

**Gottfried Wilhelm Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Praktische Informatik
Fachgebiet Software Engineering**

**Entwicklung eines Konzepts zur Beurteilung
von gemeinsamen Verständnis mittels Vision
Videos in Requirements Engineering basierend
auf einer systematischen Literatursuche**

Development of a Concept for Assessing Shared
Understanding by Means of Vision Videos in
Requirements Engineering Based on a
Systematic Literature Review

Masterarbeit

im Studiengang Informatik
von

Robert Völkner

**Prüfer: Prof. Dr. rer. nat. Kurt Schneider
Zweitprüfer: Dr. rer. nat. Jil Klünder
Betreuer: M. Sc. Oliver Karras**

Hannover, 4. November 2020

Zusammenfassung

Entwicklung eines Konzepts zur Beurteilung von gemeinsamen Verständnis mittels Vision Videos in Requirements Engineering basierend auf einer systematischen Literatursuche

Ein essentieller Aspekt für ein erfolgreiches Softwareprojekt ist das Erzeugen eines gemeinsamen Verständnisses zwischen den Stakeholdern und den Entwicklern. Dieses gemeinsame Verständnis ist eine Voraussetzung für die erfolgreiche Kommunikation der Anforderungen, die an das zu entwickelnde System gestellt werden. In der Literatur existieren hierfür bereits diverse Ansätze die beschreiben, mit welchen Methoden ein gemeinsames Verständnis erzeugt werden kann. Einige dieser Ansätze sehen dabei die Nutzung von Videos vor, da durch die visuelle Darstellung der Anforderungen an das System mögliche Sprachbarrieren umgangen werden, die sonst bei einer traditionellen textuellen Kommunikation zu Problemen führen können. Die Darstellung von visionären Nutzungsszenarien eines Systems in sogenannten Vision Videos stellt dabei eine mögliche Umsetzung dieses Ansatzes dar. In diesem Zusammenhang existieren zwar einige Publikationen die zeigen, dass durch Vision Videos ein gemeinsames Verständnis erzeugt werden kann, jedoch gibt es noch keinen Ansatz, in dem eine allgemeingültige Methode zur Beurteilung oder Messung von gemeinsamen Verständnis definiert wird.

Aus diesem Grund wird innerhalb dieser Arbeit nach Methoden, Metriken und Heuristiken gesucht, mit denen gemeinsames Verständnis gemessen beziehungsweise beurteilt werden kann, das auf Basis von Vision Videos erzeugt wurde. Zu diesem Zweck wird eine systematische Literatursuche durchgeführt, um geeignete Methoden, Metriken und Heuristiken in der Literatur zu identifizieren. Die identifizierten Methoden, Metriken und Heuristiken werden anschließend in den Kontext von Vision Videos übertragen, um sie für diese Arbeit nutzbar zu machen. Zuletzt liefert diese Arbeit ein Experimentdesign, in dem konkrete Ansatzpunkte für die Nutzung der Methoden, Metriken und Heuristiken zur Messung und Beurteilung von gemeinsamen Verständnis beschrieben werden.

Abstract

Development of a Concept for Assessing Shared Understanding by Means of Vision Videos in Requirements Engineering Based on a Systematic Literature Review

An essential aspect for a successful software project is the creation of a shared understanding between stakeholders and developers. This shared understanding is a prerequisite for the successful communication of the requirements that are posed to the system to be developed. There are many approaches in literature that describe methods with which it is possible to create a shared understanding. Some of those approaches intend to use videos, since their visual representation of the requirements of the system can bypass language barriers that would pose problems for traditional textual communication. The representation of visionary scenarios of use of a system in so called vision videos presents a possible implementation of this approach. While there are some publications in this context that show that vision videos can create shared understanding, there are no approaches that define methods for the assessment or measurement of shared understanding.

For that reason, within this work methods, metrics and heuristics are sought which can assess or measure shared understanding that was created via vision videos. For this purpose, a systematic literature review is conducted to identify fitting methods, metrics and heuristics. Furthermore, these methods, metrics and heuristics are transferred into the context of vision videos to make them usable for this work. Finally, this work presents an experiment design which describes concrete approaches for the use of the methods, metrics and heuristics to measure and assess shared understanding.

Inhaltsverzeichnis

1 Einleitung.....	1
1.1 Motivation.....	1
1.2 Ziel der Arbeit.....	2
1.3 Struktur der Arbeit.....	3
2 Grundlagen.....	4
2.1 Mentale Modelle.....	4
2.2 Shared Understanding.....	5
2.3 Videonutzung im Requirements Engineering.....	7
2.4 Vision Video.....	9
2.5 Verwandte Arbeiten.....	11
3 Systematische Literatursuche.....	14
3.1 Struktur der Suche.....	14
3.2 Manuelle Suche.....	15
3.2.1 Ergebnisse.....	17
3.2.2 Suchstring.....	19
3.3 Datenbanksuche.....	21
3.3.1 Suchprozess.....	22
3.3.2 Ergebnisse.....	24
3.4 Snowballing.....	26
3.4.1 Suchprozess.....	26
3.4.2 Ergebnisse.....	27
4 Auswertung.....	30
4.1 Ergebnisübersicht.....	30
4.2 Effizienz und Effektivität.....	31
4.3 Jahresverteilung.....	33
4.4 Mentale Modelle vs. Shared Understanding.....	34
4.5 Inhaltlich.....	35
4.5.1 Mapping Konzepte.....	36
4.5.2 Numerische Konzepte.....	47
4.6 Übertragung auf Vision Videos.....	51
4.7 Threats to Validity.....	52
4.7.1 Construct Validity.....	52
4.7.2 Internal Validity.....	54
4.7.3 External Validity.....	54
4.7.4 Conclusion Validity.....	55
5 Experimentdesign.....	56
5.1 Zielfindung.....	57
5.2 Kontext.....	59
5.3 Hypothesen.....	60
5.4 Variablen.....	61

5.4.1 Unabhängige Variablen.....	61
5.4.2 Abhängige Variablen.....	62
5.5 Probandenauswahl.....	62
5.6 Experimentdesign.....	63
5.6.1 Allgemeine Designprinzipien.....	64
5.6.2 Wahl des Designtypen.....	65
5.7 Instrumentierung.....	66
5.8 Experimentablauf.....	67
5.8.1 Hypothesen.....	67
5.8.2 Methoden.....	68
5.9 Threats to Validity.....	69
5.9.1 Construct Validity.....	69
5.9.2 Internal Validity.....	70
5.9.3 External Validity.....	71
5.9.4 Conclusion Validity.....	71
6 Fazit und Ausblick.....	72
6.1 Fazit.....	72
6.2 Ausblick.....	73
Anhang.....	74
A) Liste des Quasi-Goldstandards und der hilfreichen Publikationen.....	74
B) Kombinationen der Suchbegriffe für SpringerLink.....	76
C) Angepasste Suchstrings für die jeweilige Syntax der Datenbanken.....	77
D) Liste relevanter Publikationen aus der systematischen Literatursuche.....	79
E) Patterns für Egozentrismus nach Suzuki und Tsuda [42].....	85
F) Tabellarische Übersicht der relevanten Publikationen.....	86
G) Beispiel einer Konzepttable für den Pathfinder Algorithmus.....	91
Literaturverzeichnis.....	92
Tabellenverzeichnis.....	97
Abbildungsverzeichnis.....	98
Erklärung der Selbstständigkeit.....	99

1 Einleitung

1.1 Motivation

Einer der wichtigsten Aspekte für die erfolgreiche Durchführung eines Softwareprojekts ist die Kommunikation zwischen den Stakeholdern und den Entwicklern [71]. Da jede der beteiligten Personen verschiedene Fähigkeiten besitzt und aus verschiedenen Domänen stammt ist es entscheidend, dass die Kommunikation zwischen allen Beteiligten gefördert wird, um ein gemeinsames Verständnis bezüglich des Softwareprojekts zu erreichen [71]. Die Förderung der Kommunikation zwischen den Projektbeteiligten ist einer der zentralen Bestandteile des Requirements Engineerings, welches laut Pohl und Rupp [72] einen systematischen und disziplinären Ansatz zur Spezifikation und zum Management von Anforderungen darstellt. Die Ziele des Requirements Engineerings sind dabei, die relevanten Anforderungen zu identifizieren, bei den Stakeholdern ein gemeinsames Verständnis bezüglich der Anforderungen zu erschaffen und darüber hinaus die Anforderungen zu dokumentieren und systematisch zu managen [72]. Dabei gilt es, die Bedürfnisse und Wünsche der Stakeholder zu berücksichtigen, damit die letztendliche Umsetzung der Anforderungen den Vorstellungen der Stakeholder gerecht wird [72].

Eines der grundlegenden Probleme, welches bei der Kommunikation zwischen Stakeholdern und Entwicklern auftritt, ist, dass Stakeholder und Entwickler meistens aus verschiedenen Domänen stammen. Dadurch besteht zwischen ihnen eine natürliche Sprachbarriere, wodurch bei den beteiligten Personen ein fehlendes Verständnis für die jeweils andere Domäne hervorgerufen wird [73]. Konkret haben Entwickler in der Regel nur wenig bis gar kein Wissen über die spezifische Domäne der Stakeholder und die Stakeholder besitzen meist keinerlei Erfahrung bezüglich eines Softwareentwicklungsprozesses. Fischer [74] bezeichnet dieses fehlende Verständnis für die jeweils andere Partei als „Symmetry of Ignorance“, die durch den verschiedenen Sprachgebrauch der jeweiligen Domäne erzeugt wird. Er beschreibt, dass Kommunikation zwischen Spezialisten mit spezifischem Wissen nötig ist, um in Zukunft komplexe Probleme lösen zu können, da diese Probleme mehr Wissen benötigen, als eine einzelne Person besitzen kann [74]. In diesem Zusammenhang sieht Fischer [74] diese „Symmetry of Ignorance“ nicht als Hindernis an, sondern stattdessen als Möglichkeit für Kreativität [74]. Durch die unterschiedlichen Sichtweisen verschiedener Personen können neue alternative Lösungsmöglichkeiten gefunden und unterschwellige Probleme identifiziert werden [74]. Die „Symmetry of Ignorance“ darf jedoch nicht ignoriert werden, da es sonst zu Missverständnissen zwischen den Stakeholdern und Entwicklern kommt, wodurch verschiedene mentale Modelle des Softwareprojekts bei den beteiligten Personen erzeugt werden [73]. Als möglichen Lösungsansatz für dieses Problem schlägt Rupp [73] daher vor, den auditiven und visuellen Kanal für die Informationsverarbeitung zu kombinieren. Eine Möglichkeit für die Kombination der beiden Informationskanäle stellt dabei die Nutzung von Videos dar.

Creighton et al. [76] präsentieren mit ihrem Ansatz des „Software Cinemas“ eine mögliche Umsetzung der Nutzung von Videos im Requirements Engineering. Sie integrieren dabei die Produktion von Videos zu verschiedenen Zeitpunkten in den Softwareentwicklungsprozess, um die zu entwickelnde Software aus der Sichtweise der Endnutzer darzustellen [76]. Die Videos sollen dabei unter anderem eine unterstützende Wirkung für die Klärung von Anforderungen besitzen [76]. Diverse andere Ansätze berichten ähnliche Vorgehensweisen, bei denen Videos genutzt wurden, um visionäre Systeme oder Teile eines Systems darzustellen, wodurch wiederum Aspekte des Requirements Engineerings unterstützt werden

[56,75,77]. Diese Ansätze besitzen darüber hinaus die Gemeinsamkeit, dass die Videos spezifisch für die an dem jeweiligen Projekt beteiligten Personen produziert werden, also für die Stakeholder und Entwickler. Da diese Videos die Visionen bezüglich des jeweiligen Projekts oder Aspekten des Projekts darstellen, wurde in verschiedenen Publikationen der Begriff „Vision Video“ verwendet, um diese Art von Videos zusammenzufassen [56,58,75]. Diese Vision Videos stellen dabei eine Grundlage dar, auf deren Basis Stakeholder und Entwickler über die Vision eines Systems diskutieren können. Diese Diskussion ermöglicht es, mögliche Unterschiede zwischen den Vorstellungen verschiedener Personen zu identifizieren und Missverständnisse zu klären, wodurch ein besseres Verständnis bezüglich des Projekts erreicht werden kann. Diese Wirkung von Vision Videos wurde auch innerhalb der Arbeit von Karras et al. [62] identifiziert, in der durch eine Studie gezeigt wird, dass durch die Unterstützung einer textuellen Spezifikation mit Vision Videos Personen schneller ein besseres Verständnis für die Spezifikation entwickeln.

In der vorhandenen Literatur wird jedoch noch nicht ausreichend gezeigt, dass Vision Videos neben dem Verständnis einzelner Personen bezüglich des zukünftigen Systems auch das gemeinsame Verständnis zwischen Personen fördern. Da wie einleitend beschrieben die Erzeugung von einem gemeinsamen Verständnis zwischen Personen essentiell für den erfolgreichen Abschluss eines Softwareprojekts ist, stellt sich die Frage, wie ein solches durch Vision Videos erzeugtes gemeinsames Verständnis gemessen und beurteilt werden kann. Um diese Frage zu beantworten empfiehlt es sich, nach entsprechenden existierenden Methoden, Heuristiken und Metriken für die Beurteilung von gemeinsamen Verständnis in der Literatur zu suchen. Aus diesem Grund wird in dieser Arbeit eine systematische Literatursuche durchgeführt, die passende Methoden, Heuristiken und Metriken identifizieren soll, mit denen eine Beurteilung von gemeinsamen Verständnis ermöglicht wird. Diese Methoden müssen dabei aus ihrem ursprünglichen Kontext in den Kontext von Vision Videos übertragen werden, um sie im Sinne der Beurteilung von durch Vision Videos erzeugtem gemeinsamen Verständnis verwenden zu können. Im folgenden Unterkapitel werden die Ziele, die diese Arbeit verfolgt, genauer erläutert.

1.2 Ziel der Arbeit

Das übergeordnete Ziel, das mit dieser Arbeit verfolgt wird, ist die Identifikation von Methoden, Heuristiken und Metriken aus der Literatur, die für die Beurteilung und die Messung von gemeinsamen Verständnis verwendet werden können, welches durch Vision Videos erzeugt wurde. Aus diesen Methoden, Heuristiken und Metriken sollen dann Ansatzpunkte für eine Methode entwickelt werden, mit der das gemeinsame Verständnis beurteilt werden kann. Dieses übergeordnete Ziel lässt sich in einige Teilziele aufteilen, anhand derer der Aufbau dieser Arbeit besser nachvollziehbar wird.

Das erste Teilziel, was sich aus dem übergeordneten Ziel ergibt, ist die Identifikation von Methoden, Heuristiken und Metriken für die Beurteilung und das Messen von gemeinsamen Verständnis. Der Kontext, aus dem diese Ansätze stammen, soll dabei absichtlich nicht auf einen Zusammenhang mit Vision Videos eingeschränkt werden, damit potentiell eine größere Menge von Methoden, Heuristiken und Metriken identifiziert werden kann. Dieses erste Teilziel wird umgesetzt, indem eine systematische Literatursuche sowohl geplant als auch durchgeführt wird. Diese wird aus verschiedenen Suchmethoden zusammengesetzt, um eine möglichst diverse Menge an Methoden, Heuristiken und Metriken zu generieren.

Auf dem ersten Teilziel aufbauend ist das zweite Teilziel dieser Arbeit, die durch die systematische Literatursuche gefundenen Methoden, Heuristiken und Metriken in den

Kontext von Vision Videos zu übertragen. Hierbei gilt es festzustellen, inwiefern eine Übertragung in den Kontext von Vision Videos überhaupt möglich ist. Falls diese Übertragung möglich ist, gilt es darüber hinaus zu prüfen, inwiefern die gefundenen Methoden, Heuristiken und Metriken für den neuen Kontext angepasst werden müssen.

Auf der Basis der Ergebnisse des zweiten Ziels beinhaltet das letzte Teilziel dieser Arbeit die Definition eines Ansatzpunktes einer konkreten Methode zur Beurteilung beziehungsweise Messung von gemeinsamen Verständnis, das durch Vision Videos erzeugt wurde. Dieser Ansatzpunkt wird in dieser Arbeit in Form eines Experimentdesigns beschrieben, das auf Basis von drei ausgewählten Methoden zur Beurteilung von gemeinsamen Verständnis ein Vorgehen beschreibt, mit dem das gemeinsame Verständnis zwischen Probanden bestimmt werden kann. Hierbei ist anzumerken, dass es aufgrund des zeitlichen Rahmens dieser Arbeit vorgesehen war, nur das Experimentdesign als Ansatzpunkt einer Methode zu definieren und nicht eine mögliche Umsetzung des Experimentdesigns als Experiment durchzuführen. Das Experimentdesign ist daher eher als möglicher Ausgangspunkt für weitere Arbeiten anzusehen.

1.3 Struktur der Arbeit

Diese Arbeit besteht aus sechs aufeinander aufbauenden Hauptkapiteln. Nach dieser Einleitung, in der die Motivation für die Arbeit erklärt und die Ziele der Arbeit definiert wurden, folgt in Kapitel 2 eine Menge von Erläuterungen und Definitionen zu grundlegenden Konzepten, die Kernbestandteile dieser Arbeit darstellen und in den darauffolgenden Kapiteln verwendet werden. Dies beinhaltet darüber hinaus ein Unterkapitel, in dem verwandte Arbeiten präsentiert und von dieser Arbeit abgegrenzt werden. Das darauf folgende dritte Kapitel befasst sich mit der Definition der systematischen Literatursuche. Die systematische Literatursuche besteht dabei aus drei verschiedenen, aufeinander aufbauenden Suchmethoden, deren Definition, Durchführung und Ergebnisse jeweils beschrieben werden. Danach werden in Kapitel 4 die Ergebnisse der gesamten systematischen Literatursuche analysiert. Hierbei wird sowohl der Prozess der systematischen Literatursuche genauer betrachtet, als auch inhaltlich die gefundenen Methoden, Heuristiken und Metriken zur Beurteilung und Messung von gemeinsamen Verständnis erläutert. Zuletzt wird in diesem Kapitel untersucht, inwiefern sich die gefundenen Methoden, Heuristiken und Metriken auf den Kontext von Vision Videos übertragen lassen. Darauf folgend wird im fünften Kapitel dieser Arbeit auf Basis der Ergebnisse der vorherigen Kapitel ein Experimentdesign erarbeitet, mit dem das gemeinsame Verständnis von Probanden beurteilt beziehungsweise gemessen werden kann. Hierbei baut das Experimentdesign auf drei verschiedenen Hypothesen auf, die sich auf die Erzeugung von gemeinsamen Verständnis mittels Vision Videos beziehen. Im sechsten und damit letzten Kapitel werden die Ergebnisse dieser Arbeit in einem Fazit zusammengefasst und ein Ausblick auf die Nutzungsmöglichkeiten dieser Ergebnisse in zukünftigen Arbeiten gegeben.

2 Grundlagen

2.1 Mentale Modelle

Um den Begriff des „shared understandings“ besser zu verstehen ist es sinnvoll, sich zunächst mit dem Konzept der mentalen Modelle auseinanderzusetzen. Der Begriff „mentales Modell“ wurde im Laufe der Zeit in diversen Kontexten verwendet. Den Grundstein für das Konzept legte Craik [64], der Modelle als physische oder chemische Systeme bezeichnet, die eine ähnliche Struktur besitzen wie der Prozess, den sie imitieren.

Johnson-Laird [65] war der erste, der den Begriff „mentales Modell“ definierte. Nach seiner Definition repräsentiert ein mentales Modell eine Sachlage und die Struktur des Modells ist entweder eine direkte Repräsentation oder eine Analogie zu dieser Sachlage [65]. Durch diese Struktur werden daher die relevanten Aspekte der Sachlage in der realen Welt wiedergespiegelt [65]. In einer anderen Publikation erklärt Johnson-Laird [66] des Weiteren, dass Menschen die Welt verstehen, indem sie ein entsprechendes mentales Modell erzeugen. Er bezeichnet die Modelle als unvollständig und daher als eine simplere Imitation der realen Situationen, die sie repräsentieren [66]. Für ein mentales Modell muss eine Person daher nicht wissen, wie etwas technisch funktioniert, da das mentale Modell nur das Verhalten imitiert [66]. Ausgehend von dieser ursprünglichen Definition gab es diverse Ansätze, die dem Begriff des mentalen Modells verschiedene Eigenschaften und Verhaltensweisen zugeordnet haben.

Norman [67] beschreibt einige Eigenschaften von mentalen Modellen. Laut Norman [67] sind mentale Modelle grundsätzlich unvollständig und instabil. Die Instabilität bezieht sich darauf, dass die Personen Details des Systems vergessen können, besonders wenn diese Details oder Systeme über einen längeren Zeitraum nicht genutzt wurden [67]. Darüber hinaus besitzen mentale Modelle keine festen Grenzen, wodurch die Modelle verschiedener Systeme miteinander vermischt oder verwechselt werden können [67]. Des Weiteren sind mentale Modelle laut Norman [67] im Allgemeinen nicht wissenschaftlich, da nicht ausgeschlossen werden kann, dass Personen in ihren mentalen Modellen Verhaltensmuster beibehalten, die durch unbegründete Einflüsse wie Aberglauben geprägt werden.

Caroll und Olson [68] beschreiben in ihrem Ansatz mentale Modelle als ausgefeilte Struktur, die das Verständnis des Nutzers widerspiegelt, bezüglich dessen was das System enthält, wie es funktioniert und warum es in dieser Weise funktioniert. Das mentale Modell kann dementsprechend als Wissen über das System interpretiert werden, mit dem sich mental die Aktionen des Systems simulieren lassen, bevor diese ausgeführt werden.

Chermack [69] liefert wiederum einen anderen Ansatz bezüglich mentaler Modelle. Er beschreibt, dass Menschen auf Basis ihrer Vermutungen, Überzeugungen, Erfahrungen und Voreingenommenheiten bezüglich der Welt ununterbrochen mentale Modelle der Realität erzeugen [69]. Diese Erzeugung von mentalen Modellen kann dabei auch erfolgen, ohne dass sich die betroffene Person dessen bewusst ist [69]. Diese mentalen Modelle werden dann verwendet, um auf ihrer Basis Entscheidungen zu treffen [69]. Die konkrete Struktur der Modelle selbst wird dabei von Chermack [69] als „schwer zu verstehen“ bezeichnet.

Diese und weitere Definitionen werden in der Publikation von Hammarbäck [7] ausführlich analysiert, um den Begriff des „mental Modells“ in einer allgemeinen Form zu definieren. Dabei stellt Hammerbäck [7] fest, dass sich die einzelnen Definitionen in diversen Punkten widersprechen, da sie in verschiedenen Kontexten definiert wurden. Ein Beispiel hierfür ist,

dass Carroll und Olsen [68] beschreiben, dass mentale Modelle aus verschiedenen kleineren mentalen Modellen mit unterschiedlichem Detailgrad bestehen, während die meisten anderen Definitionen nur von einer zusammenhängenden Struktur sprechen. Hammerbäck [7] stimmt mit Rock [70] überein, dass die grundlegende Übereinstimmung aller Definitionen ist, dass ein mentales Modell eine interne Struktur ist, welche das Handeln von Personen beeinflusst. Jedoch identifiziert Hammerbäck [7] auch eine Menge anderer Funktionen und Eigenschaften, die mentalen Modellen zugeschrieben werden können. Eine dieser Funktionen ist, dass mentale Modelle dazu verwendet werden, Voraussagen über eine Situation zu treffen [7]. In diesem Zusammenhang werden mentale Modelle dazu genutzt, Schlussfolgerungen bezüglich einer Situation zu ziehen und anhand dieser die Situation zu erklären [7]. Darüber hinaus kann durch diese Voraussagen und Schlussfolgerungen mithilfe von mentalen Modellen eine Situation mental simuliert werden, ohne dass eine physische Ausführung notwendig ist [7]. Neben den geschilderten Funktionen identifiziert Hammerbäck [7] auch einige Eigenschaften von mentalen Modellen, die zwischen den meisten Definitionen konsistent bleiben. Eine dieser Eigenschaften ist, dass das mentale Modell auf struktureller Ebene ähnlich zu dem realen System ist, welches repräsentiert werden soll [7]. Eine weitere Eigenschaft, die in diesem Zusammenhang auftritt, ist, dass mentale Modelle aufgrund der biologischen Begrenzungen der mentalen Kapazität von Menschen eingeschränkt sind [7]. Dies hat zur Folge, dass die realen Systeme durch Abstraktion vereinfacht werden müssen, um eine Repräsentation durch ein mentales Modell zu ermöglichen. Neben dem Kontext, indem sie erzeugt werden, sind mentale Modelle abhängig vom vorherigen Wissen der Person [7]. Dies ist damit zu begründen, dass mentale Modelle eine Wissensstruktur darstellen und somit von dem bereits vorhandenen Wissen einer Person beeinflusst werden können [7]. Die letzte Eigenschaft, die mentalen Modellen eindeutig zugeordnet werden kann, ist, dass sie dynamisch sind [7]. Mentale Modelle können abhängig von Änderungen der Situation, für die sie erzeugt wurden, entweder angepasst oder verworfen und durch ein neues, anderes mentales Modell ersetzt werden [7].

Betrachtet man sämtliche Funktionen und Eigenschaften, die von Hammerbäck [7] als allgemeingültig identifiziert wurden, dann lässt sich für diese Arbeit die folgende Definition für mentale Modelle bestimmen:

Mentale Modelle sind eine dynamische, interne Repräsentation eines realen Systems oder einer realen Situation, die unter Einfluss von vorherigem Wissen erzeugt wird und das Handeln einer Person beeinflusst. Sie besitzen dabei eine ähnliche Struktur wie das reale Gegenstück, das sie repräsentieren und sind in ihrer Komplexität eingeschränkt. Anhand von mentalen Modellen lassen sich Rückschlüsse über reale Systeme oder Situationen ziehen. Darüber hinaus können durch mentale Modelle reale Systeme oder Situationen erklärt, Voraussagen über diese getroffen und das Verhalten mental simuliert werden.

2.2 Shared Understanding

Obwohl der Begriff des „shared understandings“ Verwendung in diversen Publikationen im Bereich Requirements Engineering findet, wird nur in den seltensten Fällen klar definiert, was unter dem Begriff „shared understanding“ zu verstehen ist. Falls definiert wird, was „shared understanding“ bedeutet, ist die Definition oftmals sehr rudimentär und stark vom Kontext der Verwendung abhängig. Ein Beispiel hierfür ist die Publikation von Humayun und Gang [17], in der definiert wird, dass „shared understanding“ auftritt, wenn verschiedene Mitglieder eines Projekts das gleiche Verständnis bezüglich jeder Anforderung haben. Des Weiteren wird beschrieben, dass „shared understanding“ wichtig für die Lösung von Konflikten und Klärung von Mehrdeutigkeiten ist [17]. Auf eine Erklärung, was unter einem gleichen Verständnis zu

verstehen ist, wird jedoch verzichtet. Da das Konzept von „shared understanding“ einen der zentralen Punkte dieser Arbeit darstellt, ist an dieser Stelle eine klare Definition des Begriffs für diese Arbeit nötig, um mögliche Unklarheiten und Missverständnisse zu vermeiden. Aus diesem Grund werden im Folgenden einige der wenigen ausgereiften Definitionen für den Begriff „shared understanding“ betrachtet.

Eine der frühesten Definitionen des Begriffs „shared understanding“ liefert 1995 Easterbrook [53]. Easterbrook [53] definiert, dass zwei oder mehr Personen ein „shared understanding“ bezüglich einer Situation besitzen, wenn sie die gleichen Erwartungen bezüglich der Situation haben. Als Erwartungen werden dabei sowohl die Erklärungen der Personen bezüglich der Situation als auch die Voraussagen, wie sich die Situation entwickeln könnte, definiert [53]. Easterbrook [53] nimmt an, dass diese Erwartungen auf einer Art von mentalem Modell der Situation beruhen, jedoch wird auf eine genaue Ergründung dieses mentalen Modells verzichtet. Ein zentraler Punkt dieser Definition ist der Bezug des „shared understandings“ auf eine bestimmte Situation. Unter einer Situation versteht Easterbrook [53] eine Folge von Interaktionen und die Umgebung, in der die Interaktionen stattfinden. Easterbrook [53] beschreibt die Rolle der Situation als entscheidend, da nicht gewährleistet werden kann, dass das „shared understanding“ der Personen außerhalb der spezifischen Situation erhalten bleibt. In einer anderen Situation kann nicht garantiert werden, dass die Personen die gleichen Erwartungen erzeugen, wodurch auch das „shared understanding“ der Personen nicht mehr garantiert ist [53]. Bleibt ein „shared understanding“ zwischen verschiedenen Situationen erhalten, dann kann es laut Easterbrook [53] als robust bezeichnet werden. Ein robustes „shared understanding“ lässt darauf schließen, dass die Personen ein identisches mentales Modell besitzen [53]. Jedoch ist für Easterbrook [53] nicht entscheidend, wie die Personen ihre Erwartungen ableiten, wodurch die präzise Ergründung des mentalen Modells irrelevant wird [53]. Darüber hinaus distanziert Easterbrook [53] den Begriff „shared understanding“ klar von Allgemeinwissen, da Allgemeinwissen darauf beruht, dass die betroffenen Personen wissen, dass das Wissen allgemein bekannt ist, was für ein „shared understanding“ nicht der Fall sein muss.

McKay et al. [63] setzen sich in ihrer Publikation ebenfalls mit der Definition des Begriffs „shared understanding“ auseinander und identifizieren drei verschiedene Ansätze, mit denen „shared understanding“ definiert werden könnte. Der erste und simpelste der Ansätze ist, dass „shared understanding“ eine Einigung zwischen den Mitgliedern einer Gruppe entspricht, die zu einer Aktion führt. Der zweite und dritte Ansatz sind insofern ähnlich, dass sie „shared understanding“ als gesteigertes zwischenmenschliches Verständnis interpretieren, jedoch werden verschiedene Folgen dieses Verständnisses angenommen. Im zweiten Ansatz wird gefolgert, dass das zwischenmenschliche Verständnis zu Einigkeit und verstärkter Handlungsbereitschaft führt [63]. Darüber hinaus werden Aktionen durch die betroffenen Personen wahrscheinlicher und einfacher umsetzbar [63]. Im Gegensatz dazu nimmt der dritte Ansatz an, dass durch das „shared understanding“ keine Implikation bezüglich einer Einigung oder Handlungsbereitschaft entsteht [63]. McKay et al. [63] legen ihren Fokus auf den zweiten Ansatz und beschreiben, dass eine genauere Differenzierung des Grades des „shared understandings“ nötig ist. Hierbei wird festgestellt, dass abhängig davon, wie abstrakt eine Situation betrachtet wird, ein „shared understanding“ zwischen Personen angenommen oder ausgeschlossen werden kann [63]. McKay et al. [63] schlussfolgern daher, dass das „shared understanding“ zwischen Personen abhängig davon ist, auf welcher Abstraktionsebene eine Situation betrachtet wird.

Aranda [54] baut auf der Definition von Easterbrook [53] auf und führt diese weiter aus. Das von Easterbrook [53] beschriebene „shared understanding“ kann nur durch Kommunikation

innerhalb eines Projektteams erzeugt werden, da die Feinheiten des Verständnisses einzelner Personen nicht in technischen Dokumenten beschrieben werden können [54]. „Shared understanding“ kann nicht durch technische Dokumente beschrieben werden, da es zu einem gewissen Grad auf informellen und anekdotischen Erfahrungen beruht, die in technischen Dokumenten nicht berücksichtigt werden [54]. Aranda [54] führt darüber hinaus den Aspekt der mentalen Modelle weiter aus und beschreibt, dass die mentalen Modelle individueller Personen oftmals unvollständig, eventuell inkonsistent oder sehr oberflächlich sind. Dies hat zur Folge, dass Personen anhand ihrer mentalen Modelle Erwartungen bezüglich einer Situation haben, welche die Situation sowohl erklären als auch Vorhersagen bezüglich der Situation erlauben, jedoch werden diese Erwartungen oft nicht bestätigt [54]. Auf Basis seiner Analyse der Definition von Easterbrook [53] formuliert Aranda [54] die folgende, präzisere Ausführung der Definition:

„Two or more participants have a shared understanding of some situation, if the elements of their mental models salient to that situation are functionally equivalent. By functional equivalence, we mean that the models will provide the same explanations and the same predictions of a situation.“ [54]

Die grundlegende Aussage dieser Definition deckt sich dabei mit der von Easterbrook [53], jedoch präzisiert Aranda [54] die Definition durch die Einführung des Begriffs der funktionalen Äquivalenz.

Die Ansätze von McKay [63] betrachten den Begriff „shared understanding“ sehr stark im Kontext von Aktionen, die auf Basis des „shared understandings“ durchgeführt werden und berücksichtigt nicht den Aspekt der mentalen Modelle. Die Definitionen von Easterbrook [53] und Aranda [54] definieren den Begriff „shared understanding“ in einem allgemeineren Kontext als die Ansätze von McKay [63] und berücksichtigen dabei das Konzept der mentalen Modelle. Darüber hinaus werden im Kontext dieser Arbeit nicht die Aktionen betrachtet, die durch das „shared understanding“ hervorgerufen werden. Stattdessen steht die Beurteilung von „shared understanding“ selbst im Fokus dieser Arbeit, was sich mehr mit den Definitionen von Easterbrook [53] und Aranda [54] deckt. Aus diesen Gründen wird in dieser Arbeit die Definition von Aranda [54] für den Begriff „shared understanding“ verwendet, da sie eine präzisere Variante der Definition von Easterbrook [53] darstellt.

