindestanforderung für eine Expertin oder einen Experten war ein Master-Abschluss in Informatik.

Zur Bewertung ihrer vorgeschlagenen Methode nutzten die Autoren Gwets AC1, ein Maß, das den Grad der Übereinstimmung zwischen zwei Experten anzeigt. In diesem Fall repräsentiert der erste Experte den Durchschnitt der Bewertungen der Experten, während der zweite Experte die Bewertung durch die vorgeschlagene Methode darstellt. Die Autoren

übertrugen die Ähnlichkeitswerte der beiden Experten auf eine Skala von 1 bis 5, wobei eine Wertung von eins ein Bewertungsergebnis von weniger als zwanzig, zwei ein Bewertungsergebnis von zwanzig bis vierzig, drei ein Ergebnis von vierzig bis sechzig, vier ein Ergebnis von sechzig bis achtzig und fünf ein Ergebnis von mehr als achtzig darstellt. Anschließend berechneten die Autoren die Übereinstimmung zwischen den beiden Experten unter Verwendung von Gwets AC1 [13].

Die Ergebnisse ihrer Bewertung zeigten, dass die vorgeschlagene Methode eine erhebliche Übereinstimmung mit der Bewertung durch den Experten aufwies. Interessanterweise stellten die Autoren fest, dass ihre vorgeschlagene Methode eine höhere Übereinstimmung mit dem Experten aufwies als die durchschnittliche Übereinstimmung zwischen den Experten. Darüber hinaus beobachteten die Autoren, dass Pädagogen tendenziell mehr Wert auf Informationen über Eigenschaften legen als auf Informationen über Beziehungen bei der Bewertung von Anwendungsfall-Diagrammen. Diese Erkenntnis könnte dazu beitragen, den Bewertungsprozess zu verbessern und eine konsistentere und objektivere Bewertung von Studentenantworten zu gewährleisten [13].

Zusammenfassend stellen die Autoren eine innovative Methode zur automatischen semantischen Bewertung von Anwendungsfall-Diagrammen vor. Ihr Ansatz umfasst die semantische Bewertung in den Dimensionen (property) und Beziehung (relationship), wobei ein Datensatz aus realen Projekten gesammelt und ein Goldstandard als Referenz etabliert wurde. Die Evaluation ihrer Methode unter Verwendung von Gwets AC1 zeigt eine erhebliche Übereinstimmung mit Expertenbewertungen. Darüber hinaus liefert ihre Erkenntnis darüber, dass Pädagogen bei der Bewertung von Anwendungsfall-Diagrammen tendenziell Eigenschaftsinformationen stärker gewichten als Beziehungsinformationen, wertvolle Einblicke zur Verbesserung des Bewertungsprozesses und zur Förderung einer konsistenten und objektiveren Bewertung von Studentenantworten.

2.4.3 Graph matching zur Bewertung von UML

Der Artikel mit dem Titel "New method for summative evaluation of UML class diagrams based on graph similarities" wurde von Outair Anas, Mohammed Ouadou und Abdelhadi Lotfi verfasst [4]. In ihrem Artikel widmen sich die Autoren der anspruchsvollen Aufgabe der Bewertung von UML-Klassendiagrammen, die von Studierenden erstellt wurden. Diese Aufgabe ist oft komplex für Lehrende, da UML-Klassendiagramme mehrere gültige Darstellungen haben können. Die Autoren schlagen ein halbautomatisches System vor, das dieses Problem durch den Einsatz eines Vergleichs von



Abbildung 2.2: Beispiel für einen Graphenabgleich [4].

syntaktischen, strukturellen und semantischen Ähnlichkeiten angeht, um Fehler von Studierenden zu identifizieren und wertvolles Feedback zu ihrem Lernprozess zu geben. Ihre Methode besteht aus drei Schritten:

1. **UML-Diagramme in Graphen umwandeln:** Die Autoren beginnen ihren Ansatz, indem sie UML-Klassendiagramme in Graphdarstellungen umwandeln und das Metamodell durch die Einführung neuer Elemente zur Erleichterung der Bewertung verbessern. Sie stellen Klassen als Knoten dar, Attribute als mit Klassen verbundene Knoten und Assoziationen als beschriftete Kanten [5].
2. **Definition von Ähnlichkeitsmaßen:** Die Autoren definieren eine Reihe von Ähnlichkeitsmaßen, die auf die transformierten UML-Graphen anwendbar sind. Sie untersuchen verschiedene Techniken zur Graphabstimmung und Metriken zur Knotenähnlichkeit, um den Vergleich von Graphen und die Fehlererkennung zu erleichtern [14].
3. **Abgleich und Vergleich von Graphen:** Unter Verwendung der zuvor definierten Ähnlichkeitsmaße führen die Autoren einen Abgleich und Vergleich der von Studierenden erstellten UML-Diagramme und der vom Lehrer bereitgestellten Referenzdiagramme durch [30] (Siehe 2.2).

Die Leistung des Systems wird anhand eines Datensatzes bewertet, der aus 100 von Studierenden erstellten Klassendiagrammen besteht, wobei

für jedes Diagramm ein entsprechendes Referenzdiagramm zur Bewertung bereitgestellt wird. Drei Übungen werden ausgewählt, um das System offline zu konfigurieren und zu bewerten, wobei iterative Verbesserungen an den Ähnlichkeitskriterien und der Systemfunktionalität auf Grundlage der Übungsergebnisse vorgenommen werden. Die Ergebnisse zeigen, dass das System eine Genauigkeitsrate von 70 % bei der Erkennung von Fehlern von Studierenden erreicht und minimalen Eingriff erfordert, um Übereinstimmungen für 80 % der verarbeiteten Diagramme zu korrigieren [4]. Die Autoren betonen, dass der in das System integrierte formative Bewertungsansatz gut geeignet ist, um UML-Klassendiagramme zu bewerten, da er den Studierenden Feedback zur Verbesserung ihres Verständnisses des Lehrstoffes liefert. Darüber hinaus hat das System das Potenzial, die Belastung der Lehrenden zu verringern, indem es den Bewertungsprozess automatisiert und ihnen ermöglicht, individuelleres Feedback an einzelne Studierende zu geben.

Zusammenfassend bietet die in diesem Artikel vorgestellte Methode eine vielversprechende Möglichkeit zur summativen Bewertung von UML-Klassendiagrammen unter Berücksichtigung der Möglichkeit mehrerer gültiger Darstellungen. Durch den umfassenden Vergleich von syntaktischen, strukturellen und semantischen Ähnlichkeiten kann das System Fehler in von Studierenden erstellten Diagrammen erkennen und zur Verbesserung ihres Lernfortschritts beitragen. Die Autoren erkennen an, dass weitere Forschung erforderlich ist, um das System zu verfeinern und seine Genauigkeit zu verbessern, aber die Ergebnisse dieser Studie legen nahe, dass es das Potenzial hat, eine wertvolle Unterstützung sowohl für Lehrende als auch für Studierende bei der Bewertung von UML-Klassendiagrammen zu sein.

2.4.4 Automatisierte Bewertung mit GReQL

In ihrem Artikel “Automated checks on UML diagrams” [39] stellen Michael Striwe und Michael Goedicke eine innovative Technik zur Evaluierung von UML-Klassendiagrammen auf der Grundlage von Graphabfragen vor. Sie betrachten UML-Klassendiagramme als Graphen und formalisieren ihren Ansatz mithilfe einer Graphabfragesprache namens GReQL, die von der JGraLab-Bibliothek bereitgestellt wird.

Graph Repository Query Language (GReQL), ähnlich wie SQL, eignet sich gut für die Implementierung regelbasierter Prüfungen von graphenbasierten Daten [38]. Sie ermöglicht die Abfrage von Elementen bestimmter Typen, die Untersuchung ihrer Verbindungen und die Überprüfung ihrer Attribute. Im Kontext von UML-Klassendiagrammen können diese Abfragen

das Vorhandensein von Diagrammelementen wie Klassen, Schnittstellen, Eigenschaften, Operationen, Parametern, Assoziationen und Generalisierungen ermitteln und ihre Beziehungen bewerten [39].

```

1  <rule type="presence" points="5">
2    <query>
3      from x : V{Class},
4      y : V{Property}, z : V{PrimitiveType}
5        with
6          isDefined(x.name) and x.name="A" and
7          x --> y and
8          isDefined(y.name) and y.name="b" and
9          y --> z and
10         isDefined(z.name) and z.name="String"
11      report 1 end
12    </query>
13    <feedback>Ein "A" soll ein Attribut für die
14      Eigenschaft "b" bereitstellen.</feedback>
15  </rule>

```

Quelltext 2.1: Codebeispiel in GReQL

Um ein UML-Diagramm mithilfe dieses Ansatzes zu evaluieren, parsen die Autoren das Diagramm zunächst in eine graphenbasierte Darstellung seiner abstrakten Syntax, in der Regel durch einen XML-Parser. Anschließend verwenden sie GReQL-Abfragen, um die Anwesenheit bestimmter Diagrammelemente und ihrer Verbindungen zu überprüfen. Die Autoren stellen eine Reihe von Regeln vor, die als GReQL-Abfragen implementiert sind und an die Anforderungen bestimmter Kurse oder Aufgaben angepasst werden können.

In ihrem Bewertungssystem werden einzelnen Regeln individuelle Punktzahlen zugewiesen, wobei unterschiedliche Gewichtungen zur Unterscheidung zwischen Korrektheits- und Qualitätsaspekten verwendet werden. Zum Beispiel könnte eine Regel, die das Vorhandensein einer Klasse mit einem bestimmten Namen sicherstellt, als wichtiger angesehen werden als eine, die das Vorhandensein eines Kommentars zu einer Klasse überprüft. Die Autoren schlagen auch eine Methode zur Berechnung von Gesamtpunkten vor, indem die Punktzahlen jeder Regel aggregiert werden [39].

Um die praktische Anwendung ihres Ansatzes zu zeigen, bieten die Autoren ein Beispiel aus einem UML-Modellierungskurs an, in dem Lösungen von Studierenden automatisch mit ihrer Methodik bewertet wurden. Das Ergebnis dieser Beispielbewertung zeigt die Effektivität ihres Ansatzes bei

der Erkennung von Fehlern und der Bereitstellung rechtzeitigen Feedbacks an die Studierenden, wodurch die Arbeitsbelastung der Dozenten reduziert wird [39].

Über UML hinaus sehen die Autoren die Vielseitigkeit ihres Ansatzes bei der Bewertung anderer Sprachen, die in eine graphenbasierte Darstellung ihrer abstrakten Syntax transformiert werden können, was praktisch alle Programmiersprachen einschließt. Sie sehen seine Anwendung in intelligenten Tutoringsystemen und automatisiertem Tutoring, da es eine sofortige Rückmeldung an Studierende ermöglicht und den Bewertungsprozess für Lehrkräfte vereinfacht.

Mögliche Einschränkungen dieses Ansatzes sind jedoch die Abhängigkeit von der Annahme einer korrekten Umwandlung von UML-Diagrammen in graphenbasierte Darstellungen, die zu Ungenauigkeiten führen kann, wenn der Umwandlungsprozess fehlerhaft ist. Die Autoren schlagen jedoch vor, den Umwandlungsprozess durch eine Reihe von Testfällen zu validieren, um die Zuverlässigkeit sicherzustellen [39].

Zusätzlich konzentriert sich der Ansatz hauptsächlich darauf, das Vorhandensein von Diagrammelementen und ihren Verbindungen zu überprüfen, bewertet jedoch nicht intrinsisch die Korrektheit des Inhalts innerhalb dieser Elemente. Zum Beispiel kann er überprüfen, ob eine Klasse einen bestimmten Namen hat, aber er überprüft nicht die Richtigkeit der zugehörigen Attribute und Methoden. Die Autoren weisen jedoch darauf hin, dass ihr Ansatz erweitert werden kann, um komplexere Regeln für die Bewertung des Inhalts von Diagrammelementen zu umfassen [38].