Im weiteren Verlauf dieser Arbeit wird der Begriff „shared understanding“ verwendet, wenn sich auf die Definition von Aranda [54] bezogen wird, um mögliche Verwechslungen mit dem im Allgemeinen nicht eindeutig definiertem Begriff des gemeinsamen Verständnisses zu vermeiden.

2.3 Videonutzung im Requirements Engineering

Die Nutzung von Videos ist eine Methode, die seit einiger Zeit im Zusammenhang mit Requirements Engineering untersucht wird. Da einer der Kernaspekte des Kontextes dieser Arbeit die Verwendung von Vision Videos ist, soll an dieser Stelle zunächst die Grundidee der Nutzung von Videos im Requirements Engineering betrachtet werden. Hierfür werden im Folgenden einige Forschungen betrachtet, die verschiedene grundlegende Charakteristiken dieser Methode identifizieren.

Einen der frühesten Ansätze für die Nutzung von Videos im Requirements Engineering liefern Creighton et al. [76] mit ihrer sogenannten „Software Cinema“ Technik. Als Ausgangspunkt für ihre Technik beschreiben Creighton et al. [76], dass eines der

gravierensten Probleme des Softwareentwicklungsprozesses die Modelle sind, welche die zu entwickelnde Software repräsentieren. Diese Modelle werden dabei meist in einer Sprache für Experten definiert, wodurch domänen-fremde Stakeholder keinen ausreichenden Überblick über das gesamte System erhalten können [76]. In diesem Zusammenhang identifizieren Creighton et al. [76] fünf zentrale Herausforderungen, die es im Umgang mit Stakeholdern zu bewältigen gilt: das Definieren der Erwartungen der Stakeholder, die Änderung von bekannten mentalen Modellen, die Kommunikation unterschiedlicher Sichtweisen der Realität, die verschiedenen Bedeutungen der Sprache und die Kommunikation von Systemvisionen. Der Ansatz von Creighton et al. [76] sieht die Nutzung von digitalen Videos vor, um die Bewältigung dieser Herausforderungen im Requirements Engineering zu unterstützen. Die präsentierte Software Cinema Technik wird dabei zur Produktion von Videos verwendet, die als eine semi-formale Repräsentation der traditionellen Softwaremodelle dienen sollen [76]. Da Videos klassischer Weise dazu verwendet werden, Gefühle an die Zuschauer zu vermitteln, wollen Creighton et al. [76] diese Eigenschaft nutzen, um die Sprachbarriere zwischen Endnutzer und Entwickler zu überbrücken. In diesem Zusammenhang sollen dabei zwei Ziele erreicht werden. Zum einen sollen die Videos ein Modell der Realität liefern, welches von allen Stakeholdern gleichermaßen verstanden werden kann. Zum anderen sollen die Videos für Entwickler eine Referenz dafür darstellen, was mit dem zu entwickelnden System erreicht werden soll [76]. Software Cinema ist dabei eine Technik die beschreibt, wie solche beschreibenden Videos erstellt und erweitert werden können, um mit ihnen eine Analyse der Anforderungen der Endnutzer zu ermöglichen [76].

Entsprechend dem Konzept eines V-Modells stellen Creighton et al. [76] den Software-Lebenszyklus dar um zu zeigen, dass an diversen Punkten dieses Prozesses Videos unterstützend verwendet werden können, dargestellt in Abbildung 1.

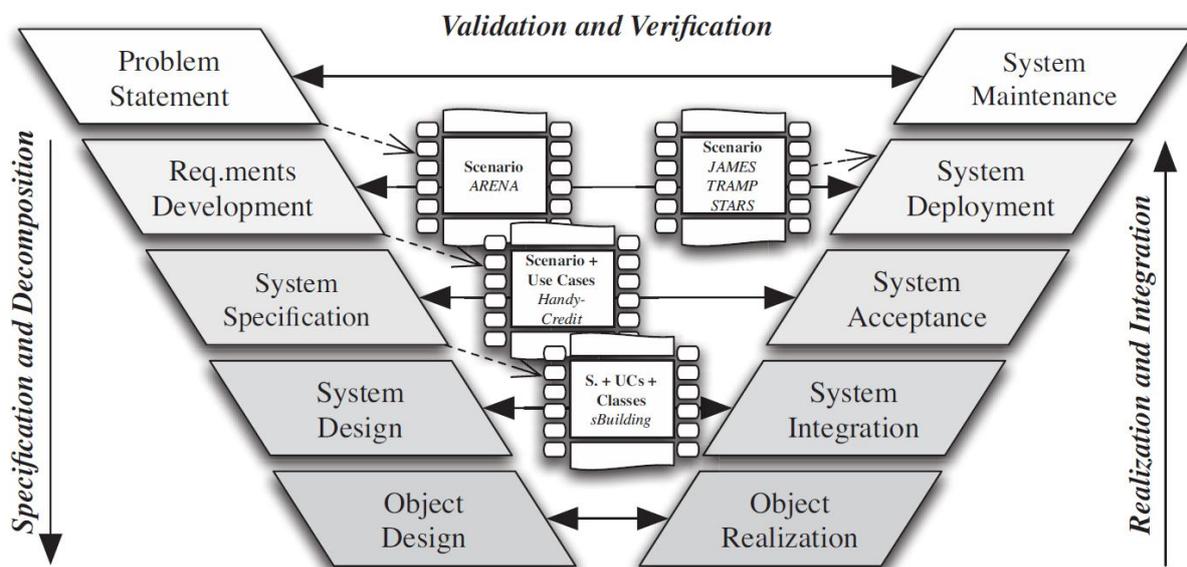


Abbildung 1: Videonutzung im Software-Lebenszyklus nach Creighton et al. [76]

Die Videos können dabei sowohl zu Beginn eines Projekts verwendet werden, um die Definition der Problemstellung und die Identifizierung der Anforderung zu unterstützen, als auch am Ende eines Projektes, um zu überprüfen, ob die Ergebnisse des Projekts mit der ursprünglichen Vision übereinstimmen [76]. Ein Beispiel dafür sind Videos von Nutzungsszenarien, mit denen sowohl Anforderungen erkannt, als auch die letztendliche Funktionsweise des Systems geprüft werden können [76]. Final stellen Creighton et al. [76]

fest, dass Videos bei der Bewältigung von jeder der fünf initial identifizierten Herausforderungen behilflich sein können.

Auch andere Forschungen kommen zu ähnlichen Erkenntnissen in Bezug auf die Nutzung von Videos im Requirements Engineering. Broll et al. [75] schlagen die Nutzung von Videos in frühen Phasen eines Projekts vor, um Anforderungen an das Projekt innerhalb von Fokusgruppen zu identifizieren. Die Videos sollen dabei die zentralen Ideen des Projekts repräsentieren, durch welche die Ziele des Projekts kommuniziert und konkrete Nutzungsszenarien visualisiert werden sollen [75]. Darüber hinaus sollen die Videos möglichst kurz sein, um die Aufmerksamkeit der Stakeholder zu garantieren [75]. Broll et al. [75] konnten durch das Feedback eines Experiments bestimmen, dass die Nutzung solcher Videos sehr effektiv ist, um die grundsätzlich eher abstrakten Produktideen für die Stakeholder verständlich darzustellen.

In einem anderen Ansatz sehen Pham et al. [56] im Kontext ihres interaktiven Multimedia-Storyboards die Nutzung von sogenannten Vision Videos vor. Diese Vision Videos sollen dabei die Vision eines Projekts repräsentieren, indem sie diverse Aktionen darstellen, die mit dem Endsystem durchführbar sein sollen [56]. Diese Videos werden dabei in Zusammenarbeit mit dem Kunden erstellt, um ein direktes Feedback zu ermöglichen [56]. Das Ziel der Erzeugung von Vision Videos ist dabei, dass die Anforderungen in einer Form dargestellt werden, die von allen Stakeholdern verstanden werden kann, um eine Basis für Diskussionen zu schaffen [56]. Darüber hinaus wird den Stakeholdern ermöglicht, neue Anforderungen zu formulieren und die Vision des Projekts so präzise wie möglich darzustellen [56]. Final stellen Pham et al. [56] fest, dass durch die von ihnen präsentierte Methode Unterschiede im Verständnis verschiedener Stakeholder identifiziert werden können, die zuvor versteckte Anforderungen aufdecken können. Die Methode fokussiert sich dabei im Allgemeinen auf die Verständlichkeit und die Ausdruckskraft der Kommunikation zwischen Requirements Engineer und Stakeholder [56].

Aus den hier präsentierten Arbeiten lässt sich letztendlich schließen, dass die Nutzung von Videos eine Methode darstellt, die für diverse Aspekte des Requirements Engineerings eine unterstützende Wirkung besitzen kann. Eine Gemeinsamkeit der Videos ist dabei, dass sie die Kommunikation zwischen Stakeholdern und Entwicklern erleichtern und im Zuge dessen ein gewisses Maß an allseitigem Verständnis für das Projekt schaffen.

2.4 Vision Video

Im vorherigen Kapitel wurden diverse Arbeiten präsentiert, die sich mit einem ähnlichen Konzept von Videos auseinandersetzen, die die Vision eines zukünftigen Systems darstellen. Jedoch existieren in der Literatur eine Vielzahl verschiedener Terminologien für diese Art von Videos. Beispiele hierfür sind „Videos von visionären Szenarien“ [76], „Videomaterial zur Visualisierung von Projektvision“ [75] oder „Vision Video“ [56]. Diese Inkonsistenz beruht darauf, dass die Terminologie abhängig vom Kontext der jeweiligen Arbeit gewählt wurde und keine allgemeingültige Definition für diese Art von Videos vorhanden ist. Da Vision Videos einen der zentralen Aspekte dieser Arbeit bilden, gilt es an dieser Stelle die Bedeutung dieses Begriffes für den Kontext dieser Arbeit zu definieren. Hierfür wird die Verwendung entsprechender Konzepte von Vision Videos in der Literatur miteinander verglichen, um sie letztendlich in einer Definition zusammenzufassen. Diese Definition soll dabei keinen Anspruch auf Allgemeingültigkeit besitzen, sondern in erster Linie den Begriff „Vision Video“ für die Verwendung in dieser Arbeit klären.

Creighton et al. [76] verwenden ihre Software Cinema Methode, um Videos zu produzieren, die visionäre Szenarien eines zukünftigen Systems darstellen. In diesen Videos sollen dabei sowohl das System selbst, als auch dessen Verhalten über einen gewissen Zeitraum hinweg modelliert werden. Das dargestellte Szenario soll dabei eine Situation widerspiegeln, die aus der Realität bekannt ist. Die auf diese Art produzierten Videos sollen als Basis von Diskussionen zwischen Stakeholdern und Entwicklern genutzt werden, die sich sowohl auf die Systemeigenschaften als auch auf die möglichen Systemabläufe beziehen.

Broll et al. [75] beschreiben die Nutzung von Videos in den frühen Phasen eines Projekts zur Identifikation von Anforderungen. Die Videos sollen dabei so kurz wie möglich sein, da sie im Kontext von Fokusgruppen verwendet werden und Broll et al. [75] die Zeit der beteiligten Personen als sehr begrenzte Ressource einstufen. Ein positiver Effekt der kurzen Dauer der Videos ist, dass die Aufmerksamkeit der beteiligten Personen erhalten bleibt. Die Videos sollen für diesen Kontext das Ziel des Projekts durch die Darstellung eines oder mehrerer konkreter Nutzungsszenarien präsentieren. Darüber hinaus soll eine mögliche technische Umsetzung des Systems dargestellt werden, die jedoch sehr oberflächlich gehalten werden muss, damit kein zu präzises Design des Systems vorgeschrieben wird. Die Videos sollen dabei in einer entsprechenden Umgebung in der echten Welt unter Verwendung von Mockups des Systems produziert werden.

Pham et al. [56] verwenden in ihrer Arbeit den konkreten Begriff „Vision Video“ und beziehen sich dabei auf die Darstellung einer Vision eines zu entwickelnden Systems, einer innovativen Idee oder einer Verbesserung eines bereits vorhandenen Systems. In ihrem Ansatz sehen sie zunächst vor, Teilaktionen im Zusammenhang mit dem System darzustellen, die dann zu einem gesamtheitlichen Vision Video zusammengesetzt werden, um dadurch den Prozess des jeweiligen Systems darzustellen [56]. Aufgrund der Art, in der die Vision Videos konstruiert werden, ist die Dauer eines einzelnen Vision Videos eher kurz und beinhaltet meistens nur eine zusammenhängende Aktion. Pham et al. [56] erläutern, dass es bei der Produktion von längeren Vision Videos zu dem Problem kommen kann, dass die einzelnen Aktionen nicht mehr voneinander abgegrenzt und dadurch dem Betrachter nicht eindeutig vermittelt werden können. Das grundlegende Ziel dieser Vision Videos ist es dabei, die an das System gestellten Anforderungen in einer aussagekräftigen Art darzustellen, um Feedback und mögliche Korrekturen der Aktionen durch die Stakeholder zu ermöglichen.

Außerhalb des Kontextes der Nutzung von Videos betrachten Fricker et al. [57] die „Best Practices“ im Requirements Engineering und definieren dabei den Begriff der Vision. Eine Vision besteht laut ihrer Definition dabei aus drei Bestandteilen: eine Definition des zu betrachtenden Problems, ein grundlegender Lösungsansatz für das definierte Problem und eine Beschreibung, inwiefern diese Lösung den momentanen Stand der Technik in Bezug auf die Zusammenarbeit von Stakeholdern und Entwicklern verbessert. Diese drei Aspekte lassen sich in den drei vorher präsentierten Ansätzen wiederfinden.

Karras et al. [58] befassen sich in ihrer Arbeit ebenfalls mit der Nutzung von Vision Videos, um die Vision eines Systems zu vermitteln und damit ein gemeinsames Verständnis zwischen allen am Projekt beteiligten Personen zu erreichen. Sie sehen dabei Visionen oder Teile von Visionen als sinnvollen Inhalt von Videos, da Visionen einen Überblick über das Projekt schaffen, mitsamt der dazugehörigen Ziele und dem Umfang des zukünftigen Systems [58]. Karras et al. [58] beziehen sich in diesem Zusammenhang auf eine vorherige Arbeit von Karras [59], welche die folgende grundlegende Definition für den Begriff „Vision Video“ liefert:

„Ein Vision Video ist ein Video, das eine Vision oder Teile einer Vision repräsentiert, um ein gemeinsames Verständnis zwischen allen Parteien zu erreichen, indem deren mentale Modelle des zukünftigen Systems offenbart, diskutiert und angeglichen werden.“ [59]

Diese Definition lässt sich dabei mit den bereits präsentierten Konzepten von Vision Videos vereinbaren und konkretisiert die Definition der Vision Videos unter Verwendung der Konzepte von mentalen Modellen und „shared understanding“, die in ihren jeweiligen Kapiteln 2.1 und 2.2 dieser Arbeit erläutert wurden. Aus diesem Grund wird diese Definition von Vision Videos im Kontext dieser Arbeit verwendet.

Die Definition lässt sich dabei jedoch um zwei Eigenschaften erweitern, die als Gemeinsamkeiten der verschiedenen Ansätze der Nutzung von Vision Videos identifiziert werden können. Eine dieser Eigenschaften ist, dass die Visionen der Vision Videos durch die Darstellung eines konkreten Nutzungsszenarios in einer realen Umgebung vermittelt werden. Die zweite der gemeinsamen Eigenschaften ist, dass sich die Länge der Vision Videos im Allgemeinen auf wenige Minuten beschränkt, um die Aufmerksamkeit des Betrachters zu erhalten und die Inhalte des Vision Videos auf das Wesentliche zu begrenzen.

2.5 Verwandte Arbeiten

Sowohl die Nutzung von Videos zur Erzeugung von „shared understanding“ als auch die Suche nach Methoden zur Beurteilung von „shared understanding“ mittels einer Literatursuche sind Thematiken, die bereits durch verschiedene Ansätze in der Literatur verfolgt wurden. Dieses Kapitel präsentiert daher einige der verwandten Arbeiten, die sich mit diesen Thematiken auseinandersetzen und dadurch Ähnlichkeiten zu dieser Arbeit aufweisen. In diesem Zusammenhang werden die grundlegenden Ideen und Konzepte der verwandten Arbeiten erläutert und von dieser Arbeit abgegrenzt, indem sowohl die Gemeinsamkeiten als auch die Unterschiede im Vergleich zu dieser Arbeit aufgezeigt werden.

Die hier präsentierten verwandten Arbeiten wurden durch die Nutzung verschiedener Quellen gefunden. Eine dieser Quellen stellt dabei eine Menge initialer Publikationen dar, die durch den Betreuer dieser Arbeit als Einführung in die Thematik bereitgestellt wurden. Eine weitere Quelle verwandter Arbeiten stellt eine allgemeine Suche der Thematik mit der Suchmaschine Google Scholar dar. Hierfür wurde nach diversen Kombinationen der Begriffe „evaluate“, „measure“, „shared understanding“, „vision video“ und „requirements engineering“ gesucht. Die letzte und wesentlich strukturiertere Suchmethode nach verwandten Publikationen stellt die systematische Literatursuche dar, die im Verlauf dieser Arbeit präsentiert wird. Im Folgenden werden einige der verwandten Arbeiten repräsentativ beschrieben.

Yusoff und Salim [60] betrachten in ihrer Arbeit das Konzept der „shared mental models“, die große Ähnlichkeit mit dem Konzept des „shared understandings“ aufweisen, das in dieser Arbeit betrachtet wird. Während der Evaluation der „shared mental models“ präsentieren sie dabei diverse Methoden, mit denen „shared mental models“ gemessen werden können und kategorisieren diese Methoden im Kontext ihrer Arbeit. Die Erklärung der Methoden durch Yusoff und Salim [60] bleibt dabei jedoch auf einem sehr oberflächlichen Level und kann daher nur als Auflistung betrachtet werden.

Mayr et al. [19] verfolgen einen ähnlichen Ansatz wie Yusoff und Salim [60] und identifizieren ebenfalls Methoden, mit denen mentale Modelle gemessen werden können. Im Gegensatz zu Yusoff und Salim [60] beschreiben Mayr et al. [19] die Funktionsweise der einzelnen Methoden genauer und geben Beispiele für die Nutzung der jeweiligen Methode.

In Bezug auf die Nutzung von Vision Videos im Kontext von „shared understanding“ fokussieren sich Schneider et al. [61] in ihrer Arbeit auf die Verfeinerung von Vision Videos. Hierfür führen sie ein Experiment durch, in dem unter anderem eine Metrik für die Messung von „shared understanding“ definiert wird. Die Metrik basiert dabei auf der Anzahl von Beiträgen in Form von neuen Ideen, Fragen, Anforderungen und Begründungen, die durch die Betrachtung eines Vision Videos beziehungsweise eines Texts hervorgerufen werden. Schneider et al. [61] kommen dabei zu dem Schluss, dass die Bereitstellung von Vision Videos zu neuen Beiträgen führt, was in diesem Kontext repräsentativ für eine Verbesserung des „shared understandings“ steht. Diese Metrik ist jedoch sehr auf das konkrete Experiment abgestimmt, in dem sie verwendet wird und daher schwer auf eine allgemeine Verwendung übertragbar.

Karras et al. [62] folgern in ihrer Arbeit, dass eine erfolgreiche Kommunikation der Anforderungen eine Voraussetzung für die Erzeugung von „shared understanding“ darstellt. In diesem Zusammenhang überprüfen sie, inwiefern Vision Videos die Kommunikation von Anforderungen durch die Spezifikation unterstützen können. Sie kommen auf Basis der Durchführung einer Studie zu dem Ergebnis, dass durch die Unterstützung mit Vision Videos Entwickler schneller ein besseres Verständnis bezüglich einer textuellen Spezifikation erhalten. Dies bedeutet, dass die Kommunikation der Anforderungen und dadurch das „shared understanding“ durch Vision Videos verbessert werden kann.

Glinz und Fricker [55] untersuchen in ihrer Arbeit die Rolle, den Wert und die Nutzung von „shared understanding“ im Software Engineering. Sie unterscheiden dabei zwischen explizitem „shared understanding“ in Form von Spezifikationen und implizitem „shared understanding“, was sich auf Fakten, Annahmen und Werte bezieht, die nicht spezifiziert wurden [55]. In diesem Zusammenhang zeigen sie diverse Methoden auf, mit denen „shared understanding“ erzeugt werden kann. Als Ergebnis ihrer Arbeit präsentieren Glinz und Fricker [55] eine Roadmap für „shared understanding“, in der sie unter anderem feststellen, dass implizites „shared understanding“ nicht ausreichend verstanden ist und noch in keiner systematischen und reflektierten Art Verwendung findet.

Creighton et al. [76] stellen ihre Software Cinema Technik vor, mit der Videos produziert werden können, die visionäre Szenarien eines Systems darstellen. Auf Basis dieser visuellen Darstellung von Anforderungen stellen sie einen allgemeinen Zusammenhang zum Softwareentwicklungsprozesses her. Unter anderem stellen sie dabei fest, dass die Videos ein Modell der Realität darstellen, durch das alle Stakeholder gleichermaßen ein Verständnis für das System entwickeln können.

Brill et al. [77] verfolgen einen ähnlichen Ansatz wie Creighton et al. [76], indem sie die Nutzung von ad-hoc Videos für die konkrete Repräsentation von Anforderungen vorschlagen. Sie führen in diesem Zusammenhang ein Experiment durch, um die Effektivität und Effizienz von günstig produzierten Videos in Bezug auf die Validierung von Anforderungen zu überprüfen und vergleichen die Videos dabei mit Use Cases. Sie kommen durch das Experiment zu dem Schluss, dass mithilfe der ad-hoc Videos mehr Anforderungen identifiziert werden konnten als mit Use Cases. Dadurch kann im Vergleich ein besseres Verständnis für das System erzeugt und Missverständnisse vermieden werden.

Broll et al. [75] nutzen ebenfalls Videos für die Anforderungsanalyse, indem sie durch die Videos die Projektvision präsentieren. Sie prüfen den Nutzen dieser Videos, indem sie diese als Diskussionsgrundlage für eine Fokusgruppe verwenden. Neben diversen Erkenntnissen

über die Produktion der Videos ging aus dem Feedback der Fokusgruppe hervor, dass die Videos als sehr effektiv empfunden wurden, um abstrakte Systeme für die Teilnehmer der Fokusgruppe verständlich zu machen. In diesem Zusammenhang wurden die Videos als gute Vorbereitung für Diskussionen wahrgenommen.

Pham et al. [56] präsentieren ein interaktives Multimedia-Storyboard, das auf Basis von Videos die Interaktion mit Stakeholdern verbessern soll. Die Videos werden dabei verwendet, um die Anforderungen des Systems zu visualisieren. Hierfür entwickeln sie das sogenannte VisionCatcher Tool, durch das die Videos mit verschiedenen anderen Medien kombiniert werden. Pham et al. [56] folgern aus ihren Ausführungen, dass Stakeholder ein besseres Verständnis für das System durch die Videos erhalten, da sie keine besondere Notation lernen müssen, um die Videos zu verstehen.

Die präsentierten Arbeiten stehen in engem Zusammenhang mit der Thematik dieser Arbeit. Während Yusoff und Salim [60] sowie Mayr et al. [19] verschiedene Methoden aufzeigen, mit denen das „shared understanding“ von Personen indirekt über deren mentale Modelle gemessen werden kann, wird durch die anderen Arbeiten deutlich, dass Videos ein effektives Mittel sind, um bei Stakeholdern ein Verständnis für ein visionäres System zu erzeugen. Insbesondere haben Schneider et al. [61] und Karras et al. [62] in ihren Ansätzen gezeigt, dass das „shared understanding“ durch die Nutzung von Vision Videos gesteigert werden kann. Eine Thematik, die jedoch von keiner der gefundenen Arbeiten aufgegriffen wurde, ist eine Kombination dieser beiden grundlegenden Ansätze, bei der allgemein verwendbare Methoden benutzt werden, um das „shared understanding“ zwischen Personen, was auf Basis von Videos erzeugt wurde, konkret zu messen und zu beurteilen. An diesem Punkt setzt diese Arbeit an, indem mit einer systematischen Literatursuche Methoden, Metriken und Heuristiken identifiziert werden, mit denen „shared understanding“ gemessen werden kann, das auf Basis der Nutzung von Vision Videos erzeugt wurde.

3 Systematische Literatursuche

Dieses Kapitel befasst sich mit dem Prozess der systematischen Literatursuche dieser Arbeit. In diesem Zusammenhang wird zunächst die allgemeine Struktur erläutert, die für die systematische Literatursuche gewählt wurde. Danach wird nacheinander auf die einzelnen, aufeinander aufbauenden Phasen der systematischen Literatursuche eingegangen, die aus einer manuellen Suche, Datenbanksuche und zuletzt Snowballing bestehen. Zu jeder dieser Phasen wird sowohl das Vorgehen beschrieben, als auch die Ergebnisse erläutert, die direkt aus der jeweiligen Phase gewonnen wurden. Als allgemeine Grundlage für die systematische Literatursuche wurde hierbei die Guideline von Kitchenham [1] verwendet.

3.1 Struktur der Suche

Damit das Ergebnis der systematischen Literatursuche möglichst aussagekräftig ist, gilt es zunächst, die passenden Suchmethoden für diese zu ermitteln. Aus einigen Quellen geht in diesem Kontext hervor, dass eine Kombination verschiedener Suchmethoden sinnvoll ist, um die Qualität und Menge der Ergebnisse der systematischen Literatursuche zu steigern [2,78,79]. Brings et al. [2] beruht dies beispielsweise darauf, dass die Schnittmenge der Ergebnisse verschiedener Suchtypen (manuelle Suche, Datenbanksuche und Snowballing) nur einen Bruchteil der Ergebnisse darstellt, die jeweils mit den einzelnen Suchtypen gefunden werden. Im Fall der manuellen Suche gibt es sogar keine Überschneidungen mit den anderen Suchtypen. Allerdings stellt Brings et al. [2] auch fest, dass sich die Suchtypen in Bezug auf Effektivität und Effizienz drastisch unterscheiden. Abbildung 2 und 3 zeigen in diesem Zusammenhang die Effektivität und Effizienz der drei betrachteten Suchmethoden. Die Effektivität bildet sich dabei aus dem Verhältnis der Anzahl der mit der jeweiligen Suchmethode gefundenen Paper und der Anzahl der insgesamt gefundenen Paper. Die Effizienz bildet sich dabei aus dem Verhältnis der Anzahl der relevanten Paper und der Anzahl der insgesamt betrachteten Paper.

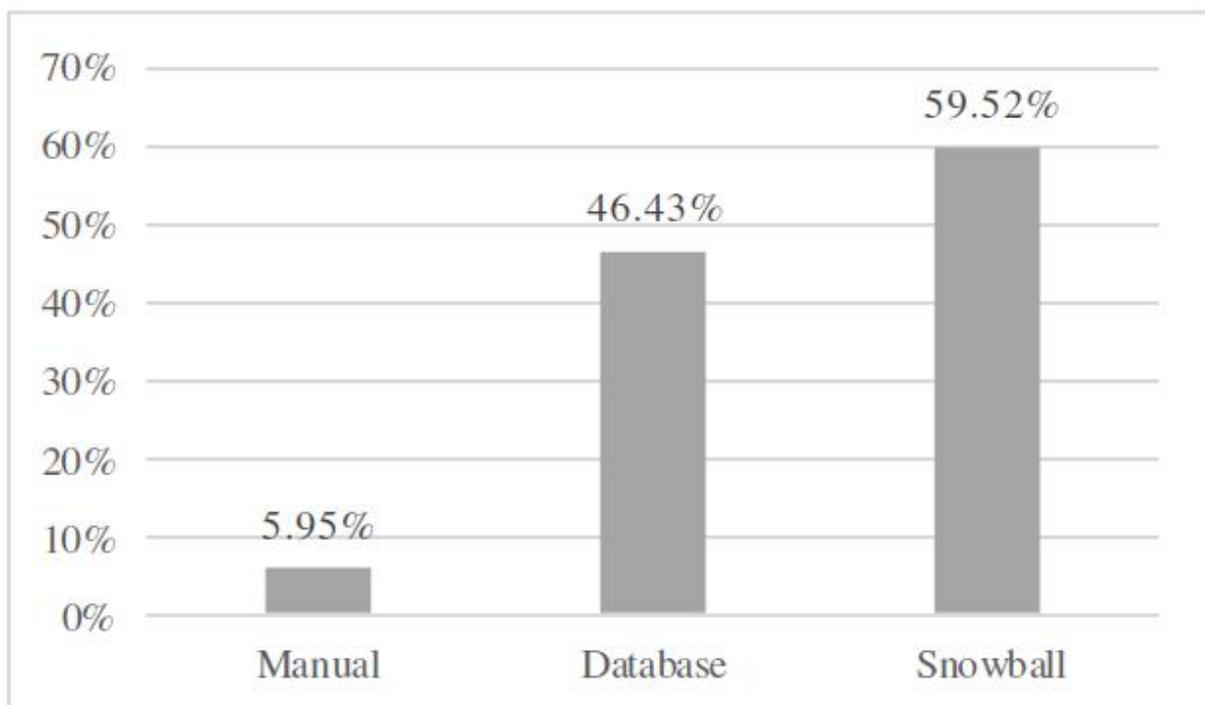


Abbildung 2: Effektivität der verschiedenen Suchmethoden nach Brings et al. [2]

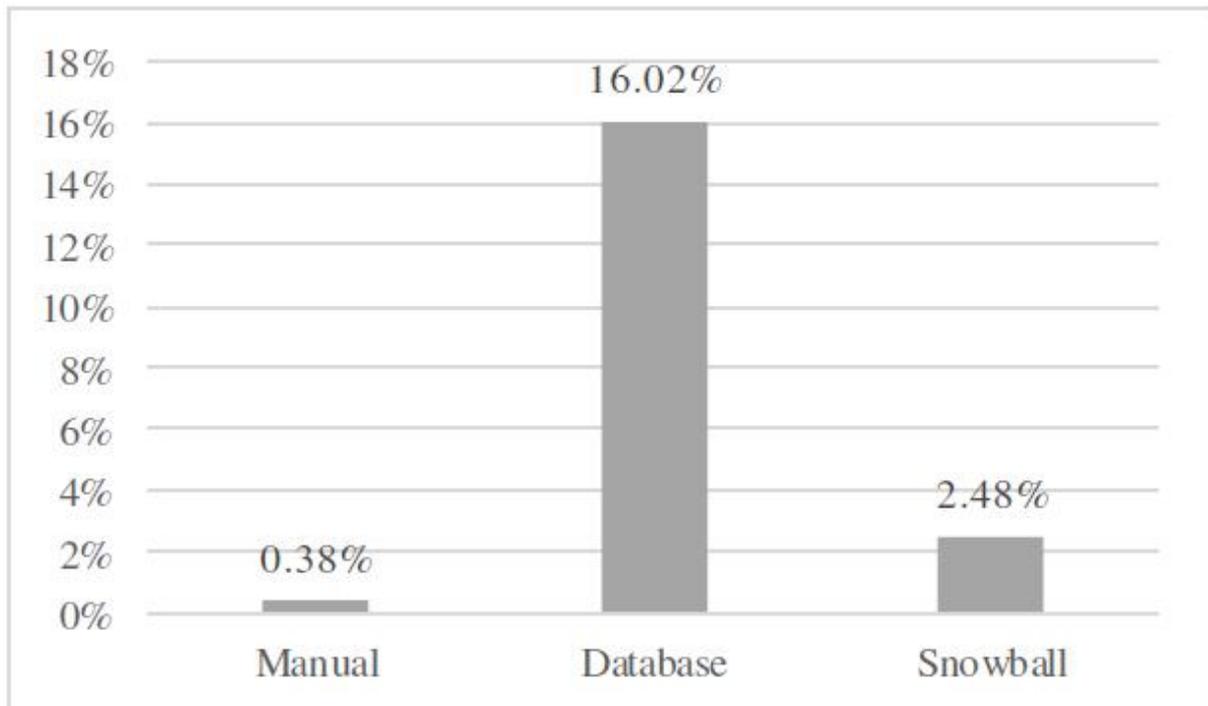


Abbildung 3: Effizienz der verschiedenen Suchmethoden nach Brings et al. [2]

Hieraus lässt sich schließen, dass Datenbanksuche und Snowballing mit Abstand sowohl effektivere als auch effizientere Suchmethoden sind als eine manuelle Suche. Darüber hinaus besitzt eine Datenbanksuche eine deutlich höhere Effizienz im Vergleich zum Snowballing oder zur manuellen Suche was bedeutet, dass sich mit der Datenbanksuche bei gleichem Aufwand eine deutlich größere Menge an Ergebnissen finden lässt. Auf der anderen Seite umfasst Snowballing mit 59,52% deutlich mehr Publikationen als die Datenbanksuche und die manuelle Suche zusammen und liefert somit die größte potentielle Ergebnismenge. Trotzdem kann anhand der Effektivität erkannt werden, dass der Anteil der manuellen Suche mit 5,95% einen relevanten Anteil der Suchergebnisse bildet, der wie vorher beschrieben, eine Menge an Ergebnissen finden kann, die komplett verschieden von der Menge der Ergebnisse der Datenbanksuche und Snowballing ist. Hieraus lässt sich letztendlich schlussfolgern, dass eine hybride systematische Literatursuche, bestehend aus den drei verschiedenen Suchmethoden, sinnvoll ist.

Aufgrund der beschriebenen Zusammenhänge zwischen manueller Suche, Datenbanksuche und Snowballing wird für diese Arbeit zunächst eine manuelle Suche durchgeführt. Das Vorgehen und die Ergebnisse der manuellen Suche werden hierfür im folgenden Kapitel erläutert. Auf Basis dieser manuellen Suche wird dann die Datenbanksuche durchgeführt. Die Ergebnisse dieser Datenbanksuche dienen dann wiederum als Ausgangspunkt für eine Suche mittels Snowballing.

3.2 Manuelle Suche

Die manuelle Suche bildet die erste Phase der systematischen Literatursuche dieser Arbeit. Das primäre Ziel dieser manuellen Suche ist es dabei, nicht nur thematisch relevante Paper zu finden, sondern auf Basis der Ergebnisse einen Suchstring zu bilden, mit dem eine Datenbanksuche durchgeführt werden kann.

Zunächst gilt es jedoch eine Menge an Inklusions- und Exklusionskriterien zu definieren, anhand derer die Auswahl der als relevant erachteten Publikationen vorgenommen wird. Im Kontext dieser Arbeit werden daher für die manuelle Suche die folgenden Inklusionskriterien definiert:

- I1: Die Publikation ist in deutsch oder englisch formuliert.
- I2: Der Inhalt der Publikation ist als wissenschaftliches Ergebnis verfasst, beispielsweise als Paper oder Buchveröffentlichung.
- I3: Die Publikation ist peer-reviewed.
- I4: Aus dem Titel und/oder der Zusammenfassung lässt sich ein Zusammenhang zum Thema „shared understanding“ ableiten.

Zusätzlich zu den Inklusionskriterien werden darüber hinaus die folgenden Exklusionskriterien definiert:

- E1: Es besteht keine Zugriffsmöglichkeit auf die gesamte Publikation .
- E2: Die Publikation fällt unter die Kategorie der grauen Literatur.

Unter grauer Literatur sind in dieser Arbeit sämtliche Publikationen zu verstehen, deren Qualität nicht anhand einer Veröffentlichung durch eine objektive Entität, wie beispielsweise durch ein Journal, bestätigt wurde. Bei diesen Kriterien ist zu betrachten, dass alle Inklusionskriterien zwingend erbracht werden müssen und gegen kein Exklusionskriterium verstoßen werden darf, damit eine Publikation berücksichtigt werden kann. Die hieraus resultierende logische Formel ist die folgende:

$$(I1 \wedge I2 \wedge I3 \wedge I4) \wedge \neg(E1 \vee E2)$$

Die Kriterien I2, I3 und E2 dienen hierbei dem Zweck, eine grundlegende Qualität der Publikationen, die für diese Arbeit betrachtet werden, zu gewährleisten. I1 muss gefordert werden, damit der Inhalt der Publikation ohne sprachliche Barrieren vollständig nachvollzogen werden kann. Das Kriterium E1 wird definiert, damit ein eventuell benötigter Kontext zur Zusammenfassung nachvollziehbar ist. Zuletzt dient I4 dazu, die thematische Relevanz der Publikation zu gewährleisten.

Nachdem die Inklusions- und Exklusionskriterien definiert wurden, muss eine Menge von Journals und Konferenzen bestimmt werden, die für die manuelle Suche relevant sind. In diesem Kontext besteht die hauptsächliche Schwierigkeit darin, eine geeignete Rangliste der Journals und Konferenzen zu erstellen. Die meisten Quellen weisen zwar den einzelnen Journals und Konferenzen eine Metrik zu, jedoch bieten sie keine zufriedenstellende Funktion, die eine thematische Kategorisierung vornehmen kann, um verwandte Journals und Konferenzen zu finden und miteinander zu vergleichen.

Die beiden Ranglisten, die im Kontext dieser Arbeit gefunden wurden und für sinnvoll erachtet werden, sind daher die Ranglisten von Feldt [3] und Google Scholar [4]. Diese beiden Ranglisten stellen dabei die relevantesten Journals und Konferenzen zum Thema Software Engineering anhand verschiedener Metriken dar. Da Requirements Engineering ein Teilgebiet von Software Engineering und „shared understanding“ wiederum ein Ziel des Requirements Engineerings darstellt, können die Ranglisten nicht direkt für diese Arbeit übernommen werden. Es wird daher anhand der generellen Thematik, mit der sich das jeweilige Journal beziehungsweise die jeweilige Konferenz beschäftigt, die Relevanz für diese Arbeit abgeschätzt.

Da die Menge aller möglichen Journals und Konferenzen, sowie die Anzahl ihrer Publikationen, so groß ist, dass es im Rahmen dieser Arbeit sowohl zeitlich nicht möglich als auch nicht zielführend wäre, sämtliche verfügbaren Quellen manuell zu durchsuchen, müssen an dieser Stelle zwei Einschränkungen der manuellen Suche vorgenommen werden. Die erste Einschränkung betrifft den Zeitbereich, in dem nach thematisch relevanten Inhalten gesucht wird. Anhand einer Menge initialer Paper, die für den Einstieg in die Thematik vom Betreuer dieser Arbeit bereitgestellt wurden, konnte unter Absprache mit dem Betreuer ein zeitlich relevanter Rahmen für die manuelle Suche bis ins Jahr 2010 festgelegt werden [54,55]. Unterstützt wird dieser zeitlich relevante Rahmen dadurch, dass die Definition von „shared understanding“ durch Aranda [54], die wie in Kapitel 2.2 beschrieben für diese Arbeit verwendet wird, ebenfalls im Jahr 2010 veröffentlicht wurde. Da die manuelle Suche, wie in Kapitel 3.1 beschrieben, primär dem Ziel dient, einen Suchstring für die Datenbanksuche zu generieren und somit keinesfalls einen Anspruch an Vollständigkeit erhebt, ist diese relativ grobe Abgrenzung in diesem Kontext ausreichend.

Die zweite Einschränkung bezieht sich auf die Menge der durchsuchten Journals und Konferenzen. Hierbei werden die Journals und Konferenzen anhand der bereits beschriebenen Abschätzung durchsucht, wobei die Auswahl mit absteigender Relevanz durchgeführt wird. Die Suche wird nach den acht relevantesten Journals und Konferenzen abgebrochen, da zum einen der zeitliche Rahmen der Arbeit einen Abschluss der Suche fordert und zum anderen neue Funde in den Journals und Konferenzen keine neuen Erkenntnisse für den Suchstring liefern.

Aufgrund der nahen Verwandtschaft von Requirements Engineering und anderen, interdisziplinären Domänen, wie beispielsweise Psychologie und Kommunikation, erschien es für diese Arbeit sinnvoll, auch interdisziplinäre Journals zu berücksichtigen. Die Betrachtung interdisziplinärer Studien könnte in diesem Kontext hilfreich sein, um Erkenntnisse aus anderen Domänen auf das Requirements Engineering zu übertragen. Da sich diese Journals jedoch außerhalb der Expertise von Informatikern befinden, kann bei der Auswahl hier nur stichprobenartig anhand des Titels der Journals vorgegangen werden. Aus diesem Grund werden daher für diese Arbeit die zwei vielversprechendsten Journals der Kommunikationsdomäne betrachtet.

Letztendlich werden die folgenden Journals und Konferenzen im Zeitbereich Januar 2010 bis Juni 2020 manuell durchsucht: International Conference of Software Engineering (ICSE), Information and Software Technology (IST), Journal of System and Software (JSS), International Symposium on Foundations of Software Engineering (FSE), Empirical Software Engineering (ESE), Transactions on Software Engineering (TSE), Transactions on Software Engineering and Methodology (TOSEM), Requirements Engineering Journal (RE), Journal of Computer-mediated Communication (JCC), Communication Theory (CT).

3.2.1 Ergebnisse

Da das Ziel dieser manuellen Suche die Generierung eines Suchstrings ist, wird die Menge der relevanten Publikationen entsprechend des Prinzips eines Quasi-Goldstandards gebildet. Das Prinzip des Quasi-Goldstandards wird von Zhang et al. [80] präsentiert. Ein Quasi-Goldstandard entspricht dabei der Menge bekannter Publikationen, die im Kontext der jeweiligen Arbeit als relevant betrachtet werden können. Der Quasi-Goldstandard ist dabei abhängig von der Menge der betrachteten Quellen, sowie dem Zeitrahmen, in dem die Quellen betrachtet werden [80]. Für die betrachteten Quellen und den betrachteten Zeitrahmen kann der Quasi-Goldstandard dabei als Grundgesamtheit angenommen werden,

während ein Goldstandard die wahre Grundgesamtheit repräsentiert [80]. Auf Basis eines Quasi-Goldstandards können entsprechend Evaluationen von Suchmethoden vorgenommen werden, indem der prozentuale Anteil der Publikationen des Quasi-Goldstandards bestimmt wird, der durch die jeweilige Suchmethode wiedergefunden wird [80]. Diese Metrik wird von Zhang et al. [80] als Quasi-Sensitivität bezeichnet. Da in dieser Arbeit keine wahre Grundgesamtheit bekannt ist, die als Goldstandard genügt, und der Rahmen, in dem die Suchmethoden durchgeführt werden, eingeschränkt ist, kann das Prinzip des Quasi-Goldstandards angewendet werden. Um diesen Quasi-Goldstandard zu gewährleisten, wird die initiale Menge an Ergebnissen nach der Vollendung des Suchprozesses abermals durchsucht, um die Relevanz der Funde zu garantieren und dadurch die Qualität des Quasi-Goldstandards zu verbessern. Durch diesen Prozess kann die Beobachtung getroffen werden, dass nur acht der Funde einem Quasi-Goldstandard entsprechen. Allerdings gibt es eine Menge von 16 Funden, die zwar inhaltlich dem Quasi-Goldstandard nicht genügen, jedoch thematisch sehr nah mit „shared understanding“ verwandt sind und Schlüsselwörter beinhalten, die für die Generierung des Suchstrings eine Bereicherung darstellen.

Aus diesem Grund wird neben dem Quasi-Goldstandard eine Menge „hilfreicher Publikationen“ definiert. Diese Menge beinhaltet alle Publikationen, die nicht der Definition des Quasi-Goldstandards genügen, jedoch thematisch ausreichend verwandt sind, um für den Suchstring berücksichtigt zu werden, sowie Schlüsselwörter enthalten, die eine Bereicherung für den Suchstring darstellen.

Um die Qualität der Ergebnisse einer systematischen Literatursuche zu verbessern, werden die einzelnen Schritte normalerweise von mindestens zwei voneinander unabhängigen Personen durchgeführt, die ihre Ergebnisse miteinander vergleichen, um den subjektiven Einfluss auf das Ergebnis zu minimieren. Da im Rahmen dieser Arbeit die Zusammenarbeit mit mehreren Personen aufgrund eines Mangels an verfügbaren Personen nicht möglich ist, wird sowohl der Quasi-Goldstandard als auch die Menge „hilfreicher Publikationen“ nach einem Zeitabstand von zwei Wochen abermals überprüft. Bei dieser Überprüfung ist die Zuordnung der Publikationen in die Kategorien Quasi-Goldstandard und „Menge hilfreicher Publikationen“ konsistent mit der initialen Zuordnung geblieben.

Tabelle 1 zeigt die initiale Verteilung der Anzahl der Publikationen abhängig von ihrem Ursprungsort sowie die darauf folgende Kategorisierung in Quasi-Goldstandard (QGS) und hilfreiche Publikationen. Eine Liste des Quasi-Goldstandards und der hilfreichen Publikationen ist in Anhang A beigelegt.

Tabelle 1: Verteilung der Anzahl der Publikationen in den Schritten der manuellen Suche

	ICSE	IST	JSS	FSE	ESE	TSE	TOSEM	RE	JCC	CT	Gesamt
betrachtet	1491	1306	2162	768	683	655	251	223	256	240	8035
initial	12	15	4	3	3	2	0	11	0	4	54
QGS	1	1	0	0	1	0	0	5	0	0	8
hilfreich	1	5	1	1	1	0	0	6	0	1	16

Letztendlich wurde über alle Journals und Konferenzen verteilt eine Menge von 8035 Publikationen durchsucht. Betrachtet man den Quasi-Goldstandard und die „Hilfreichen Publikationen“ zusammen, dann umfasst mit 11 relevanten Funden und damit einer Effektivität von 45,83% das RE Journal den mit Abstand größten Anteil der gefundenen relevanten Publikationen. Darüber hinaus ist es mit einer Effizienz von 4,9% die mit Abstand effizienteste hier betrachtete Quelle. Dies war zu erwarten, da sich das RE auf jegliche Themen zu Requirements Engineering spezialisiert hat, während die anderen Quellen einen allgemeineren Umfang zum Thema Requirements Engineering besitzen. Als weitere

beachtliche Quelle kann hier das IST identifiziert werden. Dieses Journal liefert zwar wie die meisten anderen der Journals und Konferenzen nur eine Publikation, die dem Quasi-Goldstandard genügt, jedoch war es hier möglich 5 Publikationen zu finden, die sich mit sehr nah verwandten Themen beschäftigen, unter anderem verschiedenen Modellansätzen [85,86]. In allen anderen betrachteten Journals und Konferenzen können zwischen gar keinen und 2 relevante Publikationen identifiziert werden. Abgesehen vom RE liegt die Effektivität der Suche bei allen anderen Journals und Konferenzen unter 0,5%, was auf den allgemeineren thematischen Umfang dieser zurückzuführen ist. Nach der Filterung der initialen Menge der Funde können aus TSE, TOSEM und JCMC keine Publikationen identifiziert werden, die relevant für das Ziel dieser Arbeit wären.

Um die Annahme des zeitlich relevanten Rahmens zu prüfen, der in 3.2 für die manuelle Suche von 2010 bis heute festgelegt wird, zeigt Abbildung 4 die Anzahl der relevanten Publikationen, die durch die manuelle Suche identifiziert wurden, geordnet anhand ihres Publikationsdatums.

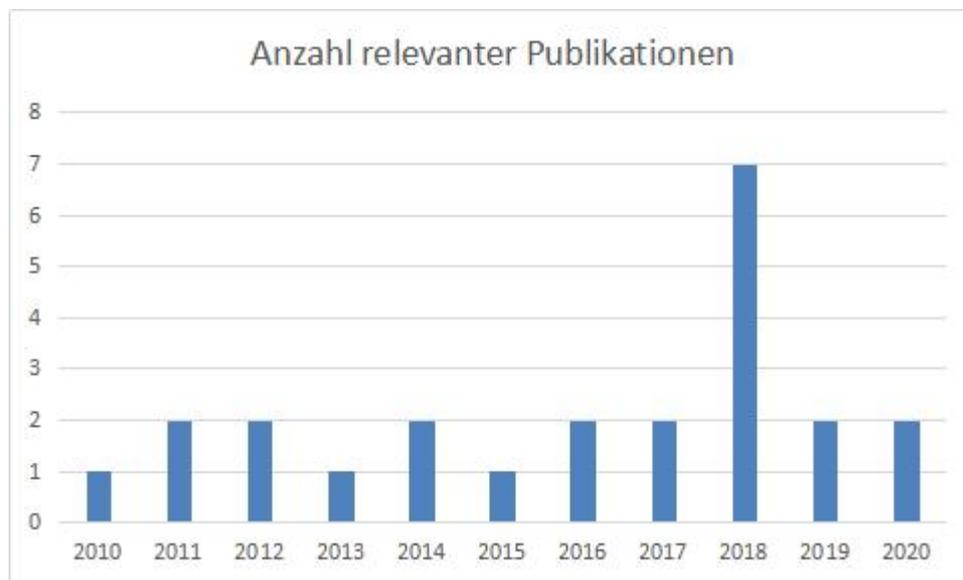


Abbildung 4: Anzahl relevanter Publikationen im Zeitraum von 2010 bis 2020

Um eine möglichst große und damit aussagekräftigere Menge an Daten zu betrachten, werden hierbei sowohl Publikationen des Quasi-Goldstandards als auch die Menge „hilfreicher Publikationen“ berücksichtigt. Aus der Verteilung lässt sich hierbei ein klarer Höhepunkt im Jahr 2018 erkennen. In allen anderen Jahren sind nur ein oder zwei Publikationen veröffentlicht worden. Es ist eine minimal abfallenden Tendenz Richtung 2010 zu erkennen, da sämtliche Jahre, in denen nur eine Publikation veröffentlicht wurden, vor 2018 liegen. Berücksichtigt man diese Tendenz, dann wird die Auswahl des Zeitrahmens von 2010 bis heute durch diese Verteilung unterstützt. Jedoch kann dies weniger als stichhaltiger Beweis, sondern eher als eine Vermutung betrachtet werden, da mit einer Anzahl von insgesamt 24 Publikationen an dieser Stelle keine durch statistische Signifikanz bewiesene Aussage getroffen werden kann.

3.2.2 Suchstring

Als finales Ergebnis der manuellen Suche gilt es noch einen Suchstring zu bilden, welcher für die Datenbanksuche genutzt werden kann. Hierfür werden sowohl die Publikationen des Quasi-Goldstandards als auch die der „Menge hilfreicher Publikationen“ nach passenden

Schlüsselwörtern durchsucht. Ein Schlüsselwort kann hierbei sowohl ein einzelnes Wort als auch eine Kombination von Wörtern sein, beispielsweise ein Adjektive und ein Nomen. Nach der Identifikation der Schlüsselwörter lassen sich diese thematisch gruppieren. Abbildung 5 zeigt die Gruppierung der gefundenen Schlüsselwörter.

Prefix	Suffix
Poor●	Understanding ●
„Lack of“●	Model●
Mutual●	Representation●
Common●	Distance●
Shared●●	Gap●
Process●	Viewpoint●
Partial●	Comprehension●
Conceptual●	
Internal●●	
Mental●●●	
Cognitive●	
Perception●	
Comprehension●	
Communication●	
Differing●●	
Varying●	

Abbildung 5: Gruppierung der Schlüsselwörter anhand von farblicher Codierung

Die Schlüsselwörter einer Gruppierung wurden zur besseren Übersicht farblich markiert, sodass alle Schlüsselwörter einer Farbe der selben Gruppierung angehören. Die Gruppierung muss vorgenommen werden, da die Kombination einiger Schlüsselwörter zu unerwünschten Suchbegriffen führen würde, die entweder grammatikalisch nicht sinnvoll sind oder im Allgemeinen einem anderen Kontext zuzuordnen sind. Aus den gleichen Gründen müssen einige Schlüsselwörter grundsätzlich ausgeschlossen werden, obwohl sie teilweise im Kontext von „shared understanding“ Anwendung finden. Ein Beispiel für ein solches vorbelastetes Schlüsselwort wäre „Design“. Dieses Schlüsselwort wird nicht mit aufgenommen, da es zwar im Kontext von Kommunikation valide wäre, jedoch in der allgemeinen Software Engineering Domäne stark vorbelastet ist und somit zu vielen Suchergebnissen führen würde, die im Kontext dieser Arbeit nicht relevant sind. Im Gegenzug hat die Gruppierung zur Folge, dass einige Schlüsselwörter mehreren Gruppierungen zugeordnet sind, da sie in verschiedenen Kontexten valide sein können.

Durch die Verkettung der Schlüsselwörter mit AND- und OR-Operatoren kann der in Abbildung 6 dargestellte Suchstring gebildet werden.

```

((shared OR mutual OR common OR differing OR varying OR „lack of“ OR poor) AND
  (understanding OR comprehension OR viewpoint)) OR
((perception OR comprehension OR communication OR cognitive OR mental) AND (gap OR
  distance)) OR
((shared OR mutual OR common OR mental OR cognitive OR internal OR process OR
  partial OR conceptual) AND (model OR representation))

```

Abbildung 6: Aus der manuellen Suche resultierender Suchstring

Der Suchstring lässt sich hierbei thematisch in drei Kategorien aufteilen, die zur besseren Übersicht farblich voneinander abgehoben sind. Die Kategorien repräsentieren dabei, auf

Basis von verschiedenen möglichen Formulierungen, jeweils eine anderen Ansatz bezüglich des „shared understandings“. Die erste Kategorie bildet die direkte Suche nach „shared understanding“ in Form von diversen synonym verwendeten Begriffen (blau gefärbt). Die zweite Kategorie befasst sich mit der genau gegenteiligen Formulierung, also „dem Fehlen von shared understanding“ (grün gefärbt). Neben Synonymen für Verständnis ist hier auch häufig die Verwendung von Begriffen zu beobachten, die sich auf die mentalen Differenzen zwischen verschiedenen Personen beziehen. Als letzte Art der Formulierung kann der Modellierungsaspekt von „shared understanding“ identifiziert werden (braun gefärbt). Dies bezieht sich auf das Prinzip der mentalen Modelle, die jede Person bildet und die in Einklang gebracht werden müssen, um ein „shared understanding“ zu erreichen.

3.3 Datenbanksuche

Auf der manuellen Suche aufbauend ist die Datenbanksuche die zweite Phase der systematischen Literatursuche. Im Gegensatz zum vorherigen Suchschritt soll das Ergebnis dieser Suche eine Menge von relevanter Literatur sein, die durch die Analyse zum Erreichen des Hauptziels der Arbeit beiträgt. Konkret muss diese Literatur Ansätze beinhalten, aus denen potentielle Metriken, Methode oder Heuristiken zum Messen von „shared understanding“ in Vision Videos abgeleitet werden können.

Aus diesem Grund müssen zunächst die Inklusions- und Exklusionskriterien, die in 3.2 definiert wurden, angepasst werden. Die neue Menge an Inklusions- und Exklusionskriterien für die Datenbanksuche beinhaltet daher die folgenden Punkte:

- I1: Die Publikation ist in deutsch oder englisch formuliert.
- I2: Der Inhalt der Publikation ist als wissenschaftliches Ergebnis verfasst, beispielsweise als Paper oder Buchveröffentlichung
- I3: Die Publikation ist peer-reviewed
- I4: Die Publikation muss einen Ansatz beinhalten, aus dem potentiell eine Metrik, Methode oder Heuristik zur Messung von „shared understanding“ von Vision Videos abgeleitet werden kann.

- E1: Es besteht keine Zugriffsmöglichkeit auf die gesamte Publikation
- E2: Die Publikation fällt unter die Kategorie der grauen Literatur

Genau wie in 3.2 ist auch hier zu betrachten, dass alle Inklusionskriterien zwingend erbracht werden müssen und gegen kein Exklusionskriterium verstoßen werden darf, damit eine Publikation berücksichtigt werden kann. Daher gilt auch hier die identische logische Formel:

$$(I1 \wedge I2 \wedge I3 \wedge I4) \wedge \neg(E1 \vee E2)$$

Die entscheidende Änderung liegt hier beim Inklusionskriterium I4 vor. Zuvor sollte lediglich ein Suchstring für die Datenbanksuche erstellt werden, daher hat ein allgemeiner Zusammenhang der Publikation zum Thema „shared understanding“ genügt. Da nun jedoch die Ergebnisse der Datenbanksuche relevante Inhalte für die Analyse liefern sollen, muss das Kriterium verschärft werden. Dies hat zur Folge, dass nun auch der Volltext für die Auswahl der Publikationen berücksichtigt werden muss.

Nachdem die Inklusions- und Exklusionskriterien definiert wurden, muss im nächsten Schritt eine Menge von Datenbanken ermittelt werden, auf denen die Suche durchgeführt wird.

Zhang und Babar [5] haben in ihrer Arbeit eine Rangliste der am häufigsten genutzten Datenbanken in systematischen Literatursuchen erstellt. Tabelle 2 zeigt hierbei die Rangliste der Top 11 Datenbanken.

Tabelle 2: Rangliste der Häufigkeit der Nutzung von Datenbanken in systematischen Literatursuchen nach Zhang und Babar[5]

Rank	Search engine	# of SLRs	% of SLRs
1	IEEE Xplore	24	92%
2	ACM digital library	21	81%
3	ScienceDirect	15	58%
4	ISI Web of Science	10	38%
5	El Compendex	9	35%
6	SpringerLink	8	31%
6	Wiley InterScience	8	31%
6	Inspec	8	31%
9	Google Scholar	6	23%
10	SCOPUS	2	8%
10	Kluwer	2	8%

Um den Umfang der Datenbanksuche für diese Arbeit in einem überschaubaren Rahmen zu halten, werden als Ziel drei spezifischere Verlage sowie zwei generelle Datenbanken vorgesehen. Da IEEE Xplore und ACM digital library mit 92% und 81% die deutliche Spitze der Rangliste bilden, müssen diese beiden Verlage auf jeden Fall berücksichtigt werden. Darauf folgend wurden sowohl ScienceDirect als auch Web of Science als generellere Datenbanken aufgenommen. Als letzter Verlag wird entgegen der Rangliste SpringerLink gewählt, obwohl El Compendex eine bessere Platzierung besitzt. Dies ist damit zu begründen, dass in dem in 3.2.1 definierten Quasi-Goldstandard der manuellen Suche sechs der acht Publikationen bei SpringerLink veröffentlicht wurden. Daher lässt sich ein thematischer Schwerpunkt in Bezug auf „shared understanding“ bei SpringerLink vermuten. Darüber hinaus liegt der Unterschied der Anzahl der systematischen Literatursuchen zwischen SpringerLink und El Compendex lediglich bei einer einzigen Publikation. Aus diesen Gründen wird SpringerLink als dritter Verlag ausgewählt.

3.3.1 Suchprozess

Nachdem eine Menge von Datenbanken ausgewählt wurde, in denen die Suche durchgeführt werden soll, ist der letzte Schritt, den in 3.2.2 definierten Suchstring auf die jeweilige Syntax der Datenbank anzupassen. Grundsätzlich unterstützt bis auf SpringerLink jede der Datenbanken die grundlegenden logischen Operatoren, die für die Konstruktion des Suchstrings verwendet wurden. Im Fall von SpringerLink ist es jedoch nicht möglich, komplexere Kombinationen von UND und ODER zu verwenden. Aus diesem Grund muss daher der Suchstring in alle möglichen Kombinationen aufgeteilt und jede der Kombinationen einzeln gesucht werden. Eine Liste der Kombinationen sowie der Anzahl der Funde der jeweiligen Kombination ist in Anhang B beigelegt. Eine weitere Problematik stellt sich bei der Suche mit Science Direct heraus. Diese Datenbank erlaubt nur eine begrenzte Anzahl an Operatoren pro Suchanfrage, die vom hier verwendeten Suchstring weit überschritten wird. Aus diesem Grund muss in diesem Fall der Suchstring in sieben Teilsuchstrings aufgeteilt werden. Diese werden so konzipiert, dass wenn alle Teilsuchstrings mit ODER-Operatoren

verknüpft werden, der resultierende String die gleiche Ergebnismenge wie der originale Suchstring liefern würde.

Ein weiteres Problem, was sich kurz nach Beginn der Datenbanksuche offenbart, ist, dass der Suchstring in der in 3.2.2 definierten Form bei weitem zu viele Ergebnisse liefert. Als Beispiel kann hier die Suche mit ACM digital library herangezogen werden. Eine Suche unter Verwendung des Suchstrings liefert über 480.000 Ergebnisse. Aus diesem Grund müssen einige Einschränkungen gemacht werden, um die Menge der Suchergebnisse in eine Größenordnung zu bringen, in der sie bearbeitet werden kann. Der grundlegende überarbeitete Suchstring sieht dabei wie folgt aus:

```
((((shared OR mutual OR common OR differing OR varying OR „lack of“ OR poor) AND
      (understanding OR comprehension OR viewpoint)) OR
((perception OR comprehension OR communication OR cognitive OR mental) AND (gap OR
      distance)) OR
((shared OR mutual OR common OR mental OR cognitive OR internal OR process OR
      partial OR conceptual) AND (model OR representation))) AND
(„Requirements Engineering“ OR „Software Engineering“ OR „Human Computer
      Interaction“)
```

Abbildung 7: Angepasster Suchstring anhand der Einschränkungen der Datenbanksuche

Die erste Anpassung betrifft den Rahmen, in dem der Suchstring angewendet wird. Auch wenn „shared understanding“ keine Thematik darstellt, die exklusiv im Kontext von Software Engineering betrachtet werden kann, liegt das primäre Interesse dieser Arbeit in diesem Bereich. Daher wurde dem Suchstring ein weiterer Term hinzugefügt, der einen Bezug zu Requirements Engineering, Software Engineering oder Human Computer Interaction fordert, hier in grün markiert. Insbesondere wurde an dieser Stelle Human Computer Interaction berücksichtigt, da in diesem Fachgebiet viele Parallelen zum Software Engineering bestehen [88] und einige der Ergebnisse der manuellen Suche sich auf dieses Gebiet beziehen.

Als zweites müssen einige Begriffe aus dem Suchstring entfernt werden, da sie in dieser Suche zu einer Vielzahl von Ergebnissen führen würden, die im Kontext dieser Arbeit nicht relevant sind, da sie sich auf einen komplett anderen Kontext beziehen und somit keine hilfreichen Ergebnisse liefern. Ein Beispiel hierfür ist das Wort „conceptual“ in Kombination mit „model OR representation“. Diese Kombination von Wörtern liefert nicht, wie zunächst angenommen, Publikationen zu konzeptionellen Modellen in Form von mentalen Ideen. Stattdessen liefert diese Kombination eher Ergebnisse im Bereich von Konzeptentwürfen in Projekten, was nicht relevant für diese Arbeit ist. Neben „conceptual“ werden daher die Wörter „common“ und „process“ entfernt, die zu ähnlichen Problemen führen würden, hier in rot markiert.

Als letzte Anpassung ist abhängig von der Datenbank der Suchbereich einzuschränken, auf den die Begriffe angewendet werden. Hierbei kann leider keine einheitliche Lösung für alle Datenbanken gefunden werden. Dies lässt sich darauf zurückführen, dass die verschiedenen Datenbanken unterschiedliche Optionen zum Filtern des Suchstrings bereitstellen. Darüber hinaus verhalten sich gleich benannte Suchoptionen abhängig von der Datenbank unterschiedlich. Ein Beispiel dafür ist die Suche nach Titel. Während anzunehmen wäre, dass eine Suche nach Titel nur die Titel der Publikationen durchsucht, scheint die Titelsuche bei ACM auch die Indizes der jeweiligen Publikation zu berücksichtigen. Aus diesen Gründen muss der Suchbereich, auf den der Suchstring angewendet wird, individuell auf jede Datenbank angepasst werden, um eine überschaubare Menge an Ergebnissen zu ermöglichen. Eine Liste der genauen Suchstrings für jede Datenbank ist in Anhang C beigelegt.

3.3.2 Ergebnisse

Das Ergebnis der Datenbanksuche ist ähnlich wie bei der manuellen Suche in einem iterativen Prozess entstanden. Tabelle 3 zeigt hierzu die Ergebnismengen in den einzelnen Phasen der Datenbanksuche abhängig von der jeweiligen Datenbank.

Tabelle 3: Verteilung der Anzahl der Ergebnisse in den Schritten der Datenbanksuche

	IEEE	ACM	SpringerLink	Science Direct	Web of Science	Gesamt
betrachtet	648	869	653	707	896	3773
initial	29	34	29	36	25	153
gefiltert	15	15	20	15	13	78
Volltext	9	7	13	8	6	43

Zunächst wurden die einzelnen Datenbanken mithilfe der jeweiligen Suchstrings durchsucht. Hierbei wurde nur anhand des Titels und der Zusammenfassung entschieden, ob die Publikation relevante Inhalte besitzt. Um den Verlust von relevanten Publikationen zu minimieren, wurden dabei auch Publikationen berücksichtigt, bei denen die Zuordnung unsicher war. In diesem Schritt wurde daher eine Menge von 3773 Publikationen durchsucht, wovon 153 als relevante Literatur vermerkt wurden.

Im nächsten Schritt wurden die als relevant erachteten Paper abermals anhand von Titel und Zusammenfassung gefiltert. Dieses doppelte Filtern wurde wie auch in 3.2.1 durchgeführt, um die Qualität der Ergebnisse zu verbessern, da normalerweise mindestens zwei Forscher bei einer systematischen Literatursuche die Auswahl relevanter Publikationen durchführen. Dadurch wurde die Menge an Ergebnissen von ursprünglich 153 auf 78 Publikationen eingeschränkt. Diese Reduzierung liegt größtenteils daran, dass beim zweiten Filterdurchgang genauer auf Feinheiten in der Formulierung geachtet wurde, durch die viele Publikationen ausgeschlossen werden konnten.

Im letzten Schritt wurden die 78 Publikationen anhand ihrer Inhalte im Volltext geprüft. Dabei wurden entsprechend des Inklusionskriteriums I4 nur diejenigen Publikationen als relevant klassifiziert, in denen mögliche Ansätze zu Metriken, Methoden oder Heuristiken zur Messung von „shared understanding“ vorhanden waren. Durch diesen letzten Schritt konnte nach Entfernung von Duplikaten eine Menge von 43 tatsächlich relevanten Publikationen identifiziert werden.

Bezüglich der einzelnen Datenbanken lässt sich feststellen, dass SpringerLink mit 13 relevanten Publikationen die meisten Ergebnisse liefert. Dies unterstützt die in 3.3 vorgenommene Bevorzugung von SpringerLink vor El Compendex. Die anderen Datenbanken sind mit Mengen von jeweils sieben bis zehn Ergebnissen relativ nah beieinander. Neben den absoluten Zahlen lassen sich auch Rückschlüsse aus der Effizienz entsprechend der Definition in 3.1 ziehen. Abbildung 8 zeigt die Effizienz der Suche in den jeweiligen Datenbanken sowie die Effizienz der gesamten Datenbanksuche.