Zusammenfassend bietet der Ansatz von Striewe und Goedicke eine vielversprechende Alternative zur Bewertung von UML-Klassendiagrammen. Indem sie diese Diagramme als Graphen behandeln und GReQL für Abfragen nutzen, bieten sie eine flexible und anpassbare Möglichkeit zur Bewertung. Ihre Methodik hat das Potenzial für breitere Anwendungen in verschiedenen Sprachen und Bildungsbereichen und verspricht, das Feedback an Studierende zu verbessern, während sie die Arbeit für Lehrende vereinfacht.

2.5 JACK

Das E-Assessment-System namens "JACK3" (dritte Version von JACK) [29], entwickelt von der Universität Duisburg-Essen, stellt eine webbasierte Plattform dar, die Pädagogen eine effektive Lösung für die Erstellung und Durchführung von Online-Übungen bietet. Diese vielseitige Plattform ermöglicht Lehrkräften die Erstellung einer breiten Palette von Übungen,

die nahtlos über eine benutzerfreundliche Oberfläche bereitgestellt werden können. Neben der Erstellung von Bewertungen verfügt JACK3 über eine Auswahl an Funktionen, die den Bewertungsprozess optimieren sollen. Diese beinhalten die automatisierte Bewertung, umfassende Berichterstellungswerkzeuge und die Echtzeitverfolgung des Lernfortschritts der Schüler.

Die Bedeutung von JACK3 reicht über die reine Bequemlichkeit hinaus, da es Lehrern ermöglicht, wertvolle Zeit und Ressourcen zu sparen, während gleichzeitig die Effizienz und Effektivität der Bewertungen gesteigert werden. Des Weiteren bereichert JACK3 das Lernerlebnis, indem es den Schülern ansprechende und personalisierte Bewertungen bietet, begleitet von sofortigem Feedback. Die Plattform erleichtert die Datensammlung zur Überwachung des Lernfortschritts der Schüler und zur Identifizierung von Verbesserungsmöglichkeiten. Sie unterstützt verschiedene Lehr- und Lernmethoden, darunter das umgekehrte Klassenzimmer, Blended Learning und Fernunterricht [29].

Darüber hinaus bietet JACK3 eine Vielzahl weiterer Vorteile, wie die Steigerung der Schülerbeteiligung und -motivation, die Möglichkeit zur schnellen Bewertung und Rückmeldung von Aufgaben sowie die Prävention von Betrug und Plagiat. Zusätzlich bietet es erweiterte Möglichkeiten zur Verfolgung der Schulentwicklung. JACK3 bietet eine breite Palette von Übungstypen an, die den unterschiedlichen Lernzielen der Pädagogen und den spezifischen Bedürfnissen ihrer Schüler gerecht werden. Zu den Beispielen gehören Multiple-Choice-Quizfragen, Wahr/Falsch-Quizfragen, Zuordnungsübungen, Lückentextübungen, Kurzantwortfragen Aufsatzfragen und vieles mehr. Des Weiteren ermöglicht JACK3 die Erstellung komplexerer Übungen wie Drag-and-Drop-Aufgaben, Hotspot-Übungen, Umfragen zur Datensammlung und Webquests, die eine tiefgehende Exploration von Themen ermöglichen [29].

Es ist erwähnenswert, dass JACK3 seine Fähigkeiten zur Unterstützung der Bewertung von Übungen zur Unified Modeling Language (UML) erweitert hat, was seine Anwendbarkeit in der Softwaretechnik-Ausbildung verbessert. Lehrer können das System nutzen, um UML-Diagramme wie Klassendiagramme, Sequenzdiagramme und Aktivitätsdiagramme automatisch zu bewerten.

Im Bereich der UML-Übungen bietet JACK3 Lehrern die Flexibilität, Übungen zu gestalten, die Klassendiagramme, Sequenzdiagramme und Aktivitätsdiagramme umfassen. Diese Übungen ermöglichen es den Schülern, die Struktur und das Verhalten von Softwaresystemen zu visualisieren, die komplexen Beziehungen zwischen Systemkomponenten zu verstehen, ihre Ideen effektiv an Kommilitonen zu kommunizieren und ihre Fähigkeiten im

Bereich Problemlösung und kritisches Denken zu schärfen [29].

Das Hauptaugenmerk dieser wissenschaftlichen Arbeit liegt auf der Analyse und Evaluierung des Instrumentariums JACK, insbesondere seiner Kapazität zur kritischen Beurteilung von UML-Diagrammen. Im darauf folgenden Abschnitt wird die spezifische Problematik in umfassender Weise erörtert und eine breite Palette möglicher Lösungsansätze erörtert.

3. Problemanalyse

In diesem Kapitel erfolgt eine gründliche Analyse der zugrunde liegenden Problematik. Die Identifikation und eingehende Diskussion der spezifischen Herausforderungen und Schwierigkeiten, die im Rahmen dieser Arbeit behandelt werden, stehen dabei im Fokus.

3.1 Beschreibung des Bewertungsprozesses

Wie bereits in den vorherigen Kapiteln erwähnt, verwendet die Fakultät für Informatik der Universität Duisburg-Essen (UDE) ein elektronisches Bewertungssystem namens "JACK" [29], um bestimmte Prüfungen und Übungen automatisch zu bewerten. Unter den verschiedenen Arten von Übungen konzentriert sich diese Masterarbeit auf Übungen vom Typ UML.

Eine Übung vom Typ UML-Klassendiagramm besteht darin, die statische Struktur eines Software-Systems mithilfe der UML-Modellierungssprache zu modellieren [33]. Den Studierenden wird in der Regel eine Beschreibung des Systems, der wichtigsten Entitäten und ihrer Beziehungen zur Verfügung gestellt, und sie müssen dann ein Klassendiagramm erstellen, das diese Elemente darstellt. In diesem Diagramm sind die Klassen die Hauptobjekte, mit ihren Attributen (Variablen) und Methoden (Funktionen), und die Beziehungen zwischen den Klassen werden durch Assoziationen, Aggregationen oder Kompositionen dargestellt. Die Studierenden müssen besonders auf die Genauigkeit der Namen, Multiplizitäten und Kardinalitäten achten, um die Struktur des Systems korrekt widerzuspiegeln [33]. Das Hauptziel dieser Übung ist es, das Verständnis der objektorientierten Modellierungskonzepte zu vertiefen und eine solide Grundlage für die Softwareentwicklung zu schaffen. In Bezug auf das JACK-System kann der Bewertungsprozess in mehrere verschiedene Phasen unterteilt werden:

Phase 1: Erstellung der Übung

In dieser ersten Phase erstellt der Lehrer die Übung, indem er eine

schriftliche Beschreibung eines zu modellierenden Systems bereitstellt. Besonderes Augenmerk wird auf die Genauigkeit der Bezeichnungen der Entitäten und die Klarheit und Explizitheit der Beziehungen zwischen ihnen gelegt. Der Lehrer hat dann einen Bereich in der JACK-Anwendung, in den er diesen Text einfügen kann, der von den Studierenden eingesehen wird.