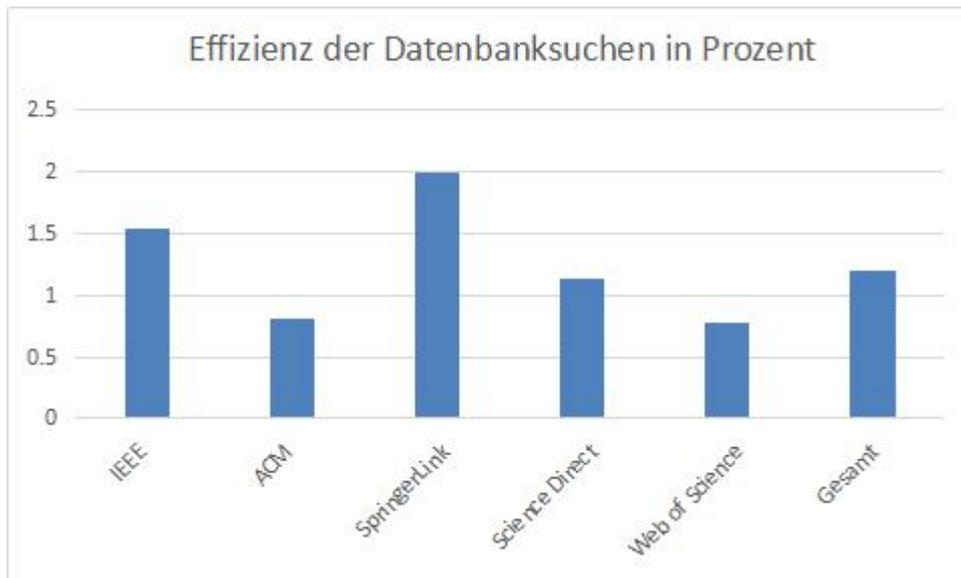


Abbildung 8: Effizienz der Datenbanksuchen in Prozent

Hierbei lässt sich die gleiche Rangfolge der einzelnen Datenbanken feststellen. Auffällig ist jedoch, dass keine der Datenbanken eine Effizienz von 2% übersteigt und die gesamte Effizienz der Suche lediglich bei 1,19% liegt. Dies ist ein extremer Unterschied zu den von Brings et al. [2] berichteten 16,02%. Dies könnte darauf zurückzuführen sein, dass der in dieser Arbeit verwendete Suchstring allgemeiner formuliert ist als im Beispiel von Brings, da die genauen Begrifflichkeiten im Kontext von „shared understanding“ noch relativ unklar waren.

Um die Qualität der durchgeführten Datenbanksuche zu überprüfen kann an dieser Stelle der Quasi-Goldstandard herangezogen werden, der durch die manuelle Suche definiert wurde. Hierfür muss die Quasi-Sensitivität bestimmt werden, die den prozentualen Anteil der wiedergefundenen Publikationen des Quasi-Goldstandards repräsentiert. Von den acht Publikationen, die durch die manuelle Suche als Quasi-Goldstandard identifiziert wurden, konnten vier Publikationen während der Datenbanksuche wiedergefunden werden. Dies führt zu einer Quasi-Sensitivität von 50%. Obwohl ein Wert von 50% zunächst vermuten lässt, dass der verwendete Suchstring nicht ausreichend war, gibt es zwei entscheidende Gründe dafür, warum die Hälfte der Publikationen des Quasi-Goldstandards nicht wiedergefunden wurden. Der erste Grund hierfür ist, dass nach der Erstellung des ursprünglichen Suchstrings einige Anpassungen an diesem vorgenommen werden mussten, um eine verwertbare Ergebnismenge zu erhalten. Es ist daher möglich, dass durch die thematische Eingrenzung des Suchstrings in den Kontext von Requirements Engineering, Software Engineering und Human Computer Interaction potentiell relevante Publikationen nicht mehr gefunden werden können. Ein zweiter Faktor, der für die niedrige Quasi-Sensitivität verantwortlich sein kann, ist, dass drei der vier Publikationen des Quasi-Goldstandards, die nicht wiedergefunden wurden, von SpringerLink publiziert wurden. Wie in Kapitel 3.3.1 war es bei SpringerLink nicht möglich, komplexe Suchanfragen zu stellen, wodurch alle möglichen Kombinationsmöglichkeiten des Suchstrings einzeln gesucht werden mussten. Darüber hinaus konnten nur Publikationen berücksichtigt werden, bei denen der jeweilige Suchbegriff im Titel enthalten war, da keine Suche in der Zusammenfassung oder den Indizes möglich war. Aus diesen Gründen lässt sich vermuten, dass einige der Publikationen des Quasi-Goldstandards nicht wiedergefunden wurden, da SpringerLink keine komplexen Suchanfragen mittels eines Suchstrings ermöglicht. Unter Berücksichtigung dieser Einschränkungen kann für die Datenbanksuche eine Quasi-Sensitivität von 50% als akzeptabel angesehen werden.

Da sowohl die Ergebnisse der Datenbanksuche als auch die Ergebnisse des anschließenden Snowballings relevant für die Auswertung sind, wird als nächstes zunächst der Prozess des Snowballings erläutert, um beide Ergebnismengen danach zusammen in Kapitel 4 zu analysieren.

3.4 Snowballing

Das Snowballing stellt den letzten Schritt des Suchprozesses der systematischen Literatursuche in dieser Arbeit dar. Snowballing ist hierbei eine Methode, die auf Basis von einer initialen Menge von relevanten Publikationen weitere thematisch verwandte Publikationen ermittelt. Hierbei gibt es zwei Arten von Snowballing, die beide in dieser Arbeit zur Anwendung kommen. Zum einen gibt es das rückwärts Snowballing, bei dem die Referenzen von relevanten Publikationen nach weiteren thematisch verwandten Publikationen durchsucht werden. Das genaue Gegenteil dazu bildet das vorwärts Snowballing, bei dem diejenigen Publikationen durchsucht werden, welche die vorliegende relevante Publikation referenzieren. Sowohl vorwärts als auch rückwärts Snowballing kann hierbei iterativ verwendet werden, um den Umfang der Ergebnismenge zu steigern.

Das Snowballing in dieser Arbeit dient dem gleichen Zweck wie die vorhergehende Datenbanksuche. Es sollen also relevante Publikationen identifiziert werden, die Ansätze beinhalten, mit welchen Metriken, Methoden oder Heuristiken zum Messen von „shared understanding“ abgeleitet werden können. Auch beim Snowballing müssen deshalb klare Inklusions- und Exklusionskriterien vorliegen, anhand derer man die Auswahl der relevanten Publikationen definiert. Da das Snowballing jedoch das gleiche Ziel wie die Datenbanksuche besitzt, können hierfür die Inklusions- und Exklusionskriterien aus 3.3 übernommen werden.

3.4.1 Suchprozess

Der konkrete Suchprozess des Snowballings in dieser Arbeit besteht aus zwei Iterationen von sowohl vorwärts als auch rückwärts Snowballing. Als initiale Menge von relevanten Publikationen wurde die Ergebnismenge von 43 Publikationen aus der Datenbanksuche verwendet, da diese aufgrund ihrer thematischen Relevanz ausgewählt wurden. Die Publikationen, die im Quasi-Goldstandard der manuellen Suche enthalten sind, werden hierbei nicht berücksichtigt, da sie als Kontrollmenge für die Ergebnisse aus der Datenbanksuche und des Snowballings dienen sollen.

Während der ersten Iteration des Snowballings wird sowohl vorwärts als auch rückwärts Snowballing auf die initiale Menge an Publikationen durchgeführt. In der zweiten Iteration wird dann wiederum auf beiden Ergebnismengen der ersten Iteration sowohl vorwärts als auch rückwärts Snowballing angewendet.

Während das rückwärts Snowballing anhand der Referenzen der jeweiligen Publikationen durchgeführt werden kann, muss für das vorwärts Snowballing eine Suchmaschine gewählt werden, die möglichst alle Publikationen findet, welche die jeweilige Publikation referenzieren. Zu diesem Zweck wurde Google Scholar verwendet. Auch wenn Google Scholar im Allgemeinen als keine optimale Suchmaschine für wissenschaftliche Arbeiten gilt [81], findet Google Scholar aufgrund des Umfangs und der Allgemeinheit der Suche die vollständigste Menge an Publikationen, die das jeweilige Paper referenzieren.

Nach der zweiten Iteration wurde entschieden, das Snowballing für diese Arbeit zu beenden. Hierfür gibt es vier voneinander abhängige Gründe, die zu dieser Entscheidung geführt haben. Zunächst ist aufgefallen, dass durch die beiden Iterationen des Snowballings zwar neue

Literatur gefunden wurde, welche sich mit den Inhalten der Ergebnismenge der Datenbanksuche decken, jedoch konnten keine Publikationen identifiziert werden, die komplett neue Ansätze präsentieren. Aus diesem Grund war der Mehrwert der Ergebnisse auf die Festigung von bereits bekannten Konzepten begrenzt. Im direkten Zusammenhang damit ist der zweite Grund, dass zwischen der Datenbanksuche und den beiden Iterationen des Snowballings bereits genügend Publikationen gefunden wurden, um die identifizierten Ansätze umfangreich zu analysieren. Somit wird der bereits begrenzte Mehrwert neuer Ergebnisse weiter gemindert. Die beiden anderen Gründe für die Beendigung des Snowballings beziehen sich auf die allgemeine zeitliche Begrenzung der Arbeit. Da diese Arbeit in einem begrenzten zeitlichen Rahmen fertiggestellt werden muss, wurde der mögliche minimale Mehrwert durch weitere Iterationen des Snowballings im Verhältnis zum Zeitaufwand als zu gering eingeschätzt. Zuletzt kam noch hinzu, dass sich in der Ergebnismenge der zweiten Iteration viele Publikationen befinden, die sehr allgemeine Konzepte zum Thema „shared understanding“ behandeln. Dies hat zur Folge, dass diese Publikationen teilweise von mehreren hundert bis tausend anderen Publikationen referenziert werden, welche in einer potentiellen dritten Iteration durchsucht werden müssten, was den Zeitaufwand nur noch weiter verstärkt. Durch das Zusammenspiel dieser vier Gründe wurde daher entschieden, dass der mögliche als minimal geschätzte Mehrwert den Zeitaufwand einer dritten Iteration nicht rechtfertigt.

3.4.2 Ergebnisse

Um den Prozess und die Ergebnisse des Snowballings darzustellen zeigt Abbildung 9 die einzelnen Iterationen des Snowballings mit den dazugehörigen Ergebnismengen.

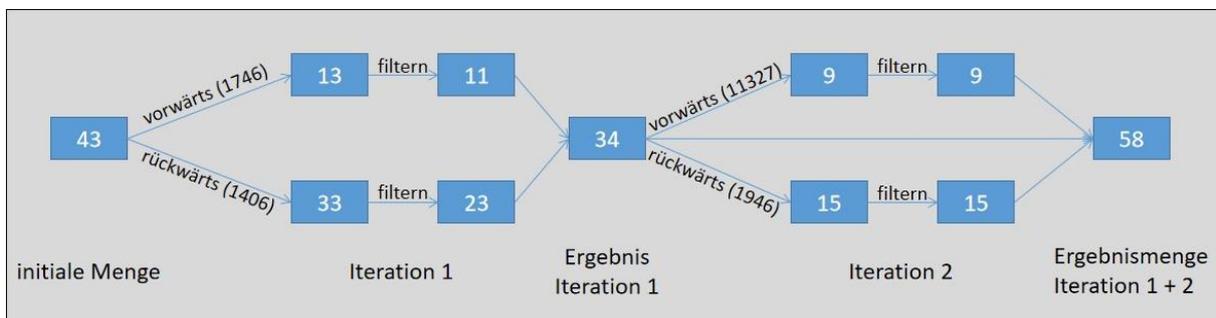


Abbildung 9: Ablauf und Ergebnisse der einzelnen Schritte des Snowballings

Die Auswahl der relevanten Publikationen der jeweiligen Iteration wurde dabei, wie auch in den vorhergehenden Suchphasen, in zwei Schritten durchgeführt. Zunächst wurde eine Vorauswahl an möglichen relevanten Publikationen anhand des jeweiligen Titels und der jeweiligen Zusammenfassung durchgeführt. Auch in diesem Fall wurden unsichere Publikationen zunächst mit berücksichtigt. Im zweiten Schritt wurden die Publikationen dann im Volltext inhaltlich überprüft, um sie letztendlich als relevant zu identifizieren. In dem Fall, dass sich Publikationen bereits in der Ergebnismenge der Datenbanksuche oder dem vorherigen Iterationsschritt befinden, wurden sie nicht nochmals berücksichtigt. Aus der ersten Iteration ging hierbei zunächst eine Menge von 33 Publikationen aus dem rückwärts Snowballing hervor, die nach dem zweiten Schritt auf 23 relevante Publikationen reduziert wurde. Das vorwärts Snowballing brachte im Gegensatz dazu 13 Publikationen hervor, die auf 11 reduziert wurden. Die zweite Iteration lieferte daraufhin mit 15 Publikationen durch rückwärts und 9 Publikationen durch vorwärts Snowballing eine geringere Ergebnismenge, bei der jedoch im zweiten Schritt keine Publikationen herausgefiltert werden konnten. Insgesamt liefert das Snowballing also eine Ergebnismenge von 58 relevanten Publikationen.

Darüber hinaus unterliegt die Ergebnismenge des Snowballings im Gegensatz zu der der Datenbanksuche keinen thematischen Einschränkungen. Dies hat zur Folge, dass hier auch einige interdisziplinäre Publikationen gefunden werden konnten, da der Rahmen des Snowballings nicht auf den Kontext des Software Engineerings begrenzt werden musste. Ein Beispiel hierfür bilden Publikationen aus der Domäne des Gesundheitswesens, die in diesem Kontext gefunden werden konnten und als thematisch relevant betrachtet wurden [22].

Um die Häufigkeit der Ergebnisse mit den anderen Suchphasen vergleichen zu können, wird auch an dieser Stelle die Effizienz der einzelnen Iterationen entsprechend der Definition in 3.1 berechnet. Abbildung 10 zeigt hierfür die Effizienz der einzelnen Iterationen sowie die Effizienz des gesamten Snowballings.

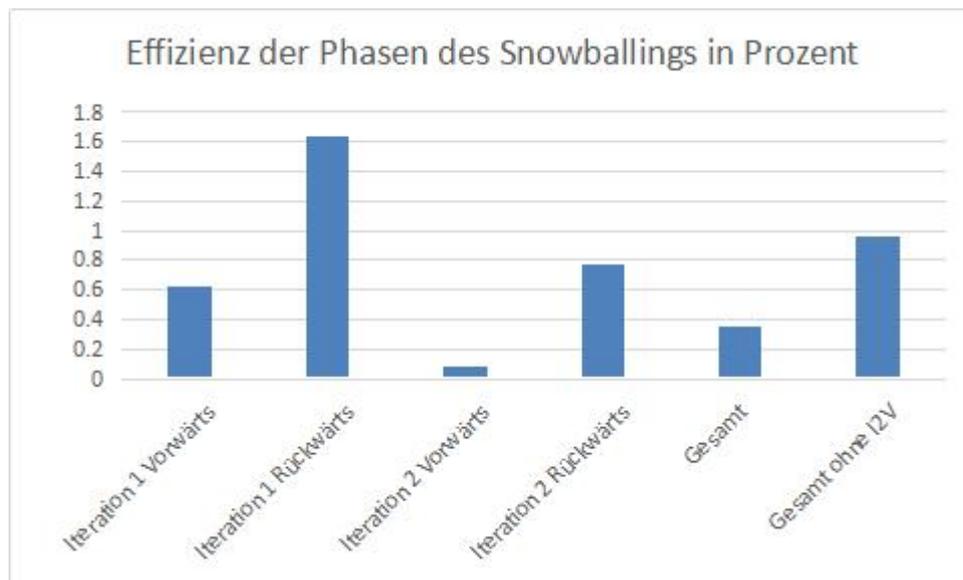


Abbildung 10: Effizienz der einzelnen Schritte des Snowballings in Prozent

Vergleicht man die beiden Iterationen, so lässt sich sowohl beim vorwärts als auch beim rückwärts Snowballing ein deutlicher Abfall der Effizienz beobachten. Das war zu erwarten, da mit jeder Iteration die Menge der noch nicht berücksichtigten relevanten Publikationen abnimmt und es somit schwieriger wird, neue relevante Publikationen zu finden. Auffällig ist hierbei noch der extrem geringe Wert für das vorwärts Snowballing in der zweiten Iteration. Dies ist darauf zurückzuführen, dass in der Menge der Publikationen dieser Iteration viele dabei waren, die grundlegende Konstrukte erklären und dadurch sehr oft referenziert wurden. Ein Beispiel hierfür wäre die Arbeit von Mathieu et al. [82], die den Einfluss von „shared mental models“ auf Teamprozesse und Teamperformance untersucht. Die Arbeit betrachtet einige grundlegende Konzepte wie mentale Modelle und gemeinsamen mentale Modelle und wurde daher 2921 mal referenziert. Um die Effizienz des gesamten Snowballings durch diese Besonderheit nicht zu sehr zu verfälschen, wurde daher zwei Effizienzwerte berechnet. Der erste Wert mit einer Effizienz von 0,35% berücksichtigt dabei alle Iterationen, während der zweite Wert mit einer Effizienz von 0,96% das vorwärts Snowballing der zweiten Iteration nicht berücksichtigt.

Vergleicht man diese Werte mit denen von Brings et al. [2] lässt sich feststellen, dass sie, wie auch schon bei der Datenbanksuche, nicht mit den 2,48% von Brings et al. [2] übereinstimmen, die Abweichung jedoch wesentlich geringer ist als bei der Datenbanksuche. Dies lässt sich damit begründen, dass zwar wie auch bei der Datenbanksuche der Rahmen der Suche komplexer ist als bei Brings et al. [2], jedoch durch die Ergebnisse der Datenbanksuche deutlich präzisiert wurde, was zu einer Verringerung der Abweichung führt. Entsprechend

dieser Beobachtung lässt sich schlussfolgern, dass die Berücksichtigung der Snowballing-Methode aufgrund der Empfehlung von Brings et al [2]. sinnvoll war.

Im folgenden Kapitel wird ein Überblick über die relevanten Publikationen geliefert, welche durch die drei verwendeten Suchmethoden gefunden wurden, sowie eine gesamtheitliche Analyse aller Ergebnisse der Suchmethoden durchgeführt.

4 Auswertung

In diesem Kapitel werden die Ergebnisse der systematischen Literatursuche aus Kapitel 3 analysiert. Hierbei wird zunächst der gesamte Prozess der systematischen Literatursuche betrachtet und die einzelnen Phasen anhand ihrer Ergebnisse miteinander verglichen. Darüber hinaus werden einige der Einschränkungen, die bei den einzelnen Suchmethoden vorgenommen wurden, durch statistische Auswertungen der Ergebnisse überprüft. Danach wird inhaltlich auf die in den Ergebnissen gefundenen Konzepte eingegangen, die sich auf Metriken, Methoden und Heuristiken zur Messung von “shared understanding” beziehen. Die gefundenen Konzepte werden dabei thematisch kategorisiert.

4.1 Ergebnisübersicht

In diesem Kapitel soll zunächst ein Überblick über die Gesamtmenge an Publikationen geschaffen werden, die im Zuge der Phase der systematischen Literatursuche betrachtet wurden. Abbildung 11 visualisiert hierfür den Prozess der systematischen Literatursuche in dieser Arbeit. In der Abbildung wurde dabei auf einige Details der einzelnen Phasen verzichtet, um die Abbildung übersichtlicher zu gestalten.

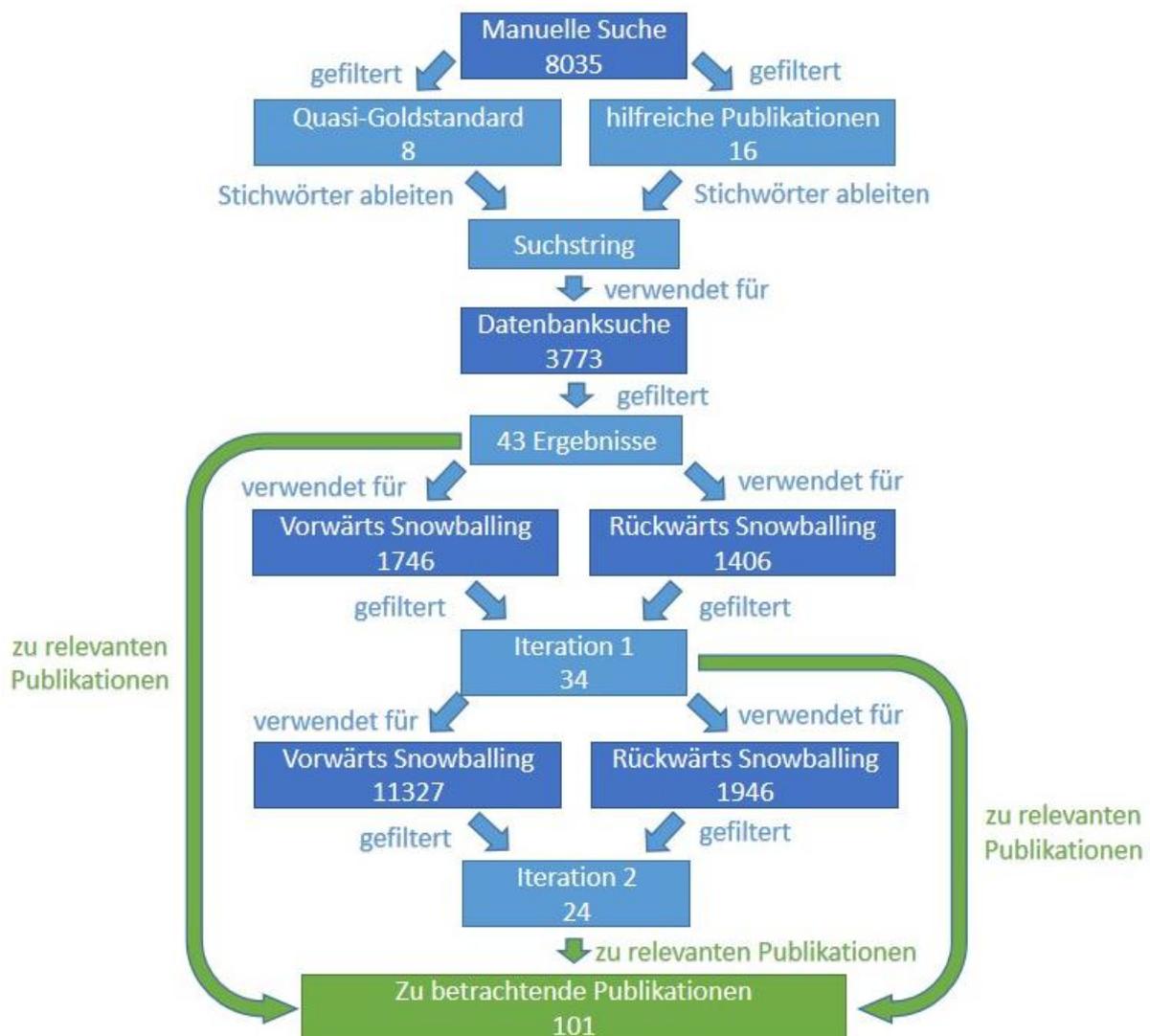


Abbildung 11: Überblick über den Ablauf der gesamten systematischen Literatursuche

Die dunkelblauen Felder stellen in Abbildung 11 die einzelnen durchgeführten Phasen dar mit der jeweiligen Anzahl der Publikationen, die während der Phase betrachtet wurden. Die Zwischenergebnisse der Phasen werden in hellblau dargestellt. Während der Durchführung der systematischen Literatursuche wurde über alle Phasen verteilt eine Menge von insgesamt 28.233 Publikationen durchsucht. Hierbei wurde primär anhand des Titels entschieden, ob die jeweilige Publikation relevant für diese Arbeit ist. In unsicheren Fällen wurde für die Kategorisierung die Zusammenfassung der jeweiligen Publikation hinzugezogen. Die einzige Einschränkung, die in diesem Zusammenhang getroffen wurde, ist, dass während der zweiten Iteration des Vorwärts-Snowballings einige Publikationen vorhanden waren, die über eintausend mal referenziert wurden. In diesen Fällen wurde die Menge der Publikationen mit Suchbegriffen eingeschränkt, die für die jeweilige Publikation geeignet sind. Lässt man also die zweite Iteration des Vorwärts-Snowballings außen vor, dann wurden während der systematischen Literatursuche insgesamt eine Menge von mindestens 16.906 Publikationen manuell durchsucht.

Aus dieser Suche ergab sich eine Menge von 101 Publikationen, die im Kontext dieser Arbeit als relevant identifiziert werden konnten, in Abbildung 11 grün hervorgehoben. Dabei stammen 43 der Publikationen aus der Datenbanksuche, während durch das Snowballing weitere 58 Publikationen hinzu kamen. Während des Snowballings wurden darüber hinaus diverse Duplikate der Publikationen aus der Datenbanksuche gefunden, die nicht in der Ergebnismenge berücksichtigt wurden. Die Ergebnisse der manuellen Suche wurden nicht in die Menge der relevanten Publikationen aufgenommen, da sie primär für die Definition des initialen Suchstrings verwendet werden sollten. Während des Verlaufs der systematischen Literatursuche wurden der Suchstring sowie die Inklusions- und Exklusionskriterien der Suchmethoden an verschiedenen Stellen verändert. Aus diesem Grund wurde die Entscheidung getroffen, die Ergebnisse der manuellen Suche nicht in die finale Menge relevanter Publikationen aufzunehmen. Eine komplette Liste der relevanten Publikationen ist in Anhang D beigelegt.

In den folgenden Kapiteln der Analyse wird grundsätzlich jede der 101 Publikationen berücksichtigt, jedoch ist es aufgrund der Menge nicht möglich, die Inhalte jeder Publikation zu erwähnen. Um einen besseren Überblick über die Ergebnismenge zu erhalten wurde nach der Durchführung der systematischen Literatursuche zunächst eine tabellarische Übersicht der Publikationen angefertigt, die in Anhang F beigelegt ist. In dieser Tabelle wurde dokumentiert, aus welchen Jahren die Publikationen stammen und welche Inhalte in den jeweiligen Publikation thematisiert wurden. Diese Tabelle dient dabei als Ausgangspunkt für die statistischen und inhaltlichen Auswertungen der Ergebnisse in den folgenden Kapiteln.

4.2 Effizienz und Effektivität

Während der Darstellung der einzelnen Suchmethoden wurden bereits einige der Ergebnisse, die aus den einzelnen Methoden hervorgingen, präsentiert. An dieser Stelle werden nun die Ergebnisse über die einzelnen Methoden hinaus miteinander verglichen. Abbildung 12 und 13 zeigen hierfür jeweils die Effektivität beziehungsweise die Effizienz der einzelnen Suchmethoden in dieser Arbeit sowie die von Brings et al. [2].

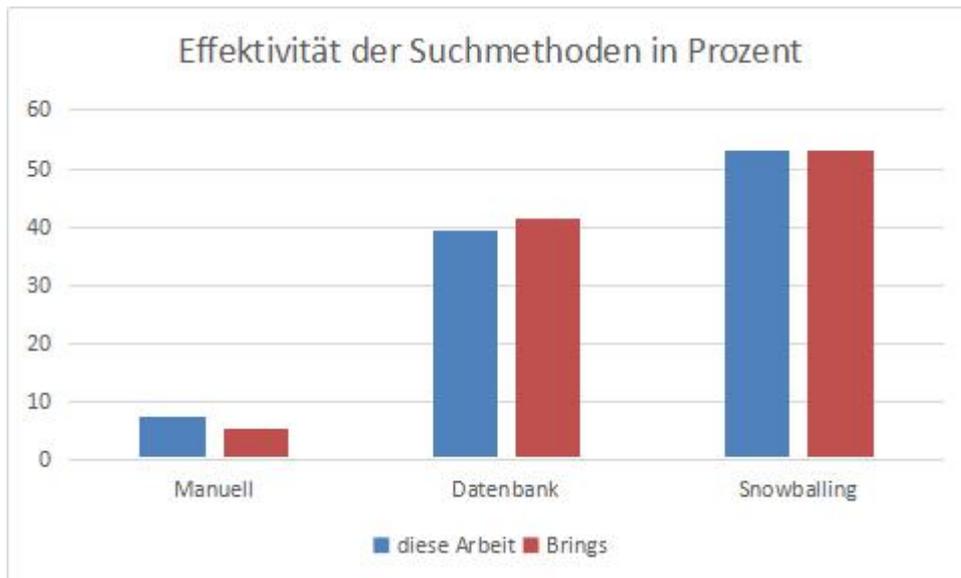


Abbildung 12: Vergleich der Effektivität der Suchmethoden mit Brings et al. [2]

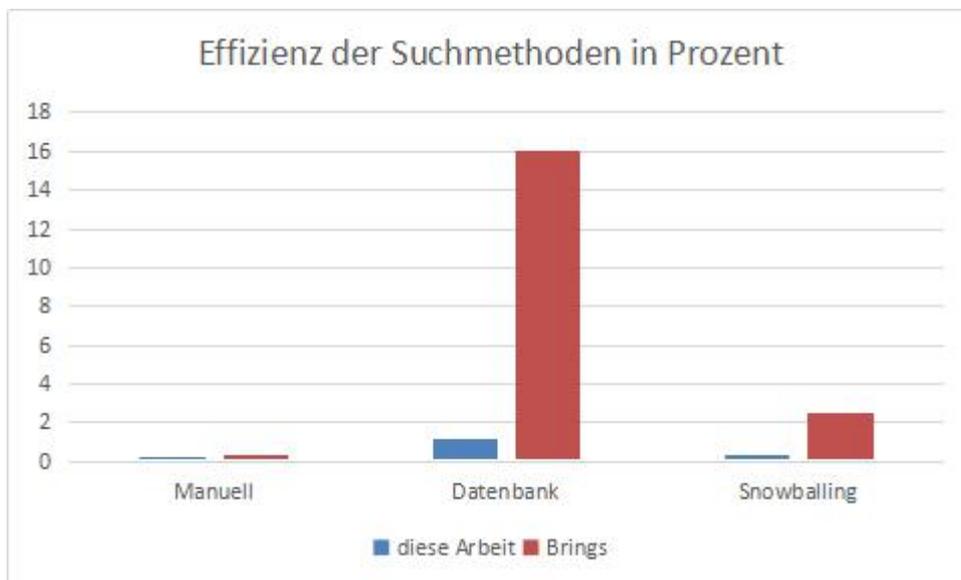


Abbildung 13: Vergleich der Effizienz der Suchmethoden mit Brings et al. [2]

Die Ergebnisse dieser Arbeit werden an dieser Stelle mit denen von Brings et al. [2] verglichen, da die Auswahl der Suchmethoden für die systematische Literatursuche in Kapitel 3.1 zu einem großen Teil anhand der Erkenntnisse von Brings et al. [2] gefolgert wurden. Durch diesen Vergleich kann entsprechend überprüft werden, wie sinnvoll die Auswahl der Suchmethoden für diese Arbeit war. Betrachtet man die Effektivität der Suchmethoden, so lässt sich feststellen, dass in dieser Arbeit eine Verteilung der Ergebnisse vorliegt, die sehr ähnlich zu der im Beispiel von Brings et al. [2] ist. Hierbei ist anzumerken, dass in dieser Arbeit das Snowballing auf den Ergebnissen der Datenbanksuche ausgeführt wurde und daher Ergebnisse, die bereits während der Datenbanksuche gefunden wurden, nicht zusätzlich in die Ergebnismenge des Snowballings aufgenommen wurden. Da Brings et al. [2] Schnittmengen von Ergebnissen erlaubt, kam es zu einer Summe der Effektivität von 111,9 Prozent. Um die Werte besser miteinander vergleichen zu können, wurden daher die Ergebnisse von Brings et al. [2] auf 100 Prozent normiert. Während die Effektivitäten für das Snowballing jeweils gerundet 53,2% betragen, verschieben sich die Anteile von manueller Suche und Snowballing jeweils nur um 2%. Dieser geringe Unterschied könnte aus verschiedenen Gründen hervorgerufen werden. Eine mögliche Erklärung wäre der Unterschied der betrachteten

Anzahl der Publikationen, da Brings et al. [2] beispielsweise nur 2665 Publikationen während der manuellen Suche betrachtet haben, im Vergleich zu den 8035 Publikationen, die in der manuellen Suche dieser Arbeit untersucht wurden. Ein anderer Grund könnte die Auswahl der betrachteten Journals und Konferenzen sein und die damit zusammenhängenden thematischen Schwerpunkte. Da sich das Beispiel von Brings et al. [2] thematisch von dem Inhalt dieser Arbeit unterscheidet, könnte die Betrachtung der gleichen Journals und Konferenzen daher verschieden große Ergebnismengen liefern. Nichts desto trotz lässt sich grundsätzlich feststellen, dass die Verteilung der Effektivität der Suchmethoden mit der von Brings et al. [2] übereinstimmt und somit Snowballing mit 53,2% die effektivste Suchmethode ist, gefolgt von der Datenbanksuche mit 39,4% Effektivität. Die manuelle Suche besitzt wie erwartet mit 7,3% den geringsten Anteil der Ergebnisse.