Phase 2: Erstellung einer Musterlösung oder Anmerkung (optional)

In dieser Phase kann der Lehrer entscheiden, ein UML-Diagramm als Musterlösung für die Übung zu erstellen. Dies erleichtert das Verständnis der Übung, ermöglicht die Überprüfung der Kohärenz und Durchführbarkeit des Systems. Alternativ kann der Lehrer die Übung lediglich annotieren, um die Schlüsselbegriffe und -elemente hervorzuheben, die in der späteren Phase nützlich sein werden.

Phase 3: Entwicklung des GReQL-Codes

In dieser Phase verwendet der Lehrer seine Anmerkungen und die Musterlösung, um den GReQL-Code zu entwickeln, der von JACK interpretiert wird, um die von den Studierenden eingereichten Lösungen zu bewerten. JACK verwendet den GReQL-Code sowie verschiedene in dieser Sprache definierte Regeln, um die UML-Diagramme zu bewerten. Wenn ein Student seine Lösung einreicht, wird eine grafische Darstellung dieser Lösung erstellt, und der GReQL-Code führt Abfragen auf dieser Darstellung aus, um eine Note für die Lösung des Studenten zu vergeben [39].

Phase 4: Einreichung der Lösung durch den Studenten

Nachdem der Student an der Lösung der Übung gearbeitet hat, lädt er ein XML-Dokument im XMI Format auf JACK hoch. Zur Generierung dieses kompatiblen XMI-Dokuments verwenden Studierende Tools wie BOUML [31] oder Software Ideas Modeler [34], mit denen sie XMI-Code aus einer zuvor erstellten grafischen UML Darstellung ableiten können. Dieses Dokument repräsentiert die von ihm entwickelte Lösung.

Phase 5: Bewertung des Diagramms

Das von den Studierenden eingereichte Diagramm wird anhand des von den Lehrern verfassten GReQL-Codes bewertet. Anschließend wird dem Studenten eine Note zugeteilt.

Diese Phasen veranschaulichen die grundlegende Funktionsweise der JACK-Plattform in Bezug auf die automatisierte Bewertung von Übungen zur Erstellung von UML-Klassendiagrammen.

3.2 Untersuchung des Problems

Im Verlauf der dritten Phase, wie im vorherigen Kapitel beschrieben, sehen sich Lehrende mit der Notwendigkeit konfrontiert, GReQL-Code zu

verfassen, der vom JACK-System zur Bewertung der Einreichungen verschiedener Studierender verwendet wird. Diese Phase stellt jedoch bereits auf verschiedenen Ebenen eine Herausforderung für die Lehrenden dar, aus folgenden Gründen:

Erforderliche Expertise für die Erstellung von GReQL-Code: Das Verfassen von GReQL-Code erfordert eine gewisse Expertise. Auf den ersten Blick mag die Syntax von GReQL nicht schwer zu verstehen sein. Dennoch kann es anspruchsvoll sein, die Feinheiten der Code-Erstellung zu beherrschen. Dies erfordert von den Lehrenden mehrere Stunden, intensives Üben und umfangreiche Tests, um einen Code zu erstellen, der präzise bewertet, insbesondere in Fällen von komplexeren Beziehungen zwischen Entitäten.

Hohe Möglichkeit von Fehlern im GReQL-Code: Selbst bei Beherrschung der Feinheiten von GReQL sind Lehrende nicht vor möglichen Fehlern gefeit. Diese Fehler können zwar geringfügig erscheinen, haben jedoch das Potenzial, das Bewertungssystem erheblich zu beeinträchtigen.

Zeitaufwand für die Erstellung von GReQL-Code: Die Bewertung jedes Diagramms erfordert erheblichen Zeitaufwand, um einen GReQL-Bewertungscode zu erstellen, insbesondere für komplexere Beziehungen zwischen verschiedenen Entitäten. Die Zeit, um die erforderliche Expertise zu erlangen, kann stark variieren und erfordert zahlreiche Stunden intensiver Übung.

Wartung und Anpassungsfähigkeit des GReQL-Codes: Nachdem der GReQL-Code erstellt und für die Bewertung von UML-Diagrammen implementiert wurde, ist kontinuierliche Wartung erforderlich. Mit steigenden Anforderungen an die Bewertung muss der GReQL-Code regelmäßig aktualisiert und angepasst werden. Dies stellt für Lehrende eine fortwährende Herausforderung dar und erfordert Zeit und Aufwand, um sicherzustellen, dass das Bewertungssystem präzise und relevant bleibt.

Bedarf an Ressourcen und technischer Unterstützung: Lehrende benötigen möglicherweise Zugang zu technischen Ressourcen und angemessener Unterstützung, um die Kunst der effizienten GReQL-Code-Erstellung zu beherrschen. Dies kann Schulungen, Orientierungsdokumente, Diskussionsforen oder andere Formen technischer Unterstützung einschließen. Die Beschaffung dieser Ressourcen kann Zeit in Anspruch nehmen und eine institutionelle Koordination erfordern.

Dies sind die verschiedenen Herausforderungen, die bei der Erstellung von GReQL-Code für die Bewertung von UML-Diagrammen auf der JACK-Plattform auftreten können.

3.3 Ableitung und Abgrenzung der Anforderungen

Das Ziel dieser Masterarbeit besteht nicht darin, Lehrkräfte von der Notwendigkeit zu befreien, GReQL-Code zu schreiben oder zu bearbeiten. Vielmehr geht es darum, den Prozess der Regeldefinition erheblich zu erleichtern. Um diesen Prozess zu vereinfachen, zielt diese Masterarbeit darauf ab, ein Verfahren zu entwickeln und in Form einer Softwareanwendung zu prototypisieren, mit dem (halb-)automatisch Bewertungsregeln aus annotierten Musterlösungen erstellt werden können. Dieses System soll in der Lage sein, folgende Ziele zu erreichen:

1. **Verminderung der Einstiegshürde:** Es strebt danach, den Prozess der Erstellung von GReQL-Regeln zu erleichtern. Selbst jemand, der keine Vorkenntnisse in GReQL hat, sollte unser Tool verwenden können, um Regeln zu erstellen, die bereits eine Bewertung von einfachen Diagrammen und Beziehungen ermöglichen.
2. **Verbesserung der Präzision und Zuverlässigkeit der GReQL-Regeln:** Durch die Nutzung von annotierten Musterlösungen zielt die Anwendung darauf ab, die Präzision und Zuverlässigkeit der generierten Regeln zu verbessern. Dies würde das Risiko von Fehlern und Inkonsistenzen in den Bewertungsregeln verringern und somit eine konsistentere Bewertung der studentischen Arbeiten gewährleisten.
3. **Optimierung von Zeit und Aufwand der Lehrkräfte:** Durch Automatisierung eines Teils des Regelbildungsprozesses zielt das System darauf ab, Zeit und Aufwand der Lehrkräfte zu sparen. Dies würde es ihnen ermöglichen, sich mehr auf das Lehren und Betreuen der Studierenden zu konzentrieren, anstatt auf mühsame administrative Aufgaben.
4. **Förderung der Skalierbarkeit und Wartbarkeit der Regeln:** Durch die Implementierung eines Mechanismus zur Pflege und Aktualisierung der generierten Regeln würde die Anwendung dazu beitragen sicherzustellen, dass die Regeln relevant und anpassungsfähig bleiben und sich den sich ändernden Anforderungen von Lehre und Bewertung anpassen.
5. **Unterstützung einer breiten Palette von Bewertungsszenarien:** Die Anwendung sollte flexibel genug sein, um eine Vielzahl von Bewertungsszenarien zu unterstützen, einschließlich solcher mit komplexen Diagrammen und Beziehungen. Dadurch würde sie eine vielseitige Lösung für Lehrkräfte bieten.

Das Ziel dieser Initiative ist es, den Prozess der Bewertung von UML-Diagrammen effizienter, zugänglicher und präziser zu gestalten, während die administrativen Arbeitslasten der Lehrkräfte reduziert werden. Dadurch soll die Qualität von Lehre und Bewertung im Bereich der Softwaremodellierung verbessert werden.

4. Entwicklung des Konzepts

In diesem Kapitel wird der Übergang von der Problemanalyse zur kreativen Konzeptentwicklung im Forschungsprozess hervorgehoben. Es markiert die zentrale Phase, in der theoretisches Wissen und praktische Erkenntnisse zusammenfließen, um innovative Lösungsansätze zu formulieren. Wir integrieren theoretische Grundlagen und Modelle aus vorangegangenen Kapiteln und erläutern die angewandten Methoden zur Entwicklung tragfähiger Lösungen. Dieses Kapitel dient als Brücke zwischen Analyse und Umsetzung, betont die kreative Synthese von Ideen und Lösungsansätzen und bietet einen detaillierten Einblick in den Prozess der Konzeptentwicklung.

4.1 Verschiedene Herangehensweisen

Vor der Ausarbeitung des endgültigen Implementierungskonzepts wurden mehrere methodische Ansätze einer eingehenden Untersuchung unterzogen. Das vorrangige Ziel dieser methodischen Herangehensweisen bestand in der Extraktion einer Reihe von Regelwerken, die später in GReQL-Ausdrücke transformiert werden sollten. Diese Regelwerke sollten aus den Kommentaren oder Anmerkungen extrahiert werden, welche von Lehrkräften in Bezug auf eine UML-Übung hinterlassen wurden. Die Wahl des Notenformats ist frei, was viele Möglichkeiten bietet.

4.1.1 YAML-basierten Annotationen

Der erste untersuchte methodische Ansatz fokussierte auf die Entwicklung eines benutzerfreundlichen Annotationsystems, das eine umfassende Modellierung der Interaktionen innerhalb eines UML-Diagramms ermöglichen sollte. Infolgedessen wurden editierbare Regelobjekte abgeleitet, welche aus diesem Annotationsystem generiert wurden. Sobald der Nutzer sämtliche Regeln, die aus dem Annotationsystem resultierten, verifizierte, konnte er GReQL-Code generieren, welcher anschließend in das JACK-System ein-



Abbildung 4.1: Beispiel einer Annotation von UML mit YAML

gefügt wurde. Die Konzeption dieses schlichten Notationssystems diene dem Zweck der Abbildung von Beziehungen innerhalb eines UML Klassendiagramms. Zu diesem Zweck wurde das YAML-Format aus mehreren Gründen präferiert:

1. Erstens, das Schreiben im YAML-Format erweist sich als unkompliziert.
2. Zweitens, es liefert Daten in einem strukturierten Format, welches leicht manipuliert werden kann.

Die Notation in YAML mag den Eindruck vermitteln, bereits eine Form der Regeldefinition darzustellen, weil der Lehrende bereits alle Interaktionen zwischen den verschiedenen Objekten im Diagramm definiert. Um jedoch die Klarheit bezüglich dieser Frage zu gewährleisten, könnte es sinnvoll sein, ein weniger formelles Notationssystem zu etablieren. Beispielsweise:

- Die Klasse A erbt von der Klasse B.
- Die Klasse A besitzt drei Attribute:
 - Attribut a vom Typ x mit öffentlicher Sichtbarkeit.
 - Attribut b vom Typ y mit öffentlicher Sichtbarkeit.
 - Attribut c vom Typ z mit privater Sichtbarkeit.

oder ein noch weniger wortreiches System:

- $A \Rightarrow B$
- A:
 - +a:x (+ für öffentlich)
 - +b:y
 - -c:z (- für privat)

Dieses System ähnelt eher herkömmlichen Kommentaren, jedoch könnte es herausfordernder sein, den darin enthaltenen Text zu analysieren und daraus anwendbare Regelvorschläge zu extrahieren. Trotzdem birgt dieser methodische Ansatz mehrere Limitationen:

1. Es existieren bereits seit geraumer Zeit vielfältige Notationssysteme und Datenformate für UML, wie beispielsweise XMI [35], UML eXchange Format (UXF) [41], Textbasierte [44] oder JSON-basierte Datenformate [7], um nur einige zu nennen.
2. Die Entwicklung eines vollständig neuen und umfassenden Notationssystems innerhalb eines begrenzten Zeitraums kann äußerst zeitintensiv sein, insbesondere wenn es die umfassende Darstellung sämtlicher Varianten eines UML-Klassendiagramms anstrebt.
3. Dieses System verlagert die Aufgabe der Ableitung von Beziehungen zwischen den verschiedenen Objekten lediglich in ein anderes Format, ohne dabei die Arbeitsbelastung für die Lehrkraft zu mindern.