Vergleicht man die Effizienz der Suchmethoden dieser Arbeit mit denen von Brings et al. [2] lassen sich jedoch große Unterschiede feststellen. Jede der Suchmethoden in dieser Arbeit ist deutlich weniger effizient als die im Beispiel von Brings et al. [2]. Der vergleichsweise geringste Unterschied liegt hier bei der manuellen Suche vor, da hier die Effizienz mit beiden Werten unter einem Prozent grundsätzlich sehr gering ist. Der größte Unterschied liegt hier bei der Datenbanksuche vor. Die Effizienz der Datenbanksuche dieser Arbeit liegt hier gerundet bei 1,2%, während sie bei Brings et al. [2] 16% beträgt. Ein ähnliches Verhältnis liegt beim Snowballing mit 0,4% zu 2,5% vor. Dies lässt vermuten, dass die Suchkriterien, die Brings et al. [2] für ihr Beispiel gewählt haben, deutlich präziser gewesen sind und somit weniger irrelevante Suchergebnisse geliefert haben. Dies bezieht sich hierbei insbesondere auf den Suchstring. Die Unterschiede in der Präzision der Suchstrings lässt sich damit erklären, dass zu Beginn dieser Arbeit die relevanten Suchbegriffe für die Thematik dieser Arbeit noch nicht klar waren. Aus diesem Grund wurde der Suchstring für diese Arbeit etwas allgemeiner formuliert, um möglichst viele mögliche Ergebnisse zu berücksichtigen, worunter die Effizienz der Suchmethoden gelitten hat. Jedoch lässt sich auch bei der Effizienz feststellen, dass die grundlegenden Verhältnisse zwischen den Suchmethoden mit denen von Brings et al. [2] übereinstimmen. Die Datenbanksuche ist, wenn auch weniger deutlich, mit gerundet 1,2% die effizienteste Suchmethode, gefolgt vom Snowballing mit 0,4% und der manuellen Suche mit 0,1%. Diese Reihenfolge stimmt mit der von Brings et al. [2] überein.

Abgesehen von den Größenordnungen in Bezug auf die Effizienz der Suchmethoden lassen sich also diverse Parallelen in Bezug auf die Effektivität und die Effizienz zwischen den Erkenntnissen dieser Arbeit und den Erkenntnissen von Brings et al. [2] feststellen. Aus diesem Grund lässt sich schlussfolgern, dass die Auswahl der Suchmethoden für die systematische Literatursuche dieser Arbeit anhand der Erkenntnisse von Brings et al. [2] sinnvoll war.

4.3 Jahresverteilung

Während der manuellen Suche in Kapitel 3.2 musste eine Einschränkung bezüglich des betrachteten zeitlichen Rahmens vorgenommen werden. Dieser wurde dabei auf den Bereich von 2010 bis heute festgelegt. Daraufhin wurde in 3.2.1 versucht, den zeitlichen Rahmen mithilfe der Ergebnisse der manuellen Suche zu stützen, was aufgrund der geringen Menge der Ergebnisse nur sehr begrenzt möglich war. Um die Auswahl des Zeitrahmens der manuellen Suche zu unterstützen wird daher an dieser Stelle die Verteilung der gemeinsamen Menge der Ergebnisse von der Datenbanksuche und dem Snowballing betrachtet. Abbildung 14 zeigt hierfür die Verteilung der relevanten Publikationen entsprechend ihrer Publikationsjahre.

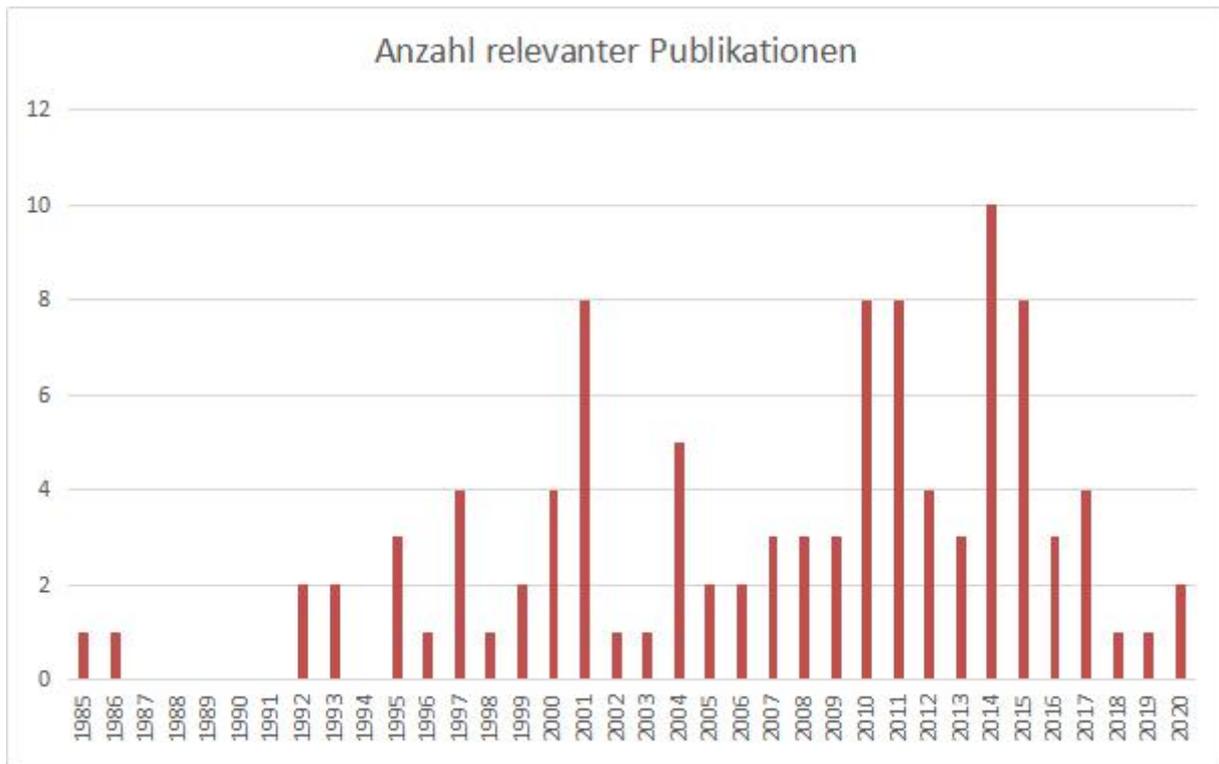


Abbildung 14: Verteilung der relevanten Publikationen anhand der Publikationsjahre

Da sowohl die Datenbanksuche als auch das Snowballing nicht auf einen bestimmten zeitlichen Rahmen begrenzt wurden, sind hierbei alle möglichen Publikationsjahre berücksichtigt worden. Anhand von Abbildung 14 ist zu erkennen, dass es zwei Intervalle gab, in denen Publikationen veröffentlicht wurden, die relevant für die Thematik „shared understanding“ sind. Das erste etwas weniger relevante Intervall befindet sich im Bereich zwischen 1997 und 2004 und hat seinen Peak mit acht Publikationen im Jahr 2001 und einer gesamten Menge von 26 Publikationen. Das zweite deutlich ausgeprägtere Intervall befindet sich zwischen 2010 und 2017. Es umfasst mit insgesamt 48 Publikationen fast die Hälfte aller gefundenen Publikationen und hat seinen Peak im Jahr 2014 mit zehn Publikationen. Da die zeitliche Einschränkung der manuellen Suche aufgrund des zeitlichen Rahmens dieser Arbeit leider nicht vermeidbar war, lässt sich anhand dieser Ergebnisse feststellen, dass der zeitliche Rahmen gut gewählt wurde, da er das wichtigste Intervall für die Thematik „shared understanding“ umfasst.

4.4 Mentale Modelle vs. Shared Understanding

Neben „shared understanding“ wurden während der systematischen Literatursuche aufgrund der thematischen Verwandtheit auch Publikationen zu mentalen Modellen berücksichtigt. Das Ziel dieser Entscheidung war es, auch diejenigen Publikationen zu finden, die zwar relevante Inhalte besitzen, jedoch nicht explizit den Begriff „shared understanding“ verwenden. Um den Erfolg dieser Maßnahme zu beurteilen zeigt Abbildung 15 den prozentualen Anteil der relevanten Publikationen, welche die Begriffe „shared understanding“ beziehungsweise mentale Modelle verwenden.

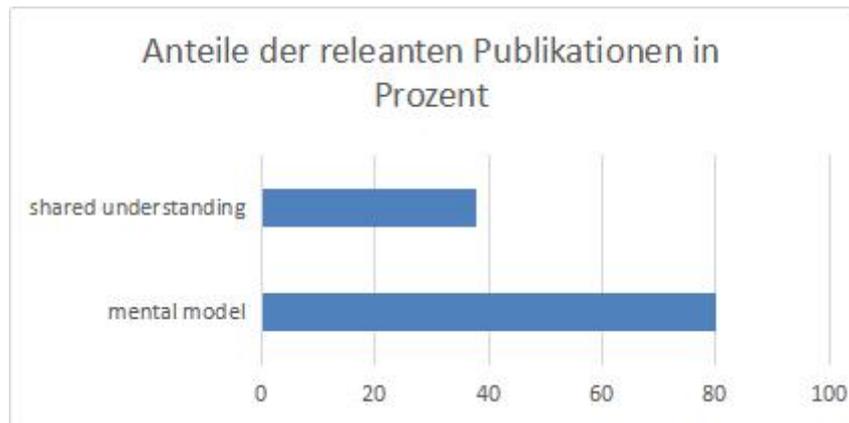


Abbildung 15: Vergleich der prozentualen Anteile der relevanten Publikationen mit den Themen „shared understanding“ beziehungsweise mentales Modell

Es ist zu erkennen, dass mit einem Anteil von etwa 80% der relevanten Publikationen der Begriff „mentales Modell“ deutlich häufiger verwendet wurde. Der Begriff „Shared understanding“ wurde im Gegensatz dazu mit einem Anteil von 38% weniger als halb so viel verwendet. Aufgrund dieses großen Unterschieds lässt sich feststellen, dass die Berücksichtigung von mentalen Modellen in der systematischen Literatursuche eine große Bereicherung für das Ergebnis der systematischen Literatursuche darstellt, da ohne diese Inklusion ein Großteil der relevanten Publikationen nicht gefunden worden wären.

Darüber hinaus lässt sich aus dieser deutlichen Diskrepanz in der Verwendung der Begriffe vermuten, dass sich der Begriff „shared understanding“ im Vergleich zu mentalen Modellen noch nicht etabliert hat. Ein weiteres Indiz hierfür ist, dass der Begriff häufig durch andere, sehr ähnliche Begriffe ersetzt wird. Beispiele hierfür wären die Begriffe „mutual understanding“ [42] oder „common understanding“ [84]. Da das Konzept von mentalen Modellen etabliert ist, wird in diesem Kontext auch oft von „shared mental models“ gesprochen, die dem Konzept von „shared understanding“ entsprechen [22,37,60].

4.5 Inhaltlich

In den vorherigen Kapiteln wurden verschiedene Erkenntnisse bezüglich des Prozesses der systematischen Literatursuche präsentiert, die anhand der Ergebnismenge relevanter Publikationen geschlussfolgert werden konnten. An dieser Stelle gilt es nun, die inhaltlichen Aspekte zu betrachten, die aus der systematischen Literatursuche gewonnen wurden. Diese Inhalte sollen dabei Ansätze liefern, die zum Erreichen des Ziels dieser Arbeit beitragen, also zum Definieren von Metriken, Methoden und Heuristiken zum Messen von „shared understanding“ in Vision Videos. Nach Betrachtung der Kategorisierungen der Ergebnismenge anhand der Tabelle in Anhang F konnten hierfür zwei übergeordnete Konzepte identifiziert werden, anhand derer die verschiedenen Ansätze klassifiziert werden können. Diese beiden Konzepte sind Mapping Konzepte und numerische Konzepte. Abhängig vom Ziel der Situation, in der diese Konzepte Anwendung finden, müssen sich diese Konzepte nicht zwangsläufig ausschließen und können teilweise kombiniert werden. In den folgenden Unterkapiteln werden sowohl die beiden Konzepte als auch einige der dazugehörigen konkreten Ansätze präsentiert. Die Auswahl der Ansätze wurde dabei anhand der Einschätzung getroffen, wie gut sie in den Kontext zur Beurteilung von „shared understanding“ durch Vision Videos überführt werden können.

4.5.1 Mapping Konzepte

Unter dem Mapping Konzept werden in dieser Arbeit all diejenigen Ansätze zusammengefasst, deren primäre Vorgehensweise darauf beruht, das mentale Modell von beteiligten Personen durch eine visuelle Darstellung zu externalisieren. Diese Externalisierungen haben zur Folge, dass das mentale Modell konkretisiert wird und dadurch mit anderen Personen kommuniziert werden kann. Auf Basis dieser visuellen Darstellung können dann weitere Analyseschritte durchgeführt werden, abhängig vom Ziel, das in der jeweiligen Situation verfolgt wird. Ein Beispiel hierfür wäre das Vergleichen der Externalisierungen von verschiedenen Personen, um Rückschlüsse bezüglich der Übereinstimmung ihrer mentalen Modelle zu schließen. Im Folgenden werden einige der Ansätze präsentiert, die verschiedene Auslegungen des Mapping Konzepts darstellen.

Cognitive Mapping

Der Begriff „cognitive Mapping“ hat in der Literatur zwei verschiedene Bedeutungen, abhängig vom Kontext, in dem der Begriff verwendet wird. Zum einen bezeichnet das „cognitive Mapping“ den Prozess, durch den Personen eine interne Repräsentation der Informationen bilden, die sie in ihrer Umgebung wahrnehmen [6]. Das Ergebnis hieraus wird dabei als „cognitive map“ bezeichnet. Anhand dieser „cognitive map“ können Person logische Schlussfolgerungen treffen und aufgrund von vergangenen Erfahrungen Situationen zu generalisieren, was dem Konzept eines mentalen Modells sehr ähnlich ist [6].

Die zweite Auslegung des Begriffs „cognitive Mapping“ bezieht sich im Allgemeinen auf eine Menge von Techniken, die eine externe Repräsentation der subjektiven Überzeugungen beziehungsweise des mentalen Modells im Rahmen der Problemdomäne erstellen [7,8,9,10,11]. Die genaue Definition des Begriffs ist dabei in jedem Anwendungsfall leicht unterschiedlich. In den meisten Fällen beinhaltet die Definition eine Form von graphischer Darstellung der „cognitive map“ [7,8,9,11]. Die graphisch dargestellten „cognitive maps“ bestehen dabei essentiell aus Knoten, die Konzepte repräsentieren und Relationen zwischen diesen Konzepten, ähnlich einer Ontologie [9]. Für diese Arbeit wird hauptsächlich die zweite Definition von „cognitive Mapping“ berücksichtigt, da sich diese auf die Techniken bezieht, die eine externe Repräsentation erzeugen. Dies passt besser zum Ziel dieser Arbeit, das sich ebenfalls auf die Identifikation von Methoden, Metriken und Heuristiken bezieht. In diesem Zusammenhang bezieht sich der Begriff „cognitive map“ auf die externe Repräsentation von mentalen Modellen.

Causal Mapping

„Causal Mapping“ stellt eine spezifischere Variante des „cognitive Mappings“ dar, welche die kausalen Zusammenhänge zwischen Konzepten aufgrund subjektiver Überzeugungen untersucht [9,13,14,15,16]. Entsprechend liegt der Fokus auf den Relationen zwischen den Konzepten, die hierbei die Richtung der kausalen Zusammenhänge definieren. Abbildung 16 zeigt ein Beispiel einer „causal map“ von Siau und Tan [9], welche die Zusammenhänge eines Kurses für objektorientierte Systemanalyse und Design aus Sicht eines Studenten darstellt.

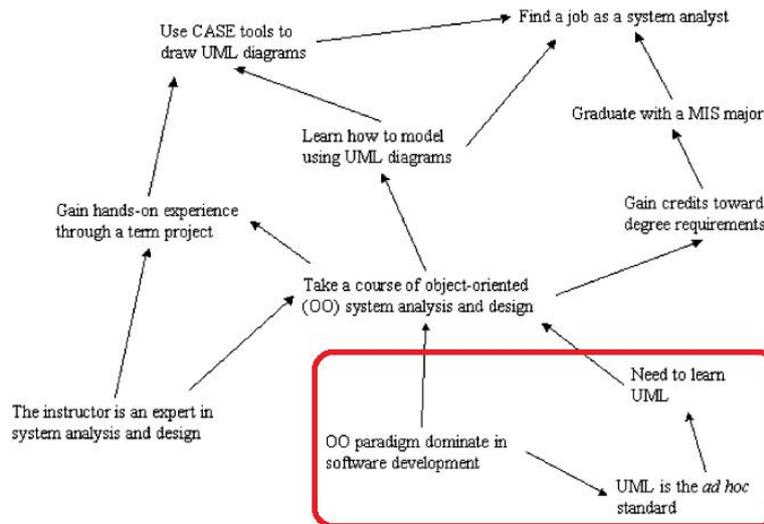


Abbildung 16: Beispiel einer „causal map“ von Siau und Tan [9]

Die Konzepte wurden hierbei durch kurze, deskriptive Phrasen repräsentiert, deren kausale Abhängigkeiten durch Pfeile dargestellt wurden. Ein einfaches Beispiel für einen kausalen Zusammenhang, der hier dargestellt wird, ist, dass UML ein ad hoc Standard ist, also muss die betroffene Person UML lernen. Diese einfachen kausalen Zusammenhängen können darüber hinaus erweitert werden, indem den Relationen zugeordnet wird, wie stark der Zusammenhang der betroffenen Konzepte ist [14,15,16]. Dies kann sowohl durch eine Ordinalskala (z.B. wenig, mittel, stark) als auch durch eine beliebige numerische Skala (z.B. Wert im Intervall [0,10]) ausgedrückt werden.

Neben der graphischen Darstellung, anhand derer „causal maps“ verschiedener Personen qualitativ miteinander verglichen werden können, lässt sich anhand der Stärke der kausalen Zusammenhänge der Unterschied zwischen „causal maps“ auch numerisch bestimmen. Hierfür kann eine $n \times n$ Adjazenzmatrix aufgestellt werden, bei der n die Anzahl der Konzepte in der „causal map“ ist [14]. Der Wert in den Zellen repräsentiert dabei dann die Existenz (wenn der Wert ungleich 0 ist), die Richtung (durch das Vorzeichen bestimmt) und die Stärke der Relation der beteiligten Konzepte. Abbildung 17 zeigt zwei Beispiele einer solchen Adjazenzmatrix, abgeleitet aus dem Beispiel von Siau und Tan [9] aus Abbildung 16. Das in den Adjazenzmatrizen präsentierte Beispiel wird dabei auf den in Abbildung 16 rot markierten Bereich beschränkt, um das Beispiel übersichtlicher zu gestalten.

Von\Nach	OO dominate	ad hoc	learn UML	Von\Nach	OO dominate	ad hoc	learn UML
OO dominate	0	3	0	OO dominate	0	1	0
ad hoc	-3	0	10	ad hoc	-1	0	8
learn UML	0	-10	0	learn UML	0	-8	0

Abbildung 17: Beispielhafte Adjazenzmatrizen für einen Ausschnitt aus Abbildung 16

Die beiden Varianten dieser Adjazenzmatrix repräsentieren hierbei die „causal maps“ von zwei verschiedenen Studenten, deren Einschätzungen der kausalen Zusammenhänge leichte Unterschiede aufweisen. Beispielhaft lässt sich aus der linken Adjazenzmatrix ablesen, dass von „ad hoc“ nach „learn UML“ mit einem Wert von 10 ein extrem starker kausaler Zusammenhang besteht. Die entgegengesetzte Relation von „learn UML“ nach „ad hoc“ wird dabei mit -10 bewertet, um die entgegengesetzte Richtung der Relation darzustellen.

Die Konvergenz dieser „causal maps“ kann beispielsweise mit der „distance ratio“ Formel bestimmt werden [16]. Diese Formel bildet die Summe aller absoluten Werte der Differenzen

zwischen den zusammengehörigen Wertepaaren der verschiedenen Matrizen und teilt diese Summe durch die maximal mögliche Differenz der beiden Matrizen. Die mathematische Formel hierfür lautet wie folgt:

$$dist = \frac{\sum_{i \in I} \sum_{j \in J} |a_{i,j} - b_{i,j}|}{I * J * y_{\max}}$$

In dieser Formel stehen I und J jeweils für die Anzahl der Zeilen beziehungsweise der Spalten der Matrizen, $a_{i,j}$ und $b_{i,j}$ für die Werte der jeweiligen Zellen der betrachteten Matrizen und y_{\max} für die maximale Differenz, welche zwischen den zwei Zellen der Matrizen möglich ist. Die maximale Differenz entspricht dabei der Differenz zwischen der Obergrenze und der Untergrenze des Wertebereichs, der für die Bewertung der kausalen Zusammenhänge verwendet wird. Das Ergebnis der Formel liegt dabei im Intervall [0,1], wobei ein Wert von 0 einer perfekten Übereinstimmung und ein Wert von 1 komplette Uneinigkeit der „causal maps“ entspricht [16]. Dies ist dabei unabhängig vom Wertebereich, der für die Bewertung der Stärke des kausalen Zusammenhangs gewählt wurde, da der Wert durch das Teilen durch die maximal mögliche Differenz normiert wird. Diese Metrik könnte dementsprechend als Ansatz zur Messung vom „shared understanding“ zwischen zwei Personen genutzt werden. Im oben präsentierten Beispiel läge der Wert der „distance ratio“ Formel bei gerundet 0.04, was einer sehr hohen Übereinstimmung der „causal maps“ entspricht und somit ein vorhandenes „shared understanding“ vermuten lässt.

Concept Mapping

Ähnlich wie das „causal Mapping“ ist auch das „concept Mapping“ eine erweiterte graphische Variante des „cognitive Mappings“. Die meist verbreitetste Variante des „concept Mappings“, welche auch in dieser Arbeit betrachtet wird, wurde dabei ursprünglich von Novak [18] entwickelt [11]. Hierbei werden die Relationen zwischen den Konzepten mit Bezeichnern versehen, welche die Art der Relation spezifizieren und die Konzepte in eine hierarchische Ordnung bringen [9,11,16,17,19,20,21,22,24]. Die Relationen werden dabei durch Pfeile dargestellt, die entweder in eine, beide oder keine Richtung zeigen [9]. Die Bezeichner der Pfeile unterliegen dabei keinen speziellen Restriktionen. Abbildung 18 zeigt ein Beispiel einer „concept map“ von Siau und Tan [9], die das Verständnis eines Nutzers bezüglich verschiedener Diagrammtypen in UML modelliert.

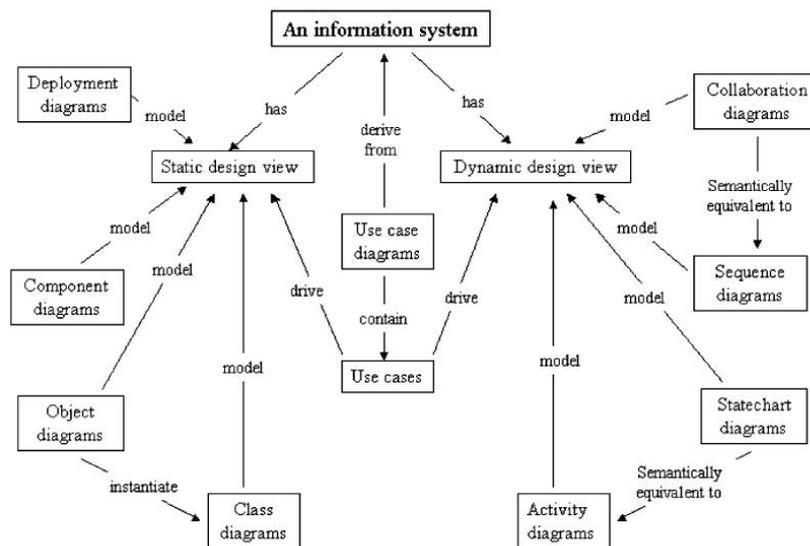


Abbildung 18: Beispiel einer „concept map“ von Siau und Tan [9]

Die Relationen können dabei, ähnlich zum „causal Mapping“, auch kausale Zusammenhänge ausdrücken [9]. Ein konkretes Beispiel aus Abbildung 18 hierfür wäre, dass ein Informationssystem laut dieser „concept map“ zwingend eine statische Designview besitzen muss.

Diese Visualisierung kann als Diskussionsbasis dienen, um Unterschiede in den mentalen Modellen beteiligter Personen und somit den Grad des „shared understandings“ qualitativ zu identifizieren und zu verbessern.

Eine Methode des „concept Mappings“ sieht vor, dass vor der Erhebung der „concept map“ die im jeweiligen Kontext relevanten Konzepte aus den entsprechenden Dokumenten, wie beispielsweise der Spezifikation, extrahiert werden [17]. Diese werden den beteiligten Personen dann als Liste mit der Aufgabe übergeben, sie mithilfe von Relationen entsprechend ihrer Vorstellung in einer „concept map“ anzuordnen [16,17,21]. Eine noch eingeschränktere Variante sieht vor, auch die grundlegende hierarchische Struktur mit Leerstellen für die Konzepte vorzugeben [23]. Dies hat zur Folge, dass die „concept maps“ der verschiedenen Personen vergleichbarer werden, da sie alle die gleichen grundlegenden Konzepte beinhalten, die für den jeweiligen Kontext relevant sind. Um einen noch größeren Nutzen aus diesem Prozess zu ziehen, kann die Methode mit anderen Interview-Techniken wie beispielsweise Think Aloud kombiniert werden, um einen genaueren Einblick in den Entscheidungsprozess zu gewinnen [20].

Eine mögliche Auswertung der Ergebnisse besteht darin, die einzelnen „concept maps“ zu einer gemeinsamen „concept map“ zu aggregieren [17]. Anhand dieser aggregierten „concept map“ kann dann das „shared understanding“ der Personengruppe analysiert werden, beispielsweise durch das Identifizieren von Übereinstimmungen und Konflikten. Numerischen lassen sich diese Konzepte durch den prozentualen Anteil übereinstimmender oder ähnlich angeordneter Konzepte quantifizieren [21]. Ebenso kann hierfür auch die Anzahl der Relationen gemessen werden, die in verschiedenen „concept maps“ identisch sind [23].

Eine weitere Möglichkeit „shared understanding“ anhand von „concept maps“ qualitativ zu messen besteht darin, bereits erstellte „concept maps“ erneut zu überarbeiten, nachdem einige Zeit im Projekt vergangen ist [19]. Durch den Vergleich der Übereinstimmung der „concept maps“ vom vorherigen Zeitpunkt mit der Übereinstimmung zu einem späteren Zeitpunkt kann geprüft werden, ob sich das „shared understanding“ im Projektverlauf verbessert hat.

Neben dem hier beschriebenen „concept Mapping“ gibt es noch andere Varianten, die an das Prinzip vom „concept Mapping“ angelehnt sind. Eine dieser Varianten von Green et al. [32] sieht die Verwendung einer abgewandelten Version des Entity-Relationship Modellings vor, welches die Konzepte durch Entities ersetzt, welche Attribute besitzen und durch Relationen miteinander verbunden sind. Auch wenn dieser Ansatz im Grunde sehr vielversprechend klingt, würde die Verwendung dieses Konzepts jedoch voraussetzen, dass die beteiligten Personen die Syntax von Entity-Relationship Diagrammen beherrschen. Da im Kontext der Beurteilung von „shared understanding“ durch Vision Videos jedoch auch Stakeholder beteiligt sind, die kein Domänenwissen bezüglich Software Engineering besitzen, kann nicht vorausgesetzt werden, dass diese Personen das Konzept in einem annehmbaren Zeitrahmen erlernen. Aus diesem Grund werden daher Ansätze, die Wissen über komplexe Konzepte wie beispielsweise Entity-Relationship Diagramme voraussetzen, nicht weiter untersucht.

Pathfinder

Die Pathfinder Methode ist eine häufig verwendete Form des „causal Mappings“, welche Konzepte als Knoten darstellt, die durch gewichtete Relationen miteinander verbunden sind [7,8,14,20,22,33,35,43,44,45,46]. Das Ziel dieser Methode ist es, die kognitiven Strukturen beziehungsweise mentalen Modelle von Personen zu analysieren und zu vergleichen [7,8]. Je ähnlicher sich dabei zwei Konzepte sind, desto näher sind sich diese Konzepte im Graph, dargestellt durch eine geringere Gewichtung ihrer Relation [22,45]. Die Relationen können dabei sowohl gerichtet als auch ungerichtet sein [43]. Die Gewichtung dieser Relationen wird dabei von den Probanden durchgeführt, indem jeweils die Verwandtschaft von Konzeptpaaren bewertet wird, was allgemein als „paired comparison rating“ oder auch „pairwise rating“ bezeichnet wird [7,8,14,33]. Die Bewertung wird dabei meistens auf Basis einer ganzzahligen Skala vorgenommen, wobei ein höherer Wert eine größere Verwandtschaft bedeutet [45]. Da die Gewichte der Relationen im Graph jedoch bei geringeren Werten größere Ähnlichkeit ausdrücken sollen, wird hierfür die Differenz aus dem maximalem Wert der Skala und der jeweiligen Bewertung der Relation verwendet [45]. Nachdem die „paired comparison ratings“ auf die Relationen des Graphen übertragen wurden, ist das Ergebnis dieses Schrittes ein vollständig verbundener Graph aller Konzepte mit allen möglichen Relationen [45]. Im nächsten Schritt werden die kürzesten Pfade zwischen allen möglichen Konzepten bestimmt. Die Länge eines Pfades ergibt sich dabei aus der Summe der Gewichte der Relationen, die in dem Pfad enthalten sind [14,46]. Anschließend werden dann alle Relationen entfernt, die in keinem der kürzesten Pfade enthalten sind [45]. Dies erlaubt der Pathfinder Methode nur die essentiellen, relevanten Relationen zu berücksichtigen [45]. Das Ergebnis dieses Schrittes ist das fertige Pathfinder Netzwerk, auch PFNet genannt [45]. Eine alternative Metrik für die Länge des kürzesten Pfades kann hierbei abhängig vom Kontext der Nutzung auch das Maximum der Gewichte der Relationen sein, die in dem Pfad enthalten sind [45]. Optional kann für die Bestimmung des kürzesten Pfades auch eine maximale Anzahl an Relationen festgelegt werden, die in dem Pfad enthalten sein dürfen [45]. Abbildung 19 zeigt ein Beispiel von Braunschweig [45], bei dem ein vollständig verbundener Graph durch die Betrachtung der kürzesten Pfade zu einem Pathfinder Netzwerk reduziert wird.

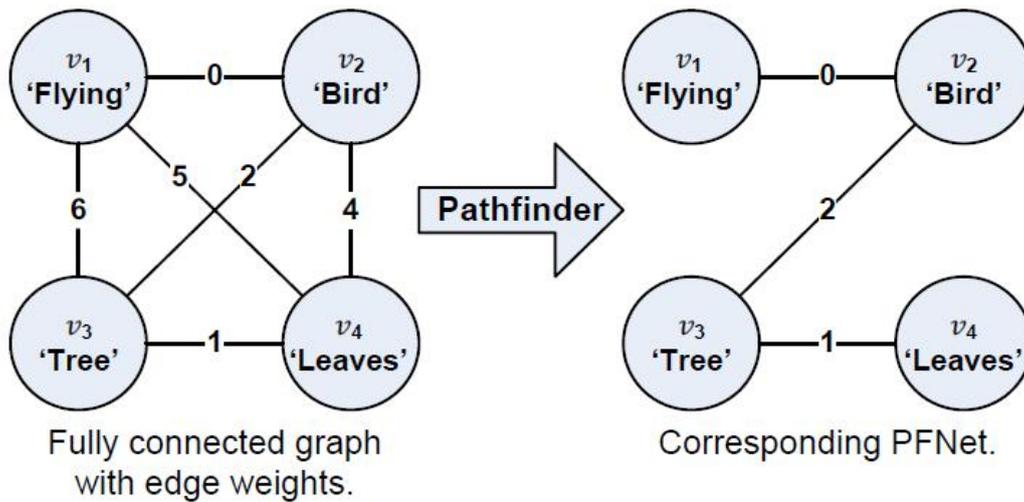


Abbildung 19: Beispiel einer Reduktion eines vollständig verbundenen Graphen zu einem Pathfinder Netzwerk von Braunschweig [45]

Ein Beispiel für die Reduktion einer Relation ist hierbei die Verbindung zwischen v_1 und v_3 mit einem Wert von 6. Da der kürzeste Pfad zwischen v_1 und v_3 mit einer Länge von 2 über v_2 läuft, wird die direkte Relation überflüssig und daher reduziert. Im dem Fall, dass mehrere kürzeste Pfade zwischen zwei Konzepten existieren, werden die Relationen aller Pfade erhalten [45].

Um die Pathfinder Netzwerke verschiedener Personen vergleichen zu können, wurde die Ähnlichkeitsfunktion NETSIM („pathfinder NETwork SIMilarity“) definiert [14]. NETSIM entspricht dabei der Anzahl gemeinsamer Relationen zwischen den Netzwerken, geteilt durch die gesamte Anzahl aller Relationen die sich in beiden Netzwerken befinden, ohne Duplikate zu berücksichtigen [14,46]. Eine mögliche mathematische Formulierung hierfür lautet dabei wie folgt:

$$NETSIM = \left(\frac{X}{T - X} \right) \quad [46]$$

Hierbei ist X die Anzahl aller gemeinsamen Relationen zwischen zwei Netzwerken und T die Summe der Anzahl aller Relationen in beiden Netzwerken darstellt, wobei hier Duplikate berücksichtigt werden. Der Wert von NETSIM liegt dabei im Intervall $[0,1]$, wobei ein Wert von 1 für komplett identische Netzwerke steht, während ein Wert von 0 bedeutet, dass bei den Netzwerken keine einzige Relation übereinstimmt [14]. NETSIM kann also als prozentualer Anteil der übereinstimmenden Relationen interpretiert werden [33]. Abbildung 20 von Lim und Klein [46] zeigt hierfür ein einfaches Beispiel zweier Pathfinder Netzwerke, anhand derer die Berechnung veranschaulicht werden kann.

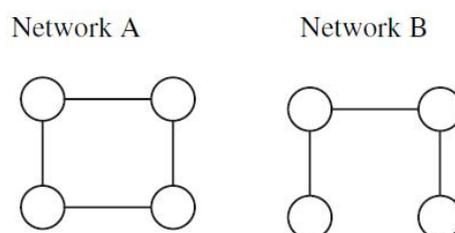


Abbildung 20: Vergleich zweier Pathfinder Netzwerke von Lim und Klein [46]

Netzwerk A und B haben hierbei drei Relationen, die miteinander übereinstimmen. Durch Verwendung der Formel für NETSIM ergibt sich entsprechend ein Wert von $(3/(7-3)) = 0.75$. Dieser Wert von 0.75 würde also beispielsweise einer 75% Übereinstimmung der Relationen zwischen den beiden Netzwerken entsprechen. Um die Ähnlichkeit eines gesamten Teams zu berechnen, kann entsprechend der durchschnittliche NETSIM-Wert zwischen allen Teammitgliedern berechnet werden [14,33].

Da die Pathfinder Methode kognitive Strukturen, also mit anderen Worten mentale Modelle, von Personen analysieren und vergleichen soll, kann die NETSIM-Metrik für die Beurteilung von „shared understanding“ zwischen zwei Personen verwendet werden. Der durchschnittliche NETSIM-Wert zwischen allen Teammitgliedern kann dann entsprechend als Metrik für das „shared understanding“ innerhalb des gesamten Teams interpretiert werden. Ähnliche Rückschlüsse werden von Lim und Klein [46] gezogen, die jedoch von geteilten mentalen Modellen anstelle von „shared understanding“ sprechen.

Eine alternative Metrik setzt den Vergleich der Relationen in den Kontext der Knoten, zu denen die Relationen gehören [45]. Hierfür wird die Anzahl der gemeinsamen Relationen eines Knotens zwischen den Netzwerken durch die Menge aller Relationen geteilt, die der Knoten in beiden Netzwerken besitzt [45]. Aus der resultierende Ähnlichkeit der Knoten wird danach der Durchschnitt über alle Knoten gebildet, um die Ähnlichkeit der Netzwerke zu bestimmen. Der Wertebereich und die Bedeutung der Werte dieser Metrik ist dabei identisch zu denen von NETSIM. Entsprechend kann auch diese Metrik zur Beurteilung des „shared understandings“ der betroffenen Personen verwendet werden.

Endlicher Zustandsautomat

Einige der relevanten Publikationen beschreiben die Nutzung von endlichen Zustandsautomaten im Kontext von mentalen Modellen [25,26,27]. Als (deterministischer) endlicher Zustandsautomat (DEA) wird hierbei ein formales, mathematisches Konstrukt bezeichnet, was in seiner einfachsten Form durch ein 5-Tupel dargestellt werden kann:

$$DEA = (Q, \Sigma, \delta, q_0, F) \text{ [83]}$$

Q bezeichnet hierbei die endliche Zustandsmenge, Σ ist das endliche Eingabealphabet, δ bezeichnet die Übergangsfunktion des endlichen Automaten, q_0 bezeichnet den Startzustand mit $q_0 \in Q$ und F ist die Endzustandsmenge mit $F \subseteq Q$ [83]. Beispielhaft wird im Folgenden näher auf die Umsetzung von Rushby [26] eingegangen, da diese sich durch die Berücksichtigung vom realem Verhalten des Systems und dem mentalen Modell des Nutzers am besten auf den Kontext dieser Arbeit übertragen lässt.

Die grundsätzliche Idee von Rushby [26] ist es, sowohl das reale Verhalten des Systems als auch das mentale Model eines Nutzers bezüglich des Verhaltens als endlichen Zustandsautomaten zu modellieren, um diese miteinander vergleichen zu können und mögliche Abweichungen zu identifizieren. Der entscheidende Vorteil dieser Methode liegt darin, dass endliche Zustandsautomaten eine endliche Menge an Zuständen besitzen [26]. Dadurch lassen sich eine Menge formaler Methoden anwenden, die alle möglichen Verhaltensweisen der endlichen Zustandsautomaten überprüfen [26]. Diese fasst Rushby [26] unter dem Begriff „model checking“ zusammen. Ein weiterer Vorteil von endlichen Zustandsautomaten ist hierbei, dass beim Auftreten eines Unterschieds zwischen zwei Ausprägungen eine Sequenz von Eingaben angegeben werden kann, die zu dem Unterschied geführt haben, da die Zustandsübergänge klar definiert sind [26,27]. Diese Informationen können laut Rushby [26] behilflich dabei sein, entweder das Design des Systems näher an das

mentale Modell des Nutzers anzupassen oder im Gegensatz dazu die Bildung eines präziseren mentalen Modells des Nutzers bezüglich des Systems zu unterstützen.

Als Problem dieser Methode identifiziert Rushby [26], dass reale Systeme oft entweder keine endliche Zustandsmenge besitzen oder die Zustandsmenge zu groß ist, um sie umfangreich zu analysieren. Der Lösungsansatz hierfür liegt dabei darin, dass reale System zu abstrahieren und zu generalisieren, da in den meisten Fällen nicht alle Details des Systems gebraucht werden, um das Ziel dieser Methode zu erreichen.

Das von Rushby [26] beschriebene Vorgehen sollte sich dabei ohne Probleme auf die Modellierung von zwei mentalen Modellen anstelle des Systemverhaltens übertragen lassen. Dadurch wäre es möglich, Übereinstimmungen und Unterschiede zwischen den mentalen Modellen von zwei Nutzern anstatt zwischen Nutzer und realem System zu identifizieren. Dies hat zur Folge, dass das Ergebnis der Methode wiederum für die qualitative Beurteilung von „shared understanding“ zwischen den beteiligten Nutzern verwendet werden kann und somit einen möglichen Ansatz für das Ziel dieser Arbeit darstellt. Für eine quantitative Beurteilung des „shared understandings“ könnte darüber hinaus der Anteil aller möglichen Pfade vom Startzustand zu einem der Endzustände bestimmt werden, der zwischen verschiedenen Nutzern übereinstimmt.

Als ein sehr ähnliches Konzept zu endlichen Zustandsautomaten werden auch Ansätze zu Petri-Netzen in den relevanten Publikationen präsentiert [25,28]. Da die Methoden, die im Zusammenhang mit Petri-Netzen präsentiert werden, sich jedoch grundlegend mit der hier beschriebenen Methode zur Nutzung endlicher Zustandsautomaten decken, werden die Methoden zu Petri-Netze nicht näher beleuchtet.

Persona Ontologien

Sim und Brouse [29,30] präsentieren in ihren aufeinander aufbauenden Arbeiten das Konzept von Persona Ontologien. Das Ziel dieses Ansatzes ist es unter anderem festzustellen, wie das Konzept von Personas verwendet werden kann, um das Verständnis von Stakeholdern und Entwicklern bezüglich der Bedürfnisse und des Verhalten von Nutzern zu verbessern [30]. Darüber hinaus soll festgestellt werden, wie mit Personas die Konzepte und Relationen von Nutzern spezifiziert werden können, um ein „shared common understanding“ unter den Stakeholdern und Entwicklern zu erschaffen [30].

Hierbei werden Personas als fiktive, spezifische und konkrete Repräsentation eines Zielnutzers definiert, die echten Personen ähnlich sein soll [30]. Zu diesem Zweck besitzen Personas konkrete persönliche Informationen, wie beispielsweise Name, Alter, Beruf, Bildungsgrad, soziale Umgebung, Ziele und Fähigkeiten [30]. Eine Ontologie auf der anderen Seite wird als eine formale, explizite Spezifikation einer geteilten Konzeptualisierung definiert, welche unter anderem das Vokabular, die essenziellen Konzepte, die Hierarchien und die Relationen zwischen Konzepten einer Domäne enthält [30].

Das von Sim und Brouse [29,30] vorgestellte Konzept ähnelt sehr dem des „concept Mappings“, da im Grunde die gleichen Bestandteile dargestellt werden. Der entscheidende Unterschied zwischen den beiden Konzepten liegt jedoch darin, worauf der Betrachtungsfokus des jeweiligen Konzepts liegt. Während sich das „concept Mapping“ auf die Darstellung des Systems fokussiert, werden bei den Persona Ontologien zunächst Personas konstruiert, die im Kontext des betrachteten Systems relevant sind. Hierfür werden im ersten Schritt des Prozesses Informationen zum System gesammelt und in Personas überführt [30]. Aus diesen Personas werden dann die Begriffe extrahiert, die essentielle Eigenschaften oder Klassen der Ontologie des Systems darstellen [30]. Diese werden dann,

nach einem eventuellen Vergleich mit bereits existierenden, verwandten Ontologien, in eine hierarchische Struktur gebracht [30]. Nachdem die Klassen durch die Erstellung von beispielhaften Instanzen auf Fehler überprüft wurden, können dann verschiedene Ontologien miteinander kombiniert werden [30]. Dieser Prozess hat dabei abermals Ähnlichkeiten zu „concept maps“, welche ebenfalls miteinander aggregiert werden können. In diesem Fall können zwei Persona Ontologien miteinander kombiniert werden, wenn sie Klassen besitzen, die durch eine oder mehrere Relationen miteinander verbunden sind [30]. Das Ergebnis dieser Methode ist eine Persona Ontologie, die sowohl die Eigenschaften der Zielnutzer und deren Zusammenhänge präsentiert, als auch die Relationen zwischen den Zielnutzern und ihrer Umgebung darstellt, welche im Fall der Softwareentwicklung durch ein Softwaresystem repräsentiert wird [30].

Im Kontext dieser Arbeit kann die Erstellung solcher Persona Ontologien durch beteiligte Personen genutzt werden, um ihr mentales Model zu repräsentieren. Durch den Vergleich der erstellten Persona Ontologien können dann, wie bei einigen bereits vorher beschriebenen Methoden, Rückschlüsse auf das „shared understanding“ zwischen den Personen getroffen werden. Der entscheidende Vorteil dieser Methode im Vergleich zu den bereits beschriebenen ähnlichen Methoden, wie beispielsweise „concept Mapping“, liegt darin, dass Personas ein Konzept darstellen, das größtenteils unabhängig von Wissen spezifischer Domänen in natürlicher Sprache beschrieben werden kann [87]. Dies hat zur Folge, dass Stakeholder, die aus keiner technisch veranlagten Domäne stammen, ohne weiteres benötigtes Wissen genauso anwenden können, wie beispielsweise Entwickler. Dies ist besonders relevant im Kontext von Vision Videos, da diese „shared understanding“ gleichermaßen zwischen allen Stakeholdern und Entwicklern schaffen sollen.

Sortiertechniken (Card Sorting)

Sortiertechniken stellen eine Gruppe von Methoden dar, bei der eine Menge von Konzepten entsprechend einer Aufgabe oder eines Themas kategorisiert beziehungsweise geordnet werden sollen [14,19,21,22]. Die Methoden sind dabei im Allgemeinen eine sehr flexible, schnelle und einfach zu verwaltende Möglichkeit, um Informationen über das mentale Modell von Personen zu sammeln [14,21,22]. Die in der Menge von relevanten Publikationen am meisten aufgetretene Variante hiervon ist das Card Sorting, auf das im Folgenden genauer eingegangen wird [14,21,22]. In der einfachsten Form werden hierbei die Konzepte auf physische Karten geschrieben und entsprechend ihrer Kategorisierung auf Stapeln geordnet [21,22]. Card Sorting kann dabei im Allgemeinen in zwei Varianten unterteilt werden, offenes und geschlossenes Card Sorting [21,22]. Offenes Card Sorting ist hierbei konzeptionell eine sehr unstrukturierte Variante, bei welcher die betroffene Person selbst die Konzepte sowie die Kategorien definiert, in welche die Konzepte eingeordnet werden [21,22]. Geschlossenes Card Sorting ist im Gegensatz dazu eine sehr strukturierte Variante, bei der sowohl die Konzepte als auch die Kategorien vordefiniert sind [21,22]. Der Übergang dieser beiden Varianten ist dabei fließend und kann somit flexibel auf die Situation angepasst werden, in der das Card Sorting verwendet wird [21].

Die Art der Ergebnisse, die aus dem Card Sorting gewonnen werden können, ist abhängig davon, wie offen oder geschlossen das Card Sorting umgesetzt wird. Wenn das Ziel darin besteht, ein qualitatives Ergebnis durch das Card Sorting zu erzielen, ist eine offenere Variante sinnvoller, da diese dem Nutzer weniger Einschränkungen auferlegt und somit die Repräsentation des mentalen Modells erleichtert [21]. Dieser Prozess kann dabei durch Interviews unterstützt werden, in denen die Auswahl der Konzepte und Kategorien sowie die jeweilige Zuordnung erklärt wird [19]. Sollen die Ergebnisse jedoch quantitativ sein,

empfiehlt sich die Nutzung eines eher geschlossenen Card Sortings [21]. Durch die Vorgabe von Konzepten und Kategorien lassen sich Metriken einfacher definieren, mit denen die Ergebnisse zwischen Personen direkt vergleichbar sind [21]. Ein Nachteil hiervon ist jedoch, dass durch das Vordefinieren von Konzepten und Kategorien nur ein sehr selektiver Einblick in das mentale Modell der betroffenen Personen ermöglicht wird [23]. Dieser kann dabei stark durch die Person beeinflusst werden, welche die Konzepte und Kategorien definiert [23].

Eine beispielhafte Methode, bei der ein quantitatives Ergebnis aus Card Sorting generiert wird, wurde von Mohammed et al. [23] beschrieben. Bei dieser Methode werden alle Paare von Konzepten jeweils mit 1 oder 0 bewertet, abhängig davon, ob sie gleich kategorisiert worden sind oder nicht. Die daraus resultierende Sequenz von Einsen und Nullen kann daraufhin mit den Sequenzen anderer Personen verglichen werden, indem die Summe der absoluten Differenzen der korrespondierenden Werte gebildet wird, die daraufhin durch die Anzahl der Paare von Konzepten geteilt wird. Der daraus resultierende Wert im Intervall von $[0,1]$ repräsentiert die Ähnlichkeit der Kategorisierungen der betroffenen Personen, wobei ein Wert von 0 für komplette Übereinstimmung und ein Wert von 1 für komplette Uneinigkeit steht. Da die Kategorisierung eine Repräsentation der mentalen Modelle darstellt, kann diese Metrik als Indiz für die Beurteilung von „shared understanding“ verwendet werden.

Ein Problem, was bei genauerer Betrachtung der von Mohammed et al. [23] beschriebenen Methode jedoch auffällig wird, ist, dass der Aufwand der Umsetzung der Methode aufgrund der Betrachtung aller möglichen Paare von Konzepten exponentiell steigt. Da dies die Nutzung dieser Methode für sehr komplexe Systeme einschränken könnte, wird an dieser Stelle eine vereinfachte Variante präsentiert, die mit linearem Aufwand umgesetzt werden kann. Hierfür werden anstatt der Paare von Konzepten die Kategorisierungen der Konzepte betrachtet. Beim Vergleich verschiedener Personen werden dabei die Kategorisierungen der einzelnen Konzepte miteinander verglichen. Bei Übereinstimmung der Kategorisierung beider Personen wird das Konzept mit einer 1 bewertet, sonst mit einer 0. Die aus dem Vergleich entstehende Sequenz kann, wie bei der Methode von Mohammed et al. [23], dann aufsummiert und durch die Anzahl der Konzepte geteilt werden. Der daraus resultierende Wert liegt dabei wieder im Intervall von $[0,1]$. Bei diesem Verfahren wird auf die Zusammenhänge zwischen den Konzepten verzichtet und stattdessen der Fokus auf die Kategorisierung dieser gelegt, um einen linearen Aufwand zu ermöglichen. Diese Metrik kann dabei, wie bei der Methode von Mohammed et al. [23], ebenfalls als Indiz für die Beurteilung von „shared understanding“ verwendet werden.

Entscheidungsbaum

Ohnishi et al. [31] stellen ein Konzept vor, bei dem Entscheidungsbaum genutzt werden, um „mutual understanding“ bei der Zusammenarbeit von Personen zu verbessern. In der folgenden Erläuterung wird „shared understanding“ synonym mit „mutual understanding“ verwendet. Die Methode beruht auf der Hypothese, dass „shared understanding“ geschaffen wird, wenn die Denkweise von anderen Personen in den eigenen Worten erklärt werden kann [31]. Dabei ist es nicht wichtig, dass die gleichen Symbole verwendet werden, solange die Verwendung verschiedener Symbole zwischen Personen konsistent miteinander ist [31].

Die Methode sieht vor, dass die beteiligten Personen jeweils einen eigenen Entscheidungsbaum erstellen, bei dem die Knoten jeweils ein Attribut besitzen, anhand dessen die Charakterisierung des jeweiligen Fallbeispiels vorgenommen wird [31]. Jede Verbindung zwischen zwei Knoten besitzt dabei eine der möglichen Ausprägungen des Attributs [31]. Anhand dieser Ketten von Entscheidungsattributen werden die Blätter des

Entscheidungsbaums erreicht, an denen die letztendliche Klasse des Fallbeispiels bestimmt wird, wobei jedes Blatt einer definierten Klasse entspricht [31]. Ein Vergleich der beiden Entscheidungsbäume findet statt, indem ein Fallbeispiel beide Entscheidungsbäume durchläuft und die jeweiligen Klassifikationen des Fallbeispiels ein Paar bilden [31]. Abbildung 21 zeigt ein Beispiel von Ohnishi et al. [31], in dem zwei Entscheidungsbäume verschiedener Personen miteinander verglichen werden.

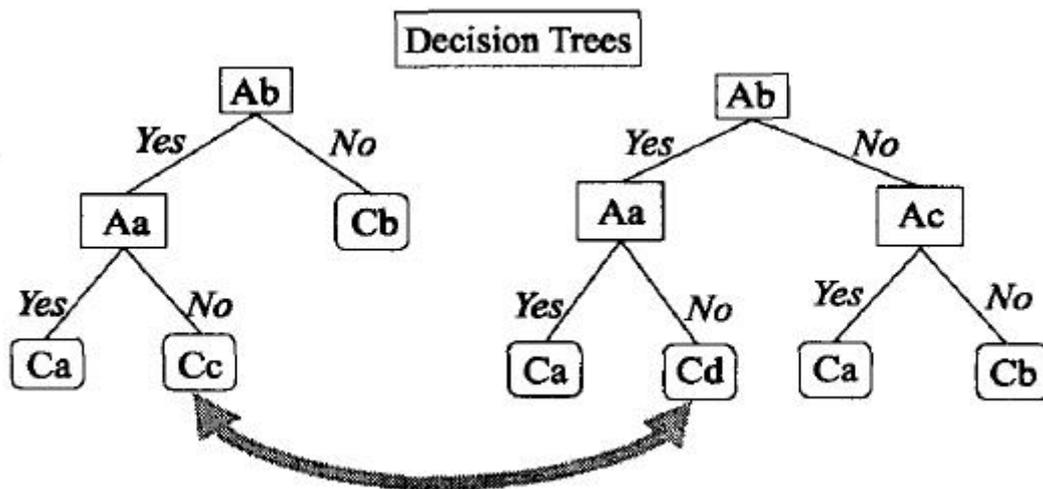


Abbildung 21: Vergleichs zweier Entscheidungsbäume von Ohnishi et al. [31]

Die Bezeichner Aa bis Ac stehen dabei für die Entscheidungsattribute, die auf die inneren Knoten verteilt sind, während Ca bis Cd die Symbole für die Klassifizierung des Fallbeispiels darstellen. In Abbildung 21 ist ein Beispiel zu sehen, in dem nach Eingabe eines Fallbeispiels mit den Attributen Ab:Yes und Aa:No die Klassifizierungen Cc und Cd als Paar identifiziert wurden, hervorgehoben durch den doppelseitigen Pfeil. Hierbei wird das eingangs erwähnte deutlich, dass korrespondierende Paare in verschiedenen Entscheidungsbäumen nicht zwangsläufig durch das gleiche Symbol gekennzeichnet sein müssen. Ohnishi et al. [31] haben hierfür eine einfache Methode definiert, anhand der man in drei Schritten die Übereinstimmung zweier Symbole prüfen kann:

Schritt 1: Durchsuche alle Paare mit verschiedenen Klassensymbolen. Nimm an, dass Y das Gegenstück zum Klassensymbol X darstellt.

Schritt 2: Überprüfe das Klassensymbol X', welches das Gegenstück zu allen Paaren mit Klassensymbol Y darstellt.

Schritt 3a: Falls $X=X'$ für alle in Schritt 1 gefundenen Paare gilt, dann beschreiben X und Y die gleiche Klasse.

Schritt 3b: Falls $X=X'$ nicht in allen Fällen aus Schritt 1 gilt, dann lässt sich der Grad der Übereinstimmung bestimmen durch die Anzahl der Paare die den Algorithmus erfüllen geteilt durch die gesamte Anzahl der Paare der Klasse. Je größer das Ergebnis ist, desto wahrscheinlicher ist die Übereinstimmung der jeweiligen Klassen.

Anhand dieser Methode der Nutzung von Entscheidungsbäumen lässt sich also berechnen, inwiefern die Klassifizierung von Fallbeispielen durch verschiedene Personen übereinstimmen. Dieses Ergebnis repräsentiert daher, ob die Entscheidungen, welche die betroffenen Personen auf Basis ihrer mentalen Modelle treffen, miteinander übereinstimmen.

Wenn die Entscheidungen übereinstimmen, die auf Basis der mentalen Modelle der Personen getroffen werden, kann davon ausgegangen werden, dass „shared understanding“ zwischen den Personen vorhanden ist. Ein Nachteil dieser Methode ist jedoch, dass nur Schlussfolgerungen anhand des Ergebnisses bezüglich der mentalen Modellen getroffen werden können und somit nicht die eigentliche Struktur der mentalen Modelle miteinander verglichen wird.

4.5.2 Numerische Konzepte

Unter numerischen Konzepten werden in dieser Arbeit all diejenigen Ansätze zusammengefasst, die im Gegensatz zu Mapping Konzepten auf eine visuelle Darstellung des mentalen Modells verzichten und stattdessen versuchen, die Inhalte des mentalen Modells mit Metriken zu quantifizieren. Die Zahlenwerte dieser Metriken können dabei sowohl die Inhalte des mentalen Modells einer einzelnen Person, als auch den Unterschied zwischen mentalen Modellen verschiedener Personen quantifizieren. Die hierbei zugrundeliegende Idee ist, dass die von einer Person getätigten Aktionen das mentale Modell dieser Person widerspiegeln. Im Gegensatz zu den Mapping Konzepten, die größtenteils Methoden umfassen, werden die numerischen Konzepte eher durch Metriken repräsentiert. Teilweise werden diese numerischen Konzepte in Kombination mit den Mapping Konzepten verwendet. Beispiele hierfür sind die „distance ratio“ Formel und das „paired comparison rating“, die in dieser Arbeit bereits im Zusammenhang mit dem „causal Mapping“ beziehungsweise der Pathfinder Methode erläutert wurden. Da Metriken oft spezifisch auf das jeweilige Ziel der Arbeit angepasst werden, in der sie Verwendung finden, konnte eine Vielzahl spezifischer Metriken in der Menge relevanter Publikationen identifiziert werden. An dieser Stelle werden daher nur einige der verallgemeinerbaren Konzepte präsentiert.

Fragebögen

Fragebögen sind eine der am meisten verwendeten Methoden, um Informationen über mentale Modelle von Personen zu sammeln [8,33]. Sie können dabei abhängig von der Art der gestellten Fragen sowohl qualitative als auch quantitative Ergebnisse liefern. Eine weit verbreitete quantitative Methode ist hierbei die Nutzung von Likert-Skalen [8,33,34,35,36,37,38]. Eine Likert-Skala ist eine Skala, welche die Zustimmung einer Person bezüglich einer gestellten Frage anhand einer Anzahl von abgestuften Punkten bestimmt [34,36,37,38]. Ein Beispiel hierfür wäre die 5-Punkt Likert-Skala, bei der ein Wert von 1 für starke Ablehnung und ein Wert von 5 für starke Zustimmung steht. Anhand von Fragebögen können daher diverse Metriken definiert, gemessen und verglichen werden, wie beispielsweise Einigkeit („Agreement“) oder Unterschiede zwischen Personen [38]. Anhand der oben genannten Likert-Skala könnte ein Unterschied zwischen Personen beispielsweise quantitative durch die Differenz der Punktebewertungen definiert werden, wobei eine kleinere Differenz einer größeren Übereinstimmung entsprechen würde [47]. „Agreement“ wird dabei oft als allgemeiner Begriff verwendet, der abhängig vom Kontext der Nutzung spezifisch definiert wird. Aus diesem Grund wird daher für diese Arbeit auf eine präzise, allgemeine Definition des Begriffes verzichtet und stattdessen eine fallspezifische Erklärung für den jeweiligen Kontext der Nutzung geliefert.

Einige der Vorteile von Fragebögen sind ihre Einfachheit und die draus folgende Skalierbarkeit und Automatisierbarkeit. Da Fragebögen im Kern ein sehr einfaches Konzept darstellen, das bei richtiger Umsetzung von jeder Person ohne großes Vorwissen verwendet werden kann, lassen sie sich sowohl auf eine sehr große Anzahl von Personen als auch auf

eine große Anzahl von Fragen skalieren [37,39]. Die Umsetzung von Online-Fragebögen unterstützt diese Skalierbarkeit und ermöglicht darüber hinaus eine automatische Verarbeitung der Fragebögen [35,37,40].

Im Kontext von den hier präsentierten numerischen Konzepten stellen Fragebögen daher eine grundlegende Methode dar, um eine große Menge an Informationen über die mentalen Modelle von Probanden zu sammeln, wodurch verschiedene Metriken zur Beurteilung von „shared understanding“ verwendet werden können.

Linguistische Analyse

Unter linguistischer Analyse werden in dieser Arbeit Methoden zusammengefasst, die anhand von Dialogen zwischen Personen Rückschlüsse über das „shared understanding“ dieser zieht. Reverdy und Vogel [41] zeigen in ihrer Arbeit einen Ansatz, der anhand der Wiederholungen, die in einer Konversation auftreten, ein Maß für das „shared understanding“ zwischen den Personen definiert. Die grundlegende Idee für dieses Konzept ist, dass durch Diskussionen zwischen Personen nach einer gewissen Zeit ein „shared understanding“ gewonnen werden kann [41]. Da jedoch allein durch Konversation in vielen Fällen nicht eindeutig bewiesen werden kann, ob der Gegenüber in einer Diskussion wirklich verstanden wurde, nutzen Reverdy und Vogel [41] Wiederholungen in einem Dialog als Indiz für die Erzeugung von „shared understanding“. Anhand der Wiederholungen definieren sie eine Proxy-Metrik für „shared understanding“, welche die auftretenden Wiederholungen in ein Verhältnis zu den in einer natürlichen Konversation auftretenden Wiederholungen setzt [41]. Diese Methode bezieht sich dabei auf einen Kontext, bei dem ein asynchroner Wissensstand vorliegt, wodurch eine Person einer anderen etwas erklären muss, wie beispielsweise den Verlauf eines Weges auf einer Karte [41]. Da sich der Kontext dieser Arbeit jedoch auf Vision Videos bezieht, bei denen alle beteiligten Personen durch das Vision Video einen synchronen Wissensstand erhalten, wird dieser Ansatz nicht tiefergehend verfolgt.

Suzuki und Tsuda [42] präsentieren einen anderen Ansatz der linguistischen Analyse, bei dem „shared understanding“ anhand des Egozentrismus des Dialogs bestimmt wird. Unter Egozentrismus wird in diesem Kontext eine Einstellung verstanden, bei der die betroffene Person ihren eigenen Standpunkt nicht überdenken kann [42]. Diese Einstellung sehen Suzuki und Tsuda [42] als Hindernis für gute Kommunikation und als ein Anzeichen für das Fehlen von „shared understanding“. Aus diesem Grund wurden Anzeichen von egozentrischen Aussagen aus Dialogen extrahiert und entsprechend des Grades von Egozentrismus gewichtet. Tabelle 4 zeigt hierfür die Klassifikation von egozentrischen Aussagen mit der dazugehörigen Gewichtung nach Suzuki und Tsuda [42]

Tabelle 4: Gewichte für die Bestimmung von Egozentrismus nach Suzuki und Tsuda [42]

Classification	Subdivision	Weight
Attack to Others	Dissatisfaction	12
	Contempt	11
Lack of Empathy	Detachment	10
	Irony	9
Priority to One's Convenience	Defiant Attitude	8
	Restriction	7
Self Validity	Conclusion	6
	Imposition	5
	Excuse	4
Belonging to Self-Profit	Demand	3
Inference	Inference	2
	Question	1
Agreement	Thanks	0

Die Gewichte reichen hierbei von „Zustimmung“ mit Wert 0 bis hin zu „Angriff auf Andere“ mit Wert 12, wobei ein höherer Wert einem stärkerem Grad von Egozentrismus gleichzusetzen ist. Mithilfe von Tabelle 4 lassen sich dann Dialoge klassifizieren, indem jedem Satz des Dialogs das entsprechende Gewicht x_i zugeordnet wird [42]. Der Egozentrismus im Dialog wird dann durch den durchschnittlichen Egozentrismus der enthaltenen Sätze bewertet [42]. Hierfür werden jedoch zwei Sonderfälle definiert. Im Fall, dass ein Satz des Dialogs mit „Zustimmung“ klassifiziert werden kann, wird der gesamte Dialog als „Zustimmung“ klassifiziert [42]. Dies wird damit begründet, dass Sätze, die einer Zustimmung folgen, in den meisten Fällen lediglich zusätzliche Erklärung zu dieser Zustimmung darstellen [42]. Der zweite Sonderfall bezieht sich auf den letzten Satz eines Dialogs. Laut Suzuki und Tsuda [42] ist in den meisten Fällen der Egozentrismus im letzte Satz eines Dialogs, definiert als 1, repräsentativ für den gesamten Dialog. Aus diesem Grund wird der Wert S des Egozentrismus im gesamten Dialog als Maximum des durchschnittlichen Egozentrismus der einzelnen Sätze und des Egozentrismus des letzten Satzes definiert:

$$S = \max\left(1, \frac{\sum x_i}{N}\right)$$

Dieser Wert kann als Metrik verwendet werden, anhand der man das „shared understanding“ der beteiligten Personen quantitativ beurteilen kann. Eine Möglichkeit der qualitativen Beurteilung der Gewichtung des Dialogs liefern Suzuki und Tsuda [42] in der Form von vier verschiedenen Klassifikationen durch Patterns. Hierfür werden die Gewichtungen der einzelnen Sätze im Verhältnis zur Zeit graphisch dargestellt. Abbildung 22 zeigt beispielhaft das erste Pattern, welches von Suzuki und Tsuda [42] aufgestellt wurde. Die verschiedenfarbigen Geraden zeigen dabei verschiedene beispielhafte Ausprägungen des Patterns auf. Die anderen drei Patterns befinden sich in Anhang E.

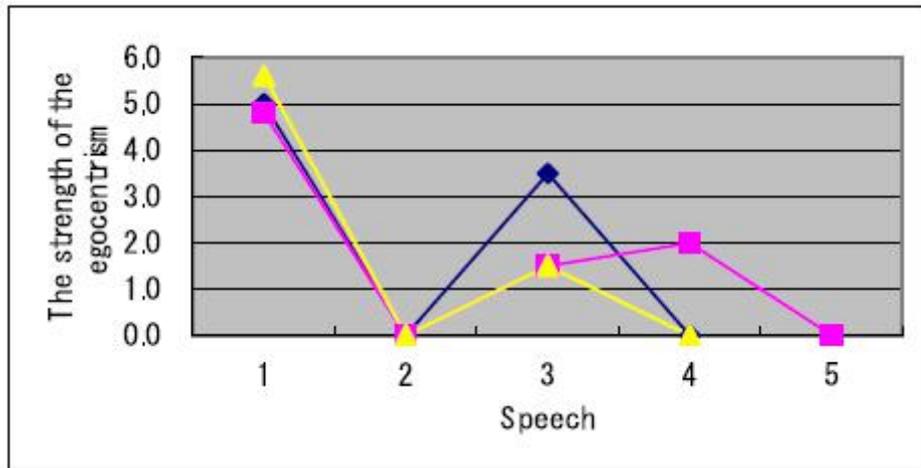


Abbildung 22: Beispielhafte Darstellung des ersten Patterns von Suzuki und Tsuda [42]

Das erste Pattern zeichnet sich dadurch aus, dass das Gewicht des letzten Satzes des Dialogs 0 entspricht. Wie vorher beschrieben lässt eine Zustimmung auf einen geringen Unterschied im „shared understanding“ der Personen vermuten [42]. Die gleiche Schlussfolgerung gilt für das zweite Pattern, bei dem der vorletzte Satz einen Wert von 0 besitzt, was auf weitere Erklärungen zur Zustimmung schließen lässt. Das dritte und vierte Pattern umfassen diejenigen Fälle, in denen entweder die Gewichtung des Egozentrismus durchgängig hoch oder der Verlauf schwankend ist und dadurch keinem der anderen Patterns zugeordnet werden kann. In beiden dieser Fälle wird von einem Fehlen von „shared understanding“ ausgegangen [42]. Die Zuordnung eines konkreten Beispiels zu einem dieser Patterns kann daher als eine qualitative Bewertung des „shared understandings“ genutzt werden.