In Anbetracht dieser diversen Überlegungen wurde von der Verfolgung dieses methodischen Ansatzes Abstand genommen, und es erfolgte die Exploration alternativer Vorgehensweisen.

4.1.2 Verwendung von Natural Language Processing Tools

Die zweite konzeptionelle Idee besteht darin, ein Natural Language Processing (NLP)-Tool zu verwenden, um die Kommentare oder Notizen des Lehrers zu analysieren. Anschließend kann das NLP-Tool verwendet werden, um Empfehlungen auf der Grundlage einer vorher festgelegten Reihe von Parametern für den Lehrer abzugeben. Diese Parameter könnten eine Reihe von UML-Regeln in einem spezifischen Format sein, so dass die KI in der Lage ist, die richtige Lösung auf der Grundlage des Kommentars zu finden. Der Vorschlag könnte dann in GReQL-Code übersetzt werden. Diese Herangehensweise hat jedoch auch einige Herausforderungen:

1. Die von dieser Methode generierten Lösungen neigen dazu, ungenau zu sein, und es sind möglicherweise zahlreiche Anpassungen erforderlich, um das gewünschte Ergebnis zu erzielen [11]. Dies könnte

möglicherweise die Arbeitsbelastung der Lernenden erhöhen, anstatt sie zu erleichtern.

2. Es wäre notwendig, die KI darauf zu trainieren, bestimmte Muster zu erkennen und sie mit vordefinierten Regeln in Verbindung zu bringen, um die Zuverlässigkeit zu erhöhen.

In ihrem aktuellen Stand erfordert diese Herangehensweise erhebliche Anstrengungen, um diese Herausforderungen zu bewältigen und eine effektive Implementierung zu erreichen. Aufgrund der potenziellen Schwierigkeiten bei der Implementierung und der möglicherweise nicht zufriedenstellenden Ergebnisse wurde dieser Ansatz auch aufgegeben.

4.2 Konzept

Bei der Entwicklung des Konzepts wurde ein radikal andersartiger Ansatz als der in dem vorherigen Abschnitt dargelegte verfolgt. Die Herangehensweise, Regeln aus Kommentaren abzuleiten, erweist sich als ein komplexer Prozess und potenziell schwierig umzusetzen. Die Gewährleistung der Zuverlässigkeit der erzielten Ergebnisse stellt ebenfalls eine bedeutende Herausforderung dar. In diesem Zusammenhang wurde eine intuitivere Herangehensweise in Betracht gezogen, nämlich die direkte Ableitung von Regeln aus der Mustervorlage.

Bevor dieses Konzept im Detail beschrieben wird, ist es unerlässlich, das Endziel in Erinnerung zu rufen, nämlich die Unterstützung von Lehrenden bei der Bewertung von UML-Übungsaufgaben (insbesondere Klassendiagrammen). Dies soll durch erhebliche Vereinfachung des Schreibens von GReQL-Code auf der JACK-Plattform erreicht werden, indem Lehrern mit Hilfe eines im Rahmen dieser Masterarbeit zu entwickelnden Tools (halb-)automatische Unterstützung geboten wird. Das Konzept zur Erfüllung dieser Aufgabe kann in drei Wichtige Schritte unterteilt werden:

Schritt 1: Erstellung einer Musterlösung mit Hilfe eines Modellierungstools

Wie im Abschnitt 3.1 erwähnt wurde, umfasst Phase 2 die optionalen Schritte zur Erstellung einer Musterlösung. In diesem Konzept ist diese Phase unerlässlich und gewinnt ihre volle Bedeutung. Nachdem ein UML-Klassendiagrammübungsaufgabe entwickelt wurde, erstellen die meisten Lehrer eine Musterlösung, oft in Form eines Klassendiagramms. Diese Lösung wird dann mit den verschiedenen Einreichungen der Studenten

verglichen, und Punkte werden gemäß den vom Lehrer festgelegten Kriterien an die einzelnen Studenten vergeben. Da in den meisten Fällen eine Musterlösung in Form eines Klassendiagramms für die Aufgabe vorhanden ist, wäre es sinnvoll, diese zur Ableitung relevanter Regeln zu verwenden. Dies stellt die erste Phase des Ansatzes dar, bei der eine Musterlösung mithilfe einer Modellierungsanwendung erstellt wird, die anschließend die Möglichkeit bietet, das Diagramm in einem leicht programmierbaren Format zu exportieren.

Schritt 2: Verarbeitung der exportierten Datei

Nachdem der Lehrer die Musterlösung erfolgreich modelliert hat, ergibt sich die Notwendigkeit, diese Vorlage in ein geeignetes Format zu exportieren, welches eine programmgesteuerte Bearbeitung ermöglicht. Dieser Transformationsprozess kann mittels diverser Datenformate wie JSON, XML oder sogar YAML realisiert werden. Sobald das exportierte Dokument verfügbar ist, eröffnet sich im Kontext eines Klassendiagramms die Möglichkeit, zunächst sämtliche in der Musterlösung enthaltenen Klassen sowie sämtliche Interdependenzen zwischen den verschiedenen Elementen zu extrahieren. Infolge dieses Phasenschritts besteht die Option zur Ableitung von "Regel"-Objekten, welche auf der Frontend-Oberfläche sichtbar sind und vom Lehrer interaktiv bearbeitet werden können. Die Bereitstellung dieser Funktionalität befähigt den Lehrer dazu, wertvolles Feedback hinzuzufügen, die zugehörigen Punktzahlen zu definieren und sämtliche erforderlichen Informationen für den nächsten Schritt und die anschließende Evaluierung umfassend zu vervollkommen.