Repertory Grid Technik

Die Repertory Grid Technik ist eine flexible Methode, die das mentale Modell von Personen bezüglich einer gegebenen Domäne bestimmen soll [48]. Die Methode besteht hierfür aus drei aufeinanderfolgenden Schritten. Im ersten Schritt sollen die Probanden Konzepte frei definieren, die ihnen relevant für die betrachtete Domäne erscheinen [14,19,48]. Eine Abwandlung dieses Schrittes sieht vor, dass den Probanden die Konzepte vorgegeben werden, um eine bessere Vergleichbarkeit zwischen den Ergebnissen verschiedener Probanden zu ermöglichen [14,49]. Im zweiten Schritt werden den Probanden nacheinander eine Reihe von Dreiergruppen von den im ersten Schritt definierten Konzepten vorgelegt [19,48]. Die Probanden sollen für jede Dreiergruppe ein Maß definieren, anhand dessen zwei der Konzepte von dem Dritten aufgrund ihrer Ähnlichkeit unterschieden werden können [14,19,48]. Diese Methode wird solange wiederholt, bis die für den Kontext relevantesten Maße identifiziert wurden [19]. Im letzten Schritt werden die Probanden dazu aufgefordert, in einer Matrix alle Kombinationen aus Konzepten und Maßen zu bewerten, bei der in den Spalten die Konzepte und in den Zeilen die Maße stehen [19,48,49]. Die Bewertungsskala kann dabei auf den jeweiligen Kontext angepasst verschieden definiert werden, beispielsweise als Likert-Skala mit einem Wertebereich von 1 bis 5 [49]. Abhängig vom Kontext der Nutzung und der gewählten Bewertungsskala können diverse statistische Analysemethoden auf die Matrix angewendet werden, um sowohl Rückschlüsse auf das mentale Modell des jeweiligen Probanden zu ziehen, als auch um die Ergebnisse verschiedener Probanden zu vergleichen. Eine einfache Möglichkeit zum Vergleichen von Personen bieten dabei diverse Distanzmetriken, mit denen der Unterschied zwischen zwei verschiedenen Matrizen bestimmt werden kann [49]. Andere Analysemethoden, die in diesem Zusammenhang Verwendung finden, sind die Cluster-Analyse und die Hauptkomponentenanalyse [19,48,49]. Bei einer

Cluster-Analyse wird überprüft, wie die Konzepte anhand ihrer Maße gruppiert sind [48]. Hierbei ist es sowohl möglich, ein einzelnes Maß, als auch den Zusammenhang zwischen einer Kombination aus mehreren Maßen zu betrachten. Ein einfaches Beispiel, bei der die Kombination von zwei Maßen betrachtet wird, ist, wenn die Konzepte gleichzeitig sowohl anhand ihres Zeitaufwands als auch anhand ihres Kostenaufwands gruppiert werden. Bei der Hauptkomponentenanalyse wird ein Graph produziert, der darstellt, welche Maße sich zwischen verschiedenen Konzepten unterscheiden [48]. Sowohl die Cluster-Analyse als auch die Hauptkomponentenanalyse sind hierbei Verfahren, deren Ergebnisse durch die Auswahl von Maßen teilweise von einer subjektiven Interpretation abhängig sind [48].

Da durch die Repertory Grid Technik Rückschlüsse auf das mentale Modell der Probanden gezogen werden und diese Rückschlüsse zwischen verschiedenen Probanden vergleichbar sind, können sowohl die qualitativen als auch quantitativen Analysemethoden auf Basis dieser Ergebnisse als Ansatz für die Beurteilung für das „shared understanding“ zwischen den jeweiligen Probanden verwendet werden.

4.6 Übertragung auf Vision Videos

Im vorhergehenden Kapitel wurden einige Ansätze präsentiert, mit denen „shared understanding“ zwischen einzelnen Personen und teilweise auch innerhalb von Teams beurteilt werden kann. Damit diese Methoden und Metriken im Kontext dieser Arbeit verwendet werden können, gilt es nun zu untersuchen, inwiefern sie auf Vision Videos angewendet werden können. Wie in Kapitel 2.4 erklärt wurde sind Vision Videos ein Medium, welches ein „shared understanding“ zwischen denjenigen Personen erzeugen soll, die das Vision Video gesehen haben. Betrachtet man im Vergleich dazu die in Kapitel 4.5 vorgestellten Methoden und Metriken, so lässt sich die Gemeinsamkeit feststellen, dass sie alle das „shared understanding“ unabhängig vom Medium bestimmen. Das soll bedeuten, dass keine der Methoden oder Metriken sich auf die Nutzung eines bestimmten Mediums bezieht, welches das „shared understanding“ erzeugt hat. Im Umkehrschluss kann man daher folgern, dass jede der Methoden und Metriken, die in Kapitel 4.5 präsentiert wurden, im Kontext der Nutzung von Vision Videos verwendet werden kann. Die Methoden sollten hierfür angewendet werden, nachdem eine Personengruppe ein Vision Video gesehen hat, um festzustellen, ob ein „shared understanding“ zwischen den Personen erzeugt wurde. Alternativ können die Methoden auch bevor und nachdem die Personen das Vision Video gesehen haben angewendet werden, um eine Veränderung im „shared understanding“ messen zu können.

Ein Beispiel für die Verwendung einer der Methoden im Kontext von Vision Videos wird hier anhand einer linguistischen Analyse des Egozentrismus vorgestellt. In diesem Beispiel soll festgestellt werden, ob ein „shared understanding“ zwischen Personen durch ein Vision Video erzeugt wurde. Hierfür wird den Personen zunächst ein entsprechendes Vision Video gezeigt. Danach sollen die Personen anhand von vorgegebenen Gesprächsthemen über das Gesehene diskutieren, um ihre jeweiligen Sichtweisen zu dem Thema zu vergleichen. Anhand einer Mitschrift dieses Dialogs kann dann entsprechend der Tabelle 4 von Suzuki und Tsuda [42] in Kapitel 4.5.2 das „shared understanding“ zwischen den Personen analysiert werden. Durch dieses Beispiel lässt sich erkennen, dass sich nichts an der Verwendung der Methode ändern würde, sollte das Vision Video gegen ein anderes Medium ausgetauscht werden. Dies zeigt die Unabhängigkeit der Methode vom Medium, mit dem das „shared understanding“ erzeugt wird. Entsprechend wird hierbei auch deutlich, dass die Methode, mit der das „shared understanding“ bestimmt wird, ebenso beliebig durch eine andere ersetzt werden kann, ohne dass an der Präsentation des Vision Videos etwas geändert werden muss. Diese Erkenntnisse

sind essentiell für die Entwicklung des Experimentansatzes, welcher in Kapitel 5 präsentiert wird.

4.7 Threats to Validity

Während der Durchführung der systematischen Literatursuche mussten aufgrund des Rahmens, in dem diese Arbeit geschrieben wird, einige Einschränkungen getroffen werden. Da nicht ausgeschlossen werden kann, dass diese Einschränkungen einen Einfluss auf das Ergebnis dieser Arbeit besitzen, muss an dieser Stelle die Validität der Ergebnisse geprüft werden. Hierzu werden all diejenigen Aspekte der Arbeit betrachtet, die eine Gefahr für die Validität darstellen, auch bekannt als „Threats to Validity“. Die Identifizierung dieser „Threats to Validity“ ist essentiell, da die Ergebnisse dieser Arbeit sonst nicht verallgemeinerbar wären. Die „Threats to Validity“ können dabei in vier grundlegende Kategorien aufgeteilt werden, auf die in den folgenden Kapiteln genauer eingegangen wird. Um möglichst alle „Threats to Validity“ zu identifizieren, die für diese Arbeit relevant sind, wurde die Arbeit von Li und Huang [50] als Ausgangspunkt verwendet. In ihrer Arbeit analysieren Lu und Huang [50] die „Threats to Validity“, die typischerweise im Kontext von systematischen Literatursuchen vorhanden sind. Dabei wurden die einzelnen „Threats to Validity“ in einer Form aufgelistet, in der sie als Richtlinie verwendet werden können, um die „Threats to Validity“ in einer spezifischen Arbeit identifizieren zu können. Es ist jedoch anzumerken, dass abhängig davon, wie die einzelnen Aspekte der Richtlinie im spezifischen Kontext der Arbeit ausgelegt werden, diese mehr als einer Kategorien zugeordnet werden könnten.

4.7.1 Construct Validity

Unter „Construct Validity“ werden all diejenigen „Threats to Validity“ klassifiziert, die sich auf den Zusammenhang zwischen der Theorie und der Umsetzung beziehen [51]. Die essentielle Aufgabe in diesem Kontext ist es, die richtigen operativen Maßnahmen für die Konzepte zu identifizieren, die in dieser Arbeit betrachtet werden [50].

Ein Problem dieser Art war, dass einige der Terminologien, die für diese Arbeit relevant sind, nicht allgemein definiert sind, wodurch diverse Abwandlungen dieser Terminologien verwendet werden. Die gravierendste Ausprägung dieses Problems bezieht sich dabei auf den Begriff „shared understanding“, für den zwar Definitionen wie die von Aranda [54] vorhanden sind, diese jedoch nicht immer Anwendung finden und dadurch alternative Begriffe wie „mutual understanding“ [42] und „common understanding“ [84] verwendet werden. Um diesem Problem entgegen zu wirken wurden daher alle Varianten von relevanten Terminologien, die in der manuellen Suche gefunden wurden, berücksichtigt und im Kontext dieser Arbeit als synonym behandelt. Darüber hinaus wurden für die auf der manuellen Suche aufbauenden Suchschritte jeder der synonymen Begriffe berücksichtigt, indem alle gefundenen Synonyme in den Suchstring integriert wurden.

Die Suchbegriffe, aus denen der Suchstring besteht, stellen eine weitere potentielle „Threat to Validity“ dar. Es ist hierbei möglich, dass entweder inkorrekte Suchbegriffe in den Suchstring aufgenommen wurden oder einige relevante Suchbegriffe für den Suchstring fehlen. Der Fall, dass inkorrekte Suchbegriffe in den Suchstring integriert wurden, konnte insofern kompensiert werden, dass dieser nach erster Anwendung in der Datenbanksuche anhand der erzeugten Ergebnisse angepasst wurde, indem irrelevante und nicht zielführende Begriffe entfernt wurden, um ein besseres Ergebnis zu erzielen. Darüber hinaus wird die Anzahl an fehlerhaften Ergebnissen aufgrund von inkorrekten Suchbegriffen dadurch minimiert, dass die

Ergebnisse des Suchstrings mehrfach manuell überprüft wurden, um ihre Relevanz für diese Arbeit zu gewährleisten. Das Risiko des Fehlens von relevanten Suchbegriffen wurde in sofern minimiert, dass eine Vielzahl verschiedener Journals und Konferenzen herangezogen wurde, um die manuelle Suche durchzuführen. Darüber hinaus wurden mit andauernder Durchführung der manuellen Suche deutlich weniger neue relevante Begriffe gefunden, was auf eine Stagnation der Suchergebnisse hinwies. Des Weiteren konnten fehlende Suchbegriffe im Suchstring durch das spätere Snowballing kompensiert werden, da diese Methode unabhängig von der Terminologie des Suchstrings ist. Letztendlich kann jedoch nicht garantiert werden, dass jeder relevante Suchbegriff berücksichtigt wurde.

Ein weiteres Problem, welches häufig bei systematischen Literatursuchen auftaucht, ist, dass die angewendete Suchmethode für den jeweiligen Kontext nicht geeignet ist. Um für diese Arbeit die Verwendung von geeigneten Suchmethoden sicherzustellen, wurde Literatur bezüglich geeigneter Suchmethoden für systematische Literatursuchen berücksichtigt und letztendlich die Ergebnisse der Ausarbeitung von Brings et al. [2] als Richtlinie benutzt, um die Suchmethode für diese Arbeit zu definieren. Ein Problem in diesem Zusammenhang ist jedoch, dass aufgrund der Art dieser Arbeit die systematische Literatursuche nur durch eine Person durchgeführt werden konnte, wodurch die Durchführung möglicherweise beeinflusst wurde. Dieses Problem ist schwer zu lösen, da eine Einzelperson immer einen gewissen subjektiven Einfluss besitzt. Jedoch konnte dieser subjektive Einfluss etwas minimiert werden, indem Zwischenschritte des Suchprozesses nach einem gewissen zeitlichen Abstand abermals überprüft wurden, um die Konsistenz der Ergebnisse zu gewährleisten.

Das letzte Problem der „Construct Validity“ bezieht sich auf zwei „Threats to Validity“, die im Fall dieser Arbeit stark voneinander abhängig sind. Das Ausgangsproblem ist hierbei die Auswahl der Journals, Konferenzen und Datenbanken, die für die einzelnen Suchschritte berücksichtigt wurden. Es ist möglich, dass hierfür einige Quellen nicht berücksichtigt wurden, die für den Kontext dieser Arbeit wichtige Inhalte besitzen. Dieses Problem wurde dadurch gelöst, dass die populärsten Journals, Konferenzen und Datenbanken anhand entsprechender Ranglisten und Metriken identifiziert und ausgewählt wurden. Darüber hinaus wurden andere Quellen, deren Relevanz bekannt war, mit in die Suche einbezogen, wie beispielsweise das Journal für Requirements Engineering. Das hauptsächliche Problem hierbei ist jedoch, dass nur eine begrenzte Anzahl an Journals, Konferenzen und Datenbanken in den jeweiligen Suchschritten berücksichtigt werden konnten. Dies hängt dabei mit dem anderen grundsätzlichen Problem zusammen, dass durch den Rahmen dieser Arbeit die zur Verfügung stehende Zeit begrenzt ist. Durch diese Zeitbegrenzung war es nötig, die Anzahl der betrachteten Journals, Konferenzen und Datenbanken einzuschränken, um die Arbeit im vorgegebenen Zeitrahmen fertigzustellen. Hierbei ist jedoch anzumerken, dass die Menge der betrachteten Publikationen in dieser Arbeit mit denen von anderen systematischen Literatursuchen im Bereich des Requirements Engineerings vergleichbar ist oder diese sogar übersteigt. Allein im Rahmen der manuellen Suche dieser Arbeit wurden 8035 Publikationen durchsucht, während beispielsweise Brings et al. [2] im Verlauf ihrer gesamten systematischen Literatursuche nur 7191 Publikationen betrachtet haben. Dies lässt vermuten, dass die systematische Literatursuche dieser Arbeit einen verhältnismäßig umfangreichen Suchprozess besitzt. Jedoch ist nicht auszuschließen, dass mit mehr verfügbarer Zeit weitere Ergebnisse auffindbar wären, die im Kontext der Beurteilung von „shared understanding“ durch Vision Videos interessant sein können.

4.7.2 Internal Validity

Unter „Internal Validity“ werden all diejenigen „Threats to Validity“ klassifiziert, die sich auf die kausale Beziehung zwischen der Umsetzung und den Ergebnissen der Arbeit beziehen [50,51]. Es muss hierbei sichergestellt werden, dass die Ergebnisse durch die beschriebene Umsetzung erzeugt wurden und nicht durch einen nicht berücksichtigten Faktor beeinflusst werden, weder kontrolliert noch gemessen wird [51].

Kulturelle Vorurteile gegenüber Publikationen sind ein Problem, die in einer Arbeit auftreten können. Es kann für diese Arbeit mit Sicherheit gesagt werden, dass während der Auswahl der Publikationen zu keinem Zeitpunkt auf den kulturellen Hintergrund einer Publikation geachtet wurde. Jedoch muss an dieser Stelle angemerkt werden, dass die Inhalte einiger Publikationen, die eventuell relevant für diese Arbeit gewesen wären, aufgrund unzureichender sprachlicher Kenntnisse seitens der Autoren so unverständlich waren, dass sie nicht berücksichtigt werden konnten.

Ein grundsätzliches Problem dieser Arbeit ist, dass die systematische Literatursuche aufgrund der Art dieser Arbeit nur von einer Person durchgeführt werden konnte. Dies führt zu einer Vielzahl möglicher Probleme in Bezug auf die „Internal Validity“. Alle im Folgenden dargestellten Probleme haben dabei auch einen gewissen Einfluss auf die „Conclusion Validity“ dieser Arbeit. Eines dieser Probleme ist, dass Kriterien für die Arbeit auf Basis von subjektiven Ansichten gewählt werden. Um dieses Problem zu minimieren, wurden die Kriterien, die an verschiedenen Stellen dieser Arbeit verwendet wurden, wenn möglich auf Basis von in der Literatur vorhandenen Richtlinien definiert. Ein Beispiel hierfür ist die Auswahl von Journals und Konferenzen für die manuelle Suche, die auf Basis des Vergleichs verschiedener Ranglisten getroffen wurde.

Ein weiteres Problem, was bei der Durchführung der systematischen Literatursuche durch eine Person entsteht, ist, dass ein subjektiver Einfluss auf die Auswahl der Publikationen und die Extrahierung der Daten besteht. Während diese Probleme nicht ausgeschlossen werden können, werden sie in dieser Arbeit durch Wiederholungen und Überprüfung der Auswahl in den jeweiligen Suchschritten minimiert. Darüber hinaus wurden Daten, die häufiger in Publikationen auftraten, eher extrahiert als diejenigen, die seltener beziehungsweise nur einmal auftraten. Entsprechend wurden Publikationen, deren Thema häufiger auftrat eher berücksichtigt als diejenigen mit sehr spezifischen, einmalig repräsentierten Themen. Bezüglich der Relevanz der einzelnen Publikationen beziehungsweise Daten war es jedoch unvermeidbar, eine subjektive Einschätzung vorzunehmen.

Die letzte Problematik, die im Zusammenhang dieses Abschnitts zu betrachten ist, bezieht sich auf die Fehlklassifizierung von Publikationen. Hierbei ist es möglich, dass sich die Art und Weise, in der Publikationen klassifiziert wurden, während der Durchführung geändert hat, wodurch es zu inkonsistenten Ergebnissen kommen würde. Dieses Problem wurde zum einen dadurch minimiert, dass vor der Durchführung der jeweiligen Suchmethoden Inklusions- und Exklusionskriterien definiert wurden, anhand derer die Auswahl der Publikationen vorgenommen wurde. Solange sich bei der Klassifizierung an diese Kriterien gehalten wurde, sollten die Ergebnisse konsistent bleiben. Zum anderen wurden die Ergebnisse der einzelnen Suchschritte überprüft, wodurch in Kombination mit den Inklusions- und Exklusionskriterien eine Fehlklassifizierung der Publikationen zum größten Teil ausgeschlossen werden kann.

4.7.3 External Validity

Unter „External Validity“ werden all diejenigen „Threats to Validity“ klassifiziert, die sich auf die Generalisierbarkeit der Arbeit beziehen [50,51]. In diesem Zusammenhang muss

geklärt werden, inwiefern die kausalen Zusammenhänge, die durch die Ergebnisse dieser Arbeit gefolgert wurden, repräsentativ für die Allgemeinheit sein können [51]. Im Kontext einer systematischen Literatursuche kann sich diese Allgemeinheit beispielsweise auf die Menge aller Publikationen beziehen, die für die Arbeit relevant wären.

Ein Problem im Zusammenhang mit „External Validity“ besteht darin, dass einige der Publikationen, die eventuell relevant für diese Arbeit gewesen wären, nicht zugänglich waren. Dieses Problem entsteht dadurch, dass es sich bei dieser Arbeit um eine studentische Arbeit handelt und sich somit der Zugriff auf Publikationen auf den Zugang über die Universität beschränkt. Aus diesem Grund konnten einige Publikationen nicht berücksichtigt werden, die anhand ihres Titels und ihrer Zusammenfassung eventuell relevante Inhalte für diese Arbeit enthalten hätten. Dieses Problem wird jedoch dadurch kompensiert, dass eine Vielzahl von Publikationen identifiziert werden konnten, die für das Ergebnis dieser Arbeit relevant sind und sich thematisch mit den Publikationen überschneiden, auf die nicht zugegriffen werden konnte.

Ein anderes mögliches Problem dieser Arbeit ist, dass die Darstellung der Informationen dieser Arbeit nicht vollständig ist. Dieses Problem bezieht sich an dieser Stelle nicht auf die objektive oder detaillierte Darstellung von Informationen wie von Li und Huang [50]. Stattdessen bezieht sich dieses Problem darauf, dass eventuell Ansätze nicht berücksichtigt wurden, da sie während des Suchprozesses nicht gefunden wurden. Ein besonders zutreffendes Beispiel hierfür sind Ansätze aus interdisziplinären Domänen, die equivalent zum „shared understanding“ sind, was in dieser Arbeit betrachtet wird. Diesbezüglich wurden zwar Journals als interdisziplinären Domänen auf relevante Publikationen untersucht, jedoch fehlt an dieser Stelle das für die Domäne spezifische Wissen um sicherstellen zu können, dass alle relevanten Ansätze berücksichtigt wurden. Insofern können die Ergebnisse dieser Arbeit mit endgültiger Sicherheit nur im Kontext von Software Engineering und damit verwandten Domänen verallgemeinert werden. Für alle anderen Domänen liefert diese Arbeit jedoch Ansätze, die in den jeweiligen Kontext übertragen werden könnten.

4.7.4 Conclusion Validity

Unter „Conclusion Validity“ werden all diejenigen „Threats to Validity“ klassifiziert, die sich auf den statistischen Zusammenhang zwischen der Umsetzung und den Ergebnissen der Arbeit beziehen [51]. Hierbei ist es essentiell, dass die Umsetzung der Arbeit reproduzierbar ist und diese Reproduktion die gleichen Ergebnisse wie die ursprüngliche Arbeit liefert [50].

Neben den „Threats to Validity“, die bereits im Abschnitt zur „Internal Validity“ vorgestellt wurden, gibt es noch ein weiteres Problem, welches sich spezifisch auf die „Conclusion Validity“ bezieht. Das Problem hierbei ist, dass die Beurteilung der extrahierten Daten aus den jeweiligen Publikationen auf Basis einer subjektiven Interpretation dieser beruht. Dies hat zur Folge, dass die Reproduktion dieser systematischen Literatursuche nicht die gleichen Ergebnisse liefern würde. Dieses Problem ist dabei, wie einige andere, darauf zurückzuführen, dass die systematische Literatursuche aufgrund der Art dieser Arbeit durch eine einzelne Person durchgeführt werden musste, wodurch ein subjektiver Einfluss nicht auszuschließen ist. Um dem entgegenzuwirken wurden einige Schritte des Suchprozesses wiederholt, um die Konsistenz der Beurteilung der Daten zu gewährleisten. Darüber hinaus wurden für die Wahl der Ansätze zur Beurteilung von „shared understanding“ in Kapitel 4.5 alle der häufig auftretenden Methoden berücksichtigt.

5 Experimentdesign

In Kapitel 4 wurden diverse Ansätze identifiziert, mit denen „shared understanding“ beurteilt werden kann. Darüber hinaus wurde festgestellt, dass jeder dieser Ansätze theoretisch auf den Kontext von Vision Videos übertragen werden kann. Im nächsten Schritt gilt es also zu überprüfen, inwiefern diese theoretischen Ansätze in der Praxis zur Beurteilung von „shared understanding“ durch Vision Videos umgesetzt werden können. Die Durchführung einer Studie in Form eines Experiments stellt hierfür eine geeignete Methode dar, um dies zu überprüfen. Aufgrund der zeitlichen Restriktionen, unter denen diese Arbeit geschrieben werden musste, ist es jedoch nicht möglich, neben der systematischen Literatursuche ein entsprechendes Experiment durchzuführen, welches einen genügend großen Umfang besitzt, um ein repräsentatives Ergebnis zu erzeugen. Um den zeitlichen Rahmen dieser Arbeit einhalten zu können wurde daher bereits in der Aufgabenstellung dieser Arbeit auf die Durchführung des Experiments verzichtet. Stattdessen wird daher an dieser Stelle ein detailliertes Experimentdesign erarbeitet, welches als Richtlinie für eine zukünftige Arbeit verwendet werden kann, die ein entsprechendes Experiment durchführt.

Experimente stellen eine Art der Studie dar, mit der man statistische Analysen auf Basis des Testens von Hypothesen durchführen kann [51]. In dieser Arbeit wurden diverse Methoden identifiziert, mit denen man eine Beurteilung des „shared understandings“ auf Basis von Vision Videos durchführen kann. Durch ein Experiment lassen sich verschiedene Hypothesen zu Ansatzpunkten definieren, um zu prüfen, inwiefern die Beurteilung von „shared understanding“ durch Vision Videos mit diesen Methoden in der Praxis möglich ist und ob Vision Videos tatsächlich einen messbaren Mehrwert in Bezug auf die Erzeugung von „shared understanding“ liefern. Um bei der Definition des Experimentdesigns ein strukturiertes Vorgehen zu gewährleisten, werden die folgenden Unterkapitel anhand der von Wohlin et al. [51] entwickelten Planungsschritte für ein Experimentdesign im Software Engineering orientiert. Abbildung 23 zeigt hierfür den grundlegenden Ablauf der Planung eines Experiments mit den dazugehörigen Schritten.

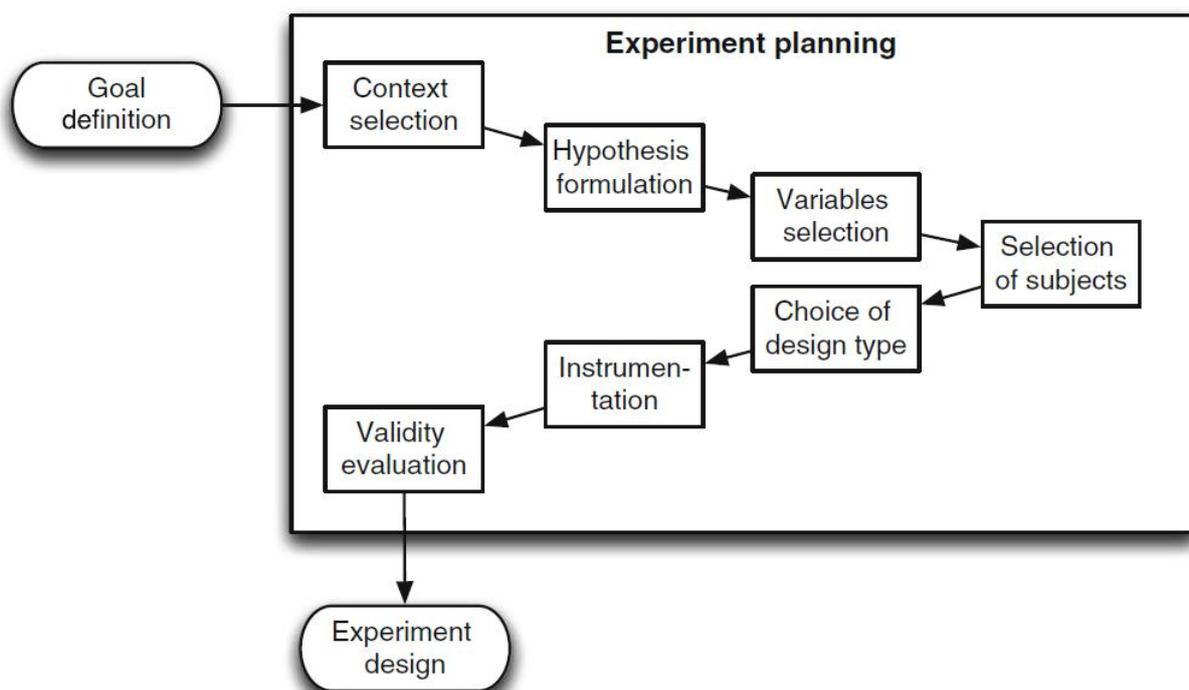


Abbildung 23: Ablauf der Schritte einer Experimentplanung nach Wohlin et al. [51]

Laut Wohlin et al. [51] gilt es hierbei zunächst, das Ziel des Experiments zu definieren, dargestellt durch die „Goal definition“. Anhand dieser Zieldefinition lässt sich dann die Planung des Experiments („Experiment planning“) mit ihren einzelnen Schritte definieren. Das Ergebnis dieser Planung des Experiments ist dann das fertige Design des Experiments, in Abbildung 23 dargestellt durch das „Experiment design“. Alle folgenden Unterkapitel dieser Arbeit beschreiben hierfür einen Schritt dieser Planung des Experiments. Falls nicht anders angegeben, wird sich dabei auf die Richtlinie von Wohlin et al. [51] bezogen.

5.1 Zielfindung

Bevor die eigentliche Planung des Experiments vollzogen wird, muss zunächst das Ziel definiert werden, auf das die Planung des Experiments ausgerichtet wird. Die Zielfindung dient somit als Grundlage für das gesamte Experiment. Wird diese Grundlage nicht sorgfältig definiert, dann kann es im Verlauf des Experiments dazu kommen, dass Teile des Experiments überarbeitet werden müssen oder im Extremfall das gesamte Experiment unbrauchbar für das eigentliche Ziel der jeweiligen Arbeit wird. Um diesem Problem entgegenzuwirken und die Wahrscheinlichkeit für eine unzureichende Zieldefinition zu minimieren, schlagen Wohlin et al. [51] die Nutzung des GQM-Templates für Zieldefinition vor, welches ursprünglich von Basili und Rombach [52] präsentiert wurde:

Analysiere <Objekt(e) der Studie>
zum Zweck <Zweck>
in Bezug auf <Qualitätsfokus>
aus der Perspektive <Perspektive>
im Kontext von<Kontext>

Das Ziel der Nutzung dieses Templates besteht darin, dass die wichtigsten Aspekte des Experiments definiert sind, bevor mit der Planung oder Umsetzung begonnen wird.

Die übergeordnete Zielsetzung der Arbeit bietet an dieser Stelle den ersten Ansatzpunkt, anhand dessen Teile der Zielsetzung definiert werden können. Das übergeordnete Ziel ist es, Methoden, Heuristiken und Metriken zu identifizieren, anhand derer Ansätze entwickelt werden können, mit denen die Beurteilung des „shared understandings“ auf Basis von Vision Videos“ möglich ist. Nachdem in Kapitel 4 entsprechende Methoden zur Bestimmung von „shared understanding“ identifiziert wurden konnte geschlussfolgert werden, dass sich diese Methoden auf den Kontext von Vision Videos übertragen lassen. Das Objekt der Analyse dieses Experimentdesigns sind dabei die Vision Videos, auf deren Basis „shared understanding“ erzeugt werden soll. Der Zweck der Analyse des Objekts der Studie ist hierbei die Beurteilung des „shared understandings“, welches durch Vision Videos erzeugt wird. Unter Berücksichtigung des Objekts der Studie lässt sich ebenfalls der Qualitätsfokus bestimmen. In diesem Zusammenhang kann entweder geprüft werden, ob eine Erzeugung von „shared understanding“ durch die Betrachtung von Vision Videos erfolgt oder wie groß der Mehrwert dieser Erzeugung in Bezug auf das „shared understanding“ der beteiligten Personen ist. In dem Fall, dass analysiert wird, ob Vision Videos ein „shared understanding“ erzeugen, liegt hierbei der Qualitätsfokus auf der Relevanz der Vision Videos für die Erzeugung von „shared understanding“. Wird stattdessen der Mehrwert der Vision Videos für die Erzeugung von „shared understanding“ betrachtet, dann liegt der Qualitätsfokus stattdessen auf der Größe des Mehrwerts der Vision Videos. Die Perspektive, aus der die Ergebnisse des Experiments betrachtet werden sollten, lässt sich hier aus dem Kontext der Vision Videos ableiten. Vision Videos werden im Rahmen von Softwareprojekten eingesetzt, bei denen

diverse Gruppen von Stakeholdern und Entwicklern miteinander zusammenarbeiten (siehe Kapitel 2.4). Entsprechend ist die Perspektive, aus der die Ergebnisse betrachtet werden, die der verschiedenen Stakeholder und der Entwickler. Zuletzt ist zu klären, in welchem Kontext das Experiment durchgeführt wird. Hierbei müssen kurz sowohl die Personen definiert werden, die am Experiment beteiligt sind, als auch die Objekte, die für das Experiment verwendet werden. Für dieses Experimentdesign werden Probanden mit verschiedenem Hintergrundwissen benötigt, um die verschiedenen Gruppen von Stakeholdern und Entwicklern repräsentieren zu können. Diesen Probanden werden dann verschiedene Vision Videos gezeigt, auf deren Basis das Experiment durchgeführt wird. Auf die Details des Kontextes und der Probandenauswahl wird dabei in den entsprechenden Unterkapiteln der Planung der Experiments genauer eingegangen.

Nach Betrachtung der einzelnen Bestandteile des GQM-Templates ist festzustellen, dass sich anhand des Qualitätsfokusses der Studie zwei verschiedene Ziele definieren lassen. Das erste Ziel bezieht sich dabei auf die Analyse, ob durch das Betrachten von Vision Videos eine Erzeugung von „shared understanding“ erfolgt. Das entsprechende Ziel wird hierfür als Ziel 1 definiert und lautet wie folgt:

Ziel 1:

Analysiere Vision Videos
zum Zweck der Beurteilung
in Bezug auf ihre Relevanz für die Erzeugung von „shared understanding“
aus der Perspektive der am Projekt beteiligten Stakeholder und Entwickler
im Kontext von Probanden, die verschiedene Gruppen von Stakeholdern und
Entwicklern repräsentieren, denen Vision Videos gezeigt werden, um die
Vision des Systems zu klären, indem sie ihre mentalen Modelle präsentieren,
diskutieren und aneinander anpassen

Das zweite Ziel, welches sich im Kontext dieser Arbeit ableiten lässt, bezieht sich auf die Fragestellung, wie groß der Mehrwert in Bezug auf „shared understanding“ ist, der durch Vision Videos erzeugt werden kann. Dieses Ziel wird dabei unter der Annahme verfolgt, dass „shared understanding“ auf Basis von Vision Videos erzeugt werden kann. Dieses Ziel wird im Kontext dieser Arbeit als Ziel 2 definiert und lautet wie folgt:

Ziel 2:

Analysiere Vision Videos
zum Zweck der Beurteilung
in Bezug auf die Größe ihres Mehrwerts für die Erzeugung von „shared
understanding“
aus der Perspektive der am Projekt beteiligten Stakeholder und Entwickler
im Kontext von Probanden, die verschiedene Gruppen von Stakeholdern und
Entwicklern repräsentieren, denen Vision Videos gezeigt werden, um die
Vision des Systems zu klären, indem sie ihre mentalen Modelle präsentieren,
diskutieren und aneinander anpassen

In den folgenden Unterkapiteln wird die Planung eines entsprechenden Experiments beschrieben, welches diese zwei Ziele umsetzen kann. Da es sich im Fall dieser Arbeit um ein Experimentdesign handelt, welches nicht im Rahmen dieser Arbeit durchgeführt wird, werden an einigen Stellen alternative Vorgehensweisen beschrieben, die abhängig vom Rahmen, in dem das Experiment durchgeführt wird, ausgewählt werden können.