Schritt 3: Generierung des GReQL-Codes

Der abschließende Schritt dieses Vorgehens umfasst die Generierung von GReQL-Code aus den Regeln, die in der vorherigen Phase extrahiert und/oder hinzugefügt wurden. Nachdem die Konfiguration abgeschlossen ist, muss der GReQL-Code unter Verwendung verschiedener XMI-Vorlagen generiert werden. Der resultierende Code kann vom Lehrer auf der JACK-Plattform verwendet werden.

Durch dieses Konzept (siehe 4.2) und seine verschiedenen Schritte ist es möglich, von einer Musterlösung zur Generierung des erforderlichen GReQL-Codes für die Bewertung von Diagrammen durch JACK zu gelangen. Dieser Ansatz hat das Potenzial, den Bewertungsprozess von Diagrammen über die JACK-Plattform erheblich zu vereinfachen, was das zentrale Ziel dieser Masterarbeit ist. Die tatsächliche Umsetzung dieses



Abbildung 4.2: Repräsentatives schema des konzepts.

Konzepts hängt jedoch eng vom gewählten Workflow und der Art und Weise ab, wie das Konzept umgesetzt wird. Dieser Ausblick leitet zum nächsten Kapitel über, das die Implementierung behandelt.

5. Implementierung

Das vorliegende Kapitel widmet sich der umfassenden Dokumentation des Implementierungsprozesses des zuvor beschriebenen Konzepts. Es bietet eine detaillierte Aufarbeitung der technischen Umsetzung und des Entwicklungsprozesses, der im Rahmen dieser Masterarbeit durchgeführt wurde. Beginnend mit einer umfassenden Beschreibung der verwendeten Technologien und Werkzeuge sowie einer ausführlichen Begründung für die Wahl dieser spezifischen Technologien, wird dieses Kapitel einen tiefen Einblick in die präzise Umsetzung des Konzepts gewähren. Die Implementierung ist ein entscheidender Schritt zur Realisierung des in den vorherigen Kapiteln skizzierten Ansatzes zur Bewertung von UML-Diagrammen. Durch die Dokumentation dieses Schrittes wird das Verständnis für die technischen Aspekte des Projekts vertieft und ermöglicht eine transparente Darstellung des Entwicklungsprozesses.

6. Evaluation

7. Diskussion

8. Zusammenfassung und Ausblick

Literatur

- [1] O. M. 2.0. *XML Metadata Interchange (XMI)*. 2007.
URL: <http://www.omg.org/spec/XMI/2.1.1/> ..
- [2] W. Abramowicz.
Business Information Systems: 16th International Conference, BIS 2013, Poznań, Poland, June 19-21, 2013, Proceedings. Bd. 157.
Springer, 2013.
- [3] N. Alruwais, G. Wills und M. Wald.
„Advantages and challenges of using e-assessment“.
In: *International Journal of Information and Education Technology* 8.1 (2018), S. 34–37.
- [4] O. Anas, T. Mariam und L. Abdelouahid. „New method for summative evaluation of UML class diagrams based on graph similarities“.
In: *International Journal of Electrical and Computer Engineering* 11.2 (2021), S. 1578–1590.
- [5] L. Auxepaules, M. Alonso, M. Alonso, L. Auxepaules, D. Py und D. Py. „Diagram, a Learning Environment for Initiation to Object-Oriented Modelling with UML Class Diagrams“. In: (2015).
- [6] J. Azevedo, E. P. Oliveira und P. D. Beites.
„E-assessment and multiple-choice questions: a literature review“.
In: *Handbook of Research on E-Assessment in Higher Education* (2019), S. 1–27.
- [7] T. Benson, G. Grieve, T. Benson und G. Grieve.
„UML, XML and JSON“. In: *Principles of Health Interoperability: FHIR, HL7 and SNOMED CT* (2021), S. 399–426.
- [8] Blackboard.
Educational Technology Services — Blackboard — North America.
Accessed on September 6, 2023. 2023.
URL: <https://www.blackboard.com/>.

- [9] G. Booch. „The History of Software Engineering“. In: *IEEE Software* 35.5 (2018), S. 108–114.
DOI: 10.1109/MS.2018.3571234.
- [10] Y. Boubekur, G. Mussbacher und S. McIntosh.
„Automatic assessment of students’ software models using a simple heuristic and machine learning“. In: *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems: Companion Proceedings*. 2020, S. 1–10.
- [11] K. Chowdhary und K. Chowdhary. „Natural language processing“. In: *Fundamentals of artificial intelligence* (2020), S. 603–649.
- [12] B. Eilers, S. Gruttmann und H. Kuchen.
„Konzeption eines integrierbaren Systems zur computergestützten Lernfortschrittskontrolle“. In: *E-learning-management* (2008), S. 213–232.
- [13] R. Fauzan, D. Siahaan, S. Rochimah und E. Triandini. „A different approach on automated use case diagram semantic assessment“. In: *International Journal of Intelligent Engineering and Systems* (2021).
- [14] R. Fauzan, D. Siahaan, S. Rochimah und E. Triandini.
„Class diagram similarity measurement: a different approach“. In: *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*. IEEE. 2018, S. 215–219.
- [15] B. GmbH. *OPAL LERNPLATTFORM*.
Accessed on September 6, 2023. 2023.
URL: <https://www.bps-system.de/opal-lernplattform/>.
- [16] H. Gregersen und C. S. Jensen.
„Temporal entity-relationship models-a survey“. In: *IEEE Transactions on knowledge and data engineering* 11.3 (1999), S. 464–497.
- [17] S. Gross, B. Mokbel, B. Hammer und N. Pinkwart.
„Feedback provision strategies in intelligent tutoring systems based on clustered solution spaces“. In: (2012).
- [18] S. Gruttmann. „Formatives E-Assessment in der Hochschullehre: computerunterstützte Lernfortschrittskontrollen im Informatikstudium“. Diss. Zugl.: Münster (Westfalen), Univ., Diss., 2010, 2009.

- [19] J. Holland, N. Baghaei, M. Mathews und A. Mitrovic.
„The effects of domain and collaboration feedback on learning in a collaborative intelligent tutoring system“.
In: *International Conference on Artificial Intelligence in Education*. Springer. 2011, S. 469–471.
- [20] J. Hollingsworth. „Automatic Graders for Programming Classes“.
In: *Commun. ACM* 3.10 (Okt. 1960), S. 528–529. ISSN: 0001-0782.
DOI: 10.1145/367415.367422.
URL: <https://doi.org/10.1145/367415.367422>.
- [21] C. R. Huyck und S. L. Lytinen.
„Efficient heuristic natural language parsing“. In: *Proceedings of the eleventh national conference on Artificial intelligence*. 1993, S. 386–391.
- [22] R. P. Jayawardena, G. D. Thiwanthi, P. S. Suriyaarachchi, K. I. Withana und C. Jayawardena. „Automated Exam Paper Marking System for Structured Questions and Block Diagrams“.
In: *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*. 2018, S. 1–5.
DOI: 10.1109/ICIAfS.2018.8913351.
- [23] R. P. Jayawardena, G. D. Thiwanthi, P. S. Suriyaarachchi, K. I. Withana und C. Jayawardena. „Automated Exam Paper Marking System for Structured Questions and Block Diagrams“.
In: *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*. 2018, S. 1–5.
DOI: 10.1109/ICIAfS.2018.8913351.
- [24] T. Khmour.
„A semantic assessment framework for e-learning systems“.
In: *International Journal of Knowledge and Learning* 13.2 (2020), S. 110–122.
- [25] S. Kocdar, A. Karadeniz, R. Peytcheva-Forsyth und V. Stoeva.
„Cheating and plagiarism in e-assessment: Students’ perspectives“.
In: *Open Praxis* 10.3 (2018), S. 221–235.
- [26] Q. Li und Y.-L. Chen. „Data flow diagram“.
In: *Modeling and Analysis of Enterprise and Information Systems*. Springer, 2009, S. 85–97.
- [27] Moodle. *Moodle LMS - Moodle-Lernplattform*.
Accessed on September 6, 2023. 2023. URL: <https://moodle.de/>.

- [28] D. Moody. „The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering“. In: *IEEE Transactions on software engineering* 35.6 (2009), S. 756–779.
- [29] B. Mschypula M. Striewe. *JACK - WIKI*. Accessed on September 6, 2023. 2023. URL: <https://wiki.uni-due.de/jack/index.php?title=Hauptseite>.
- [30] A. Outair, M. Tanana und A. Lyhyaoui. „TOWARDS AN AUTOMATIC EVALUATION OF UML CLASS DIAGRAMS BY MEASURING GRAPH SIMILARITY.“ In: *Journal of Theoretical & Applied Information Technology* 95.4 (2017).
- [31] B. Pagès. *BOUML*. 2021. URL: <https://www.bouml.fr/index.html>.
- [32] C. A. Petri und W. Reisig. „Petri net“. In: *Scholarpedia* 3.4 (2008), S. 6477.
- [33] G. Reggio, M. Leotta, F. Ricca und D. Clerissi. „What are the used UML diagrams? A Preliminary Survey.“ In: *EESSMod@ MoDELS*. 2013, S. 3–12.
- [34] I. D. Rodina. *Software Ideas Modeller*. 2023. URL: <https://www.softwareideas.net/>.
- [35] D. Skogan. „UML as a schema language for XML based data interchange“. In: *Proceedings of the 2nd International Conference on The Unified Modeling Language (UML'99)*. Citeseer. 1999.
- [36] R. Sousa und J. P. Leal. „A structural approach to assess graph-based exercises“. In: *Languages, Applications and Technologies: 4th International Symposium, SLATE 2015, Madrid, Spain, June 18-19, 2015, Revised Selected Papers 4*. Springer. 2015, S. 182–193.
- [37] U. Stöðberg. „A research review of e-assessment“. In: *Assessment & Evaluation in Higher Education* 37.5 (2012), S. 591–604. DOI: 10.1080/02602938.2011.557496. eprint: <https://doi.org/10.1080/02602938.2011.557496>. URL: <https://doi.org/10.1080/02602938.2011.557496>.
- [38] M. Striewe. „Automated Assessment of Software Artefacts-A Use Case in E-Assessment“. In: *University of Duisburg-Essen* (2014).
- [39] M. Striewe und M. Goedicke. „Automated checks on UML diagrams“. In: *Proceedings of the 16th annual joint conference on Innovation and technology in computer science education*. 2011, S. 38–42.

- [40] T. Sturm, J. von Voss und M. Boger.
„Generating code from UML with velocity templates“.
In: *International Conference on the Unified Modeling Language*.
Springer. 2002, S. 150–161.
- [41] J. Suzuki und Y. Yamamoto.
„Making UML models interoperable with UXF“.
In: *International Conference on the Unified Modeling Language*.
Springer. 1998, S. 78–91.
- [42] M. Ullrich et al.
„Automated assessment of conceptual models in education“.
In: *Manuskript unter Begutachtung* (2021).
- [43] S. Valenti, F. Neri und A. Cucchiarelli.
„An overview of current research on automated essay grading“.
In: *Journal of Information Technology Education: Research* 2.1
(2003), S. 319–330.
- [44] H. Washizaki, M. Akimoto, A. Hasebe, A. Kubo und Y. Fukazawa.
„TCD: A text-based UML class diagram notation and its model
converters“. In: *Advances in Software Engineering: International
Conference, ASEA 2010, Held as Part of the Future Generation
Information Technology Conference, FGIT 2010, Jeju Island, Korea,
December 13-15, 2010. Proceedings*. Springer. 2010, S. 296–302.
- [45] S. A. White. „Introduction to BPMN“. In: *Ibm Cooperation* (2004).

A. Textbausteine

A.1 Erster Baustein

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe. Ich habe alle Stellen, die ich aus den Quellen wörtlich oder inhaltlich entnommen habe, als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Michael Vivian Mboni Saha, Campus Essen, den 30.03.2024