5.2 Kontext

Nachdem die Ziele definiert wurden, die durch das Experiment erreicht werden sollen, ist der erste Schritt der eigentlichen Planung des Experiments die Definition des Kontextes, in dem das Experiment stattfinden sollte. Um durch das Experiment Ergebnisse zu erhalten, die in einem möglichst großen Rahmen verallgemeinerbar sind, sollten Experimente im optimalen Fall in einem großen, echten Softwareprojekt mit professionellen Mitarbeitern durchgeführt werden. Die Durchführung eines solchen „on-line“ Experiments birgt dabei jedoch sowohl für das Experiment selbst als auch für das Softwareprojekt gewisse Risiken. In einem realen Umfeld können unkontrollierte Variablen einen Einfluss auf das Ergebnis des Experiments nehmen. Seitens des Softwareprojekts kann eine neue Methode, die durch das Experiment untersucht werden soll, weniger performant sein als erwartet und dadurch Verzögerungen im Zeitplan des Projekts erzeugen. Eine Alternative hierzu bieten „off-line“ Experimente, die nicht in einem realen Projekt durchgeführt werden. Stattdessen wird ein fiktives Projekt in einer Umgebung durchgeführt, die einfacher zu kontrollieren ist. Durch diese „off-line“ Experimente werden die mit einem „on-line“ Experiment zusammenhängenden Risiken minimiert, jedoch steigen dafür die Kosten des Experiments. Eine Möglichkeit zur Minimierung von Kosten besteht darin, Studenten anstelle von professionellen Mitarbeitern eines Projekts für das Experiment einzusetzen. Experimente mit Studenten sind dabei im Allgemeinen eher auf einen bestimmten Kontext fokussiert, da sie weniger Erfahrung besitzen als professionelle Mitarbeiter eines Projekts. Darüber hinaus werden bei „off-line“ Experimenten mit Studenten oftmals aufgrund ihrer Größenordnung keine realen Probleme betrachtet sondern eher kleinere Probleme, die von Wohlin et al. [51] als „toy-Problem“ bezeichnet werden. All diese verschiedenen Alternativen und Einschränkungen bestimmen zusammen, inwiefern die Ergebnisse des Experiments verallgemeinerbar oder nur auf einen spezifischen Kontext übertragbar sind.

Wie von Wohlin et al. [51] beschrieben, wäre es für dieses Experimentdesign optimal, das Experiment im Kontext verschiedener realer Softwareprojekte durchzuführen. Um jedoch ein repräsentatives Ergebnis zu erhalten, müsste das Experiment in einer Vielzahl von realen Softwareprojekten durchgeführt werden. Da diese Art der Umsetzung des Experiments in einem überschaubaren Zeitrahmen unrealistisch ist, wird an dieser Stelle eine „offline“ Durchführung des Experiments angenommen. Dies erlaubt eine kontrollierte Durchführung des Experiments in einem überschaubaren Zeitrahmen. Darüber hinaus können die Vision Videos, die als Beispiele für das Experiment verwendet werden, durch den Experimentator kontrolliert werden, ohne dass sie dem Kontext eines realen Softwareprojekts entsprechen müssen. Entsprechend bezieht sich der Umfang der jeweiligen Vision Videos eher auf den eines „toy-Problems“ anstatt eines realen Problems. Durch den geringeren Umfang der Vision Videos können im Verhältnis zu einem realen Kontext mehr Vision Videos betrachtet werden, wodurch ein repräsentativeres Ergebnis erzeugt werden kann. Ob für die Durchführung des Experiments professionelle Arbeitskräfte oder Studenten eingesetzt werden, hängt hierbei stark vom Rahmen ab, in dem das Experiment durchgeführt wird. Im Allgemeinen sind hierbei professionelle Arbeitskräfte aufgrund ihrer Erfahrung Studenten vorzuziehen. Da jedoch abhängig von der Anzahl der zu betrachtenden Vision Videos eine Vielzahl von Probanden benötigt werden kann und die Verfügbarkeit von professionellen Arbeitskräften in den meisten Fällen sehr begrenzt ist, kann dies in den wenigsten Fällen umgesetzt werden. Studenten, die Stakeholder oder Entwickler repräsentieren, sind daher im Fall dieses Experimentdesigns eine annehmbare und wahrscheinlich einfacher umsetzbare Alternative für Probanden. Weitere Details zur Probandenauswahl werden diesbezüglich in Kapitel 5.5 beschrieben.

5.3 Hypothesen

Hypothesen bilden die Basis der statistischen Analysen, die anhand der Ergebnisse eines Experiments durchgeführt wird. Das Testen dieser Hypothesen bildet daher einen Kernbestandteil der Durchführung eines Experiments. Aus diesem Grund müssen Hypothesen vor der Durchführung des Experiments klar definiert werden, um sie dann mit den während des Experiments gesammelten Daten entweder zu verifizieren oder zu falsifizieren. Im Fall, dass die Hypothese falsifiziert wird, lassen sich dann weiterführende Schlussfolgerungen über das Experiment ziehen. Eine Hypothese besteht dabei aus zwei Teilhypothesen, einer Nullhypothese (H_0) und einer Alternativhypothese (H_1 beziehungsweise H_a). Die Nullhypothese sagt hierbei im Allgemeinen aus, dass keine signifikanten Trends oder Muster in den Ergebnissen des Experiments erkennbar sind. Die einzigen minimalen Unterschiede, die im Fall einer Nullhypothese zu beobachten sind, werden dabei durch reinen Zufall erzeugt. Das allgemeine Ziel jedes Experiments ist es daher, die Nullhypothese zu Gunsten der Alternativhypothese mit einer möglichst hohen Signifikanz zu falsifizieren. Die Alternativhypothese besitzt hierbei eine Aussage, die gegensätzlich zu der Aussage der Nullhypothese ist.

Um die Hypothesen für diese Arbeit herzuleiten werden an dieser Stelle die Ziele 1 und 2 betrachtet, die in Kapitel 5.1 definiert wurden. Ziel 1 kann hierbei direkt in die folgende Hypothese übertragen werden:

Tabelle 5: Hypothese zum Ziel 1: Erzeugung von „shared understanding“

Ziel	Ziel 1: Erzeugung von „shared understanding“
Nullhypothese	$H_{0.1}$: Die Betrachtung von Vision Videos erzeugt kein „shared understanding“ bei den Probanden
Alternativhypothese	$H_{1.1}$: Die Betrachtung von Vision Videos erzeugt ein messbares „shared understanding“ bei den Probanden

Die Hypothesen zu Ziel 2 müssen etwas differenzierter betrachtet werden. Zum einen kann untersucht werden, inwiefern sich das „shared understanding“ der Probanden verändert hat, nachdem sie das Vision Video gesehen haben. Zum anderen ist jedoch auch noch die Frage interessant, wie groß der Mehrwert für das „shared understanding“ durch Beschreibungen des jeweiligen Softwareprojekts anhand anderer Medien als Vision Videos ist, nachdem die Probanden bereits ein Vision Video des jeweiligen Projekts gesehen haben. Ein Beispiel für ein solches Medium wäre eine textuelle Beschreibung der Vision des Projekts, die im Normalfall in einer Spezifikation enthalten ist. Aus diesen beiden Ansätzen lassen sich daher die folgenden Hypothesen herleiten:

Tabelle 6: Erste Hypothese zum Ziel 2: Mehrwert bezüglich des „shared understandings“

Ziel	Ziel 2: Mehrwert bezüglich des „shared understandings“
Nullhypothese	$H_{0.2}$: Die Betrachtung von Vision Videos erzeugt keinen positiven Mehrwert für das „shared understanding“ der Probanden
Alternativhypothese	$H_{1.2}$: Nach der Betrachtung von Vision Videos ist im Vergleich zu vor der Betrachtung der Vision Videos ein messbarer, positiver Mehrwert bezüglich des „shared understandings“ bei den Probanden feststellbar

Tabelle 7: Zweite Hypothese zum Ziel 2: Mehrwert bezüglich des „shared understandings“

Ziel	Ziel 2: Mehrwert bezüglich des „shared understandings“
Nullhypothese	H _{0,3} : Nach Betrachtung von Vision Videos kann kein Mehrwert durch andere Medien in Bezug auf das „shared understanding“ der Probanden erzeugt werden
Alternativhypothese	H _{1,3} : Andere Medien können das „shared understanding“ von Probanden messbar positiv beeinflussen, nachdem diese bereits ein Vision Video zu dem jeweiligen Projekt gesehen haben

Die Überprüfung, ob „shared understanding“ vorhanden ist beziehungsweise wie groß der Mehrwert von Vision Videos oder anderen Medien für das „shared understanding“ ist, wird in diesem Experimentdesign durch die in der systematischen Literatursuche identifizierten Methoden gemessen. Die genaue Umsetzung der jeweiligen Methoden wird hierbei im Experimentdesign in Kapitel 5.6 beschrieben.

5.4 Variablen

Nachdem die Hypothesen für das Experiment definiert wurden, müssen die Variablen bestimmt werden, die für das Experiment relevant sind. Hierbei wird allgemein zwischen unabhängigen und abhängigen Variablen unterschieden. Unabhängige Variablen sind diejenigen Variablen, die in einem Experiment verändert und kontrolliert werden können. Um die richtigen unabhängigen Variablen zu identifizieren wird dabei in den meisten Fällen spezifisches Domänenwissen benötigt. Die identifizierten unabhängigen Variablen sollten dabei in irgendeiner Form einen Effekt auf die abhängigen Variablen besitzen und müssen daher kontrollierbar sein. Auf der anderen Seite werden durch die abhängigen Variablen die Effekte des Experiments erkennbar. In den meisten Fällen existiert dabei nur eine abhängige Variable, die aus den Hypothesen abgeleitet werden kann. Oft ist diese abhängige Variable nicht direkt messbar und muss deshalb indirekt gemessen werden. Da die jeweilige Metrik hierfür indirekt ist, muss sie sorgfältig validiert werden, damit korrekte Rückschlüsse bezüglich des Ergebnisses des Experiments gezogen werden können. Sowohl unabhängige als auch abhängige Variablen sollten wenn möglich mit den dazugehörigen Wertebereichen und Maßstäben definiert werden. In den beiden folgenden Unterkapiteln werden die konkreten unabhängigen und abhängigen Variablen definiert, die für die Umsetzung dieses Experimentdesigns relevant sind.

5.4.1 Unabhängige Variablen

Die folgende Auflistung von Variablen konnte für dieses Experimentdesign als unabhängige Variablen identifiziert werden:

Unabhängige Variablen:

- Länge der Vision Videos
- Zeit für die Betrachtung der Vision Videos
- Umfang der Visionsbeschreibung des Projekts
- Zeit für das Lesen der Visionsbeschreibung
- Rolle, die von den Probanden repräsentiert wird
- Methode zur Beurteilung des „shared understandings“

An dieser Stelle ist jedoch anzumerken, dass nicht alle der aufgelisteten unabhängigen Variablen verändert werden. Für einige dieser Variablen ist es insbesondere wichtig, dass sie zwischen verschiedenen Beispielen und Probanden konstant bleiben, um die Vergleichbarkeit

der gesammelten Daten zu gewährleisten. Zu diesen Variablen zählen die Länge der Vision Videos, der Umfang der Visionsbeschreibung und die Zeit für das Betrachten der Vision Videos und das Lesen der Visionsbeschreibung. Unter dem Umfang der Visionsbeschreibung ist hierbei eine Menge an Text zu verstehen, die zwischen verschiedenen Visionsbeschreibungen ähnlich groß sein sollte, um einen einheitlichen Informationsgehalt der Visionsbeschreibungen zu gewährleisten.

Die Methoden, die für die Beurteilung des „shared understandings“ der Probanden benötigt werden, entsprechen denen, die im Kapitel 4.5 präsentiert wurden. Um den Umfang dieses Experimentdesigns in einem überschaubaren Rahmen zu halten, werden an dieser Stelle beispielhaft drei verschiedene Methoden gewählt. Hierbei ist anzumerken, dass entsprechend der Schlussfolgerungen in Kapitel 4.6 jede der präsentierten Methoden im Rahmen dieses Experimentdesigns angewendet werden könnte. Die drei Methoden, die für dieses Design gewählt werden, sind die Pathfinder Methode, Card Sorting und die linguistische Analyse auf Basis von Egozentrismus. Diese Methoden wurden an dieser Stelle ausgewählt, da sie sowohl in ihrer Vorgehensweise als auch in ihrer Komplexität sehr verschieden sind. Mit Pathfinder und Card Sorting sind hierbei zwei Mapping Methoden berücksichtigt, die einen sehr großen Unterschied in der Komplexität ihrer Anwendung aufweisen. Im Gegensatz dazu stellt die linguistische Analyse einen rein numerischen Ansatz dar, der das Verhalten zwischen Personen berücksichtigt. Wie genau die Methoden umgesetzt werden wird im Experimentdesign in Kapitel 5.6 beschrieben.

5.4.2 Abhängige Variablen

Als einzige abhängige Variable, die für dieses Experiment relevant ist, kann an dieser Stelle der Grad des „shared understandings“ zwischen den Probanden identifiziert werden. Da das „shared understanding“ ein Konzept ist, welches nicht direkt messbar ist, werden hierfür die Methoden zur Beurteilung von „shared understanding“ herangezogen. Der Wertebereich ist daher abhängig von der Methode, die für die Beurteilung genutzt wird. Im Fall der Pathfinder Methode wird das „shared understanding“ durch einen Wert im Intervall $[0,1]$ repräsentiert, wobei ein Wert von 1 einer vollständigen Übereinstimmung zwischen den jeweiligen Probanden entsprechen würde. Im Fall von Card Sorting liegt das gleiche Intervall vor, jedoch ist hierbei die Bedeutung umgekehrt, sodass ein Wert von 0 vollständiger Übereinstimmung entspricht. Bei der linguistischen Analyse auf Basis von Egozentrismus wird der Wert für das „shared understanding“ im Intervall $[0,12]$ angegeben, wobei hier wie beim Card Sorting ein geringerer Wert höhere Übereinstimmung bedeutet.

5.5 Probandenauswahl

Ein wichtiger Schritt der Planung eines Experiments ist die Wahl der Probanden, mit denen das Experiment letztendlich durchgeführt werden soll. Die richtige Wahl der Probanden ist ein entscheidender Faktor dafür, wie verallgemeinerbar das Ergebnis eines Experiments ist. Um ein möglichst verallgemeinerbares Ergebnis zu erhalten, sollten die Probanden daher so ausgewählt werden, dass sie möglichst repräsentativ für die Population sind, auf die anhand der Ergebnisse Rückschlüsse gezogen werden soll. Die Menge der Probanden wird dabei auch als Stichprobe der Population bezeichnet. Die Wahrscheinlichkeit dafür, ob eine Person als Proband ausgewählt wird, kann dabei entweder bekannt oder unbekannt sein. Abhängig davon können verschiedene Methoden verwendet werden, um die Probanden auszuwählen.

Für die Planung dieses Experimentdesigns gibt es abhängig vom Rahmen, in dem das Experiment später umgesetzt wird, zwei mögliche Personengruppen, aus denen die Probanden

ausgewählt werden können. Die erste Personengruppe bilden professionelle Arbeitskräfte, die aufgrund ihrer Erfahrungen aus verschiedenen Softwareprojekten die optimale Besetzung der Probanden darstellen. Dies bezieht sich hierbei sowohl auf Entwickler, als auch auf Stakeholder, die bereits an einem Softwareprojekt beteiligt waren. Da jedoch die Anzahl der professionellen Arbeitskräfte, die für ein Experiment zur Verfügung stehen, meist sehr begrenzt ist, stellen Studenten eine einfachere Möglichkeit dar, um die Rolle der Probanden zu übernehmen.

Im Kontext von Softwareprojekten und der Verwendung von Vision Videos gibt es grundsätzlich zwei Probandengruppen, die in jedem Fall berücksichtigt werden müssen. Zum einen ist dies die Gruppe der Entwickler. Im Fall, dass professionelle Arbeitskräfte als Probanden vorhanden sind, sollte diese Rolle entsprechend auch von Entwicklern besetzt werden. Für den Fall, dass Studenten als Probanden eingesetzt werden, bieten sich hier Studenten aus Studiengängen an, die einen starken Bezug zu Informatik besitzen. Als Kriterium, ob ein Student als Proband qualifiziert ist, können hier erfolgreich abgeschlossene Veranstaltungen vorausgesetzt werden, die entsprechende Expertise vermitteln, wie beispielsweise das Softwareprojekt.

Die zweite Probandengruppe, die berücksichtigt werden muss, sind die Stakeholder eines Softwareprojekts. Wird das Experiment in einem kleineren Rahmen durchgeführt, dann können diese durch Personen repräsentiert werden, die kein Wissen bezüglich Softwareentwicklung besitzen. Falls das Experiment jedoch in einem größeren Rahmen stattfindet, könnte es interessant sein, verschiedene Gruppen innerhalb der Stakeholder zu identifizieren. Ein Beispiel für verschiedene Gruppen könnten hierbei das Management und die Endnutzer sein. Entsprechend der spezifischen Gruppen von Stakeholdern können hierbei ebenfalls entweder professionelle Arbeitskräfte aus dem jeweiligen Berufsfeld oder Studenten ausgewählt werden. In dem Fall, dass Studenten als Probanden gewählt werden, wäre es auch an dieser Stelle sinnvoll, eine gewisse Expertise vorauszusetzen. Als Beispiel hierfür könnte eine Voraussetzung für die Repräsentation des Managements sein, dass die jeweiligen Studenten aus einem Studiengang kommen, dessen Fokus auf Wirtschaft liegt. Die jeweiligen Stakeholdergruppen, die für das Experiment berücksichtigt werden, hängen dabei stark von den Beispielen ab, die während des Experiments betrachtet werden.

Abhängig davon, wie viele mögliche Probanden zur Verfügung stehen, kann dann eine Methode zur Auswahl der Probanden gewählt werden. In dem unwahrscheinlichen Fall, dass eine zu große Menge an Probanden zur Verfügung steht, bieten sich Methoden an, die zufällig Stichproben wählen. Die einfachste Variante sieht hierbei vor, Probanden zufällig aus der Menge aller möglichen Probanden auszuwählen. Da oftmals jedoch ein Mangel an Probanden besteht, bietet sich das „Quota sampling“ an. Hierbei werden alle einfach verfügbaren Probanden herangezogen, die den Maßstäben des Experiments genügen. Es wird jedoch berücksichtigt, dass eine diverse Menge an Personen gewählt wird, was sich im Fall dieses Experimentdesigns unter anderem auf die verschiedenen beschriebenen Probandengruppen bezieht. Im Allgemeinen gilt dabei, je größer die Anzahl der Probanden ist, desto geringer wird der Fehler bei der Verallgemeinerung der Ergebnisse. Entsprechend sollte auch für dieses Experimentdesign eine möglichst große Menge an Probanden herangezogen werden.

5.6 Experimentdesign

Damit später relevante Schlussfolgerungen anhand der Ergebnisse des Experiments gezogen werden können, muss das Experimentdesign sorgfältig geplant werden. Davon hängt ab, welche statistischen Analysemethoden später auf die während des Experiments gesammelten

Daten angewendet werden können. Ein Experiment kann als Reihe von Tests der Treatments gesehen werden. Ein Treatment ist dabei eine spezifische Ausprägung der veränderbaren unabhängigen Variablen. Um die Auswirkungen eines Treatments auf die abhängigen Variablen eindeutig zu identifizieren, müssen alle anderen unabhängigen Variablen konstant gehalten werden. Da ein Experiment eine Reihe von Tests darstellt, beschreibt das Experimentdesign, wie diese Tests durchgeführt werden. Für die Definition eines Experimentdesigns müssen dabei die zuvor aufgestellten Hypothesen berücksichtigt werden. Dadurch wird gewährleistet, dass durch das Experiment entsprechende statistische Analysen durchführbar sind, anhand derer die Hypothesen entweder verifiziert oder falsifiziert werden können. Um ein qualitativ hochwertiges Experimentdesign zu erzeugen sind einige allgemeine Designprinzipien zu berücksichtigen, die im folgenden Kapitel erläutert und identifiziert werden. Danach wird der Designtyp gewählt, der sich am besten dazu eignet, die für dieses Experimentdesign aufgestellten Hypothesen zu überprüfen.

5.6.1 Allgemeine Designprinzipien

Die drei allgemeinen Designprinzipien, die für ein Experiment relevant sein können, sind Randomization, Blocking und Balancing. Die meisten Experimentdesigns benutzen eine Kombination dieser Prinzipien, um die Qualität des Experiments zu verbessern. Im Folgenden werden die einzelnen Prinzipien erläutert und ihre Ausprägung für dieses Experimentdesign präsentiert.

Als Randomization wird das Prinzip bezeichnet, nach dem die Zuordnung von Probanden und Testobjekten vorgenommen wird, sowie die Definition der Reihenfolge, in der die Tests durchgeführt werden. Eine Gemeinsamkeit, die alle statistischen Methoden besitzen, ist die Voraussetzung, dass die beobachteten Daten von unabhängigen Zufallsvariablen stammen. Um diesen Zufall zu gewährleisten wird Randomization verwendet. Durch Randomization werden daher mögliche unerwünschte Effekte von Variablen ausgeglichen, die sonst das Ergebnis des Experiments beeinflussen würden. Wie in Kapitel 5.5 beschrieben wird daher auch, wenn möglich, die Auswahl der Probanden zufällig bestimmt.

In diesem Experimentdesign kann Randomization an verschiedenen Stellen Anwendung finden. In dem Fall, dass ein Überfluss an Probanden bei der Umsetzung des Experiments vorhanden ist, kann die Auswahl der Probanden zufällig durchgeführt werden. Unabhängig davon sollte jedoch in jedem Fall die Zuweisung der Probanden zu den jeweiligen Treatments des Experiments zufällig geschehen, um eine mögliche Beeinflussung der Ergebnisse des Experiments zu verhindern.

Ein weiteres Designprinzip ist das Blocking. Die Grundidee dabei ist, bekannte kontrollierbare Effekte von Variablen, die für die Ergebnisse des Experiments unerwünscht sind, durch das Experimentdesign zu beseitigen. Dabei werden die betroffenen Treatments in einem Block gesammelt, um die unerwünschten Effekte der Variablen zu isolieren und die Ergebnisse der Treatments innerhalb eines Blocks betrachten zu können. Durch die Verwendung dieser Technik werden die Ergebnisse eines Experiments im Allgemeinen präziser.

Innerhalb dieses Experimentdesigns findet Blocking an zwei verschiedenen Stellen Verwendung. Zum einen werden diejenigen Variablen, die für ein Treatment nicht relevant sind, bewusst konstant gehalten, um die Vergleichbarkeit der Treatments zu gewährleisten. Zum anderen werden den Probanden verschiedene Rollen zugeteilt, um die Ergebnisse im

Kontext dieser Rolle zu betrachten und dadurch mögliche Effekte aufgrund ihrer Expertise zu berücksichtigen.

Das letzte der Designprinzipien ist das Balancing. Beim Balancing geht es darum, dass jedes Treatment gleichermaßen oft verwendet wird, um ein ausgeglichenes Design zu garantieren. Dadurch wird die Durchführung der statistischen Analyse auf den Daten des Experiments sowohl einfacher als auch aussagekräftiger.

Um Balancing in diesem Experimentdesign zu gewährleisten, sollte eine gleiche Anzahl an Probanden für jede der Rollen gewählt werden. Darüber hinaus sollte jedes der Treatments von der gleichen Anzahl an Probanden jeder Rolle angewendet werden und gleichermaßen jeder Proband die gleiche Anzahl von Treatments durchführen. Darüber hinaus sollte im Optimalfall jede der Methoden für die Beurteilung von „shared understanding“ auf jedes Treatment angewendet werden. Das genaue Ausmaß des Balancings dieses Experimentdesigns hängt dabei stark vom Umfang der letztendlichen Umsetzung des Experiments ab.

5.6.2 Wahl des Designtypen

Die Wahl des Designtypen eines Experiments ist ein zentraler Schritt in der Planung des Experiments. Der Designtyp wird dabei anhand der Anzahl der Variablen und Treatments im Experiment bestimmt, sowie der Reihenfolge, in der die Treatments durchgeführt werden. Für dieses Experimentdesign kann als grundlegender Designtyp der Typ „one factor with more than two treatments“ identifiziert werden. Abhängig von der Umsetzung des Experiments kann dieser Designtyp mit dem „randomized complete block design“ erweitert werden.

Der einzelne Faktor, der an dieser Stelle betrachtet wird, ist das „shared understanding“ der Probanden. Um ein repräsentatives Ergebnis zu erhalten, sollten dafür mehr als zwei Treatments verwendet werden. Die Treatments sind dabei die verschiedenen beispielhaften Projekte, die durch die Vision Videos vermittelt werden. Die Anzahl der Treatments kann dabei abhängig vom Rahmen des Experiments beliebig hoch skaliert werden. Bei der Wahl der Treatments ist zu beachten, dass sie eine möglichst ähnliche Komplexität besitzen, damit die Ergebnisse zwischen den Treatments vergleichbar bleiben. Die ähnliche Komplexität umfasst dabei auch die Länge der Vision Videos und den Umfang der Visionsbeschreibung des Projekts aus der Spezifikation.

Das „randomized complete block design“ sieht vor, dass jeder Proband jedes der Treatments in einer zufälligen Reihenfolge umsetzt. Das Ziel hiervon ist es, die Variabilität zwischen den einzelnen Probanden zu minimieren. Die zufällige Reihenfolge verhindert hierbei, dass ein bestimmtes Treatment von einem möglichen Lerneffekt mehr profitiert als andere. Darüber hinaus kann bei einer geringeren Anzahl an Probanden der Umfang der Ergebnismenge deutlich gesteigert werden, wenn ein Proband mehrere Treatments anwendet. Außerdem wird dadurch, dass mehrere Probanden das gleiche Treatment umsetzen, der Fehler minimiert, der bei der Umsetzung eines einzelnen Treatments auftreten kann. Die Verwendung dieser Methode wird an dieser Stelle grundsätzlich als sinnvoll angesehen. In dem Fall, dass eine große Anzahl von Treatments verwendet wird, ist es jedoch nicht möglich, dass jeder Proband jedes Treatment in einem angemessenen Zeitrahmen umsetzt. Daher muss bei der Umsetzung des Experiments der Mehrwert der Verwendung des „randomized complete block designs“ und der Anzahl der umgesetzten Treatments abgewogen werden.

Tabelle 8 zeigt beispielhaft eine mögliche Aufteilung der Treatments auf die einzelnen Probanden. Diese Aufteilung berücksichtigt dabei nur eine der Hypothesen und müsste entsprechend equivalent für jede der Hypothesen umgesetzt werden. Für dieses Beispiel wird angenommen, dass fünf verschiedene Treatments verwendet werden und dass es zeitlich möglich ist, dass ein Proband zwei Treatments durchführt. Ein Kreuz gibt dabei an, dass das jeweilige Treatment vom Probanden umgesetzt wird, während ein Bindestrich angibt, dass der Proband das Treatment nicht durchführt.

Tabelle 8: Beispielhafte Aufteilung der Treatments

Proband	T1	T2	T3	T4	T5
P1	X	X	-	-	-
P2	X	-	X	-	-
P3	X	-	-	X	-
P4	X	-	-	-	X
P5	-	X	X	-	-
P6	-	X	-	X	-
P7	-	X	-	-	X
P8	-	-	X	X	-
P9	-	-	X	-	X
P10	-	-	-	X	X

Die Aufteilung der Treatments in Tabelle 8 entspricht einer minimalen Version, bei der jedes Paar von Treatments genau einmal umgesetzt wird und jedes Treatment genau gleich oft vorhanden ist. Die Reihenfolge, in der die Treatments durchgeführt werden, kann hierbei einen möglichen Lerneffekt bei den Probanden erzeugen. Deshalb sollte jedes Treatment gleichermaßen als erstes beziehungsweise als zweites von den Probanden betrachtet werden, um den durchschnittlichen Lerneffekt gleichmäßig auf alle Treatments zu verteilen. Die einfachste Möglichkeit, um den Lerneffekt auszugleichen, besteht in diesem Fall darin, die doppelte Anzahl an Probanden zu verwenden. Der zweiten Gruppe an Probanden werden die Treatments der einzelnen Treatmentpaare dann in genau umgekehrter Reihenfolge gezeigt, um den durchschnittlichen Lerneffekt auszugleichen.

Bei der Aufteilung der Treatments ist es darüber hinaus wichtig, dass jedes Treatment von einer geraden Anzahl an Personen durchgeführt wird, damit die Methode der linguistischen Analyse auf Basis von Egozentrismus verwendet werden kann. Eine genauere Beschreibung der Umsetzung der Methoden zur Bestimmung von „shared understanding“ für dieses Experimentdesign wird in Kapitel 5.8 präsentiert.

5.7 Instrumentierung

Die Instrumente eines Experiments sind grundsätzlich in die drei Kategorien Objekte, Richtlinien und Messinstrumente einzuteilen. Diese Instrumente sind dabei spezifisch für das Experiment und werden vor der Durchführung entwickelt. Es ist dabei wichtig, dass die Instrumente selbst keinen Einfluss auf das Ergebnis des Experiments besitzen und theoretisch durch ein beliebiges anderes Instrument ersetzbar sind, welches die gleiche Funktionalität erfüllt.

Die Objekte, die für dieses Experimentdesign benötigt werden, sind die Vision Videos sowie die Visionsbeschreibungen für die jeweiligen Beispielprojekte, die als Treatments verwendet werden. Darüber hinaus muss für die Verwendung der Pathfinder Methode, des Card Sortings

und der linguistischen Analyse auf Basis von Egozentrismus die Menge der Konzepte definiert werden, die für das jeweilige Beispielprojekt relevant sind.

Als Richtlinie für die Probanden wird für dieses Experimentdesign ein Dokument benötigt, das den Ablauf des Experiments beschreibt. Durch dieses Dokument sollte den Probanden erklärt werden, was sie im jeweiligen Schritt des Experiments zu tun haben. Es ist dabei empfehlenswert, den Experimentablauf zunächst sehr allgemein zu halten, damit der Fokus der Probanden nicht durch einen Überfluss an Informationen gestört wird. In den einzelnen Schritten des Experiments sollten die Probanden dann eine detaillierte Beschreibung der Aufgabe erhalten, wie sie den jeweiligen Schritt umsetzen sollen.

Bezüglich der Messinstrumente müssen für das Experiment einige Vorlagen erstellt werden, die für die Verwendung der Methoden zur Beurteilung von „shared understanding“ benötigt werden. Für die Pathfinder Methode wird eine Tabelle benötigt, in welche die Probanden das „paired comparison rating“ der jeweiligen Konzeptpaare eintragen können. Für das Card Sorting werden Karten mit den vordefinierten Konzepten benötigt, die durch die Probanden sortiert werden sollen. Für die linguistische Analyse auf Basis von Egozentrismus wird für das Experiment selbst ein beliebiges Gerät benötigt, mit dem der Dialog der Probanden für die spätere Analyse aufgezeichnet werden kann.

5.8 Experimentablauf

Wohlin et al. [51] sehen in ihrer Experimentplanung keinen Schritt für eine detaillierte Beschreibung des Experimentablaufs vor, sondern beschrieben den Ablauf stattdessen in einem separaten Abschnitt zur allgemeinen Durchführung eines Experiments. Um mögliche Missverständnisse in Bezug auf die Wortwahl der „Durchführung“ zu vermeiden, da in dieser Arbeit keine Durchführung des Experiments sondern stattdessen die Definition eines Experimentdesigns stattfindet, wird an dieser Stelle die Beschreibung des Experimentablaufs in die Planung des Experiments integriert. Der Ablauf des Experiments ist davon abhängig, welche der in Kapitel 5.3 definierten Hypothesen überprüft wird und welche Methode für die Beurteilung des „shared understandings“ zwischen den Probanden verwendet wird. Um den Ablauf des Experiments übersichtlicher und verständlicher zu erläutern, werden zunächst die Unterschiede in den Experimentabläufen bezüglich der einzelnen Hypothesen erklärt und danach die Umsetzung der drei Methoden präsentiert.

5.8.1 Hypothesen

Für die Überprüfung der ersten Hypothese, ob durch Vision Videos „shared understanding“ erzeugt wird, wird dem jeweiligen Probanden entsprechend der Zuteilung der Treatments das Vision Video des entsprechenden Beispielprojekts gezeigt. Danach wird anhand der für dieses Experimentdesign ausgewählten Methoden das „shared understanding“ im Vergleich zu anderen Probanden beurteilt, die das gleiche Treatment betrachtet haben. Kann durch die Methoden ein signifikanter Nachweis für ein „shared understanding“ erbracht werden, dann lässt sich die Nullhypothese $H_{0.1}$ falsifizieren.

Um die zweite Hypothese bezüglich des positiven Mehrwerts von Vision Videos auf das „shared understanding“ der Probanden zu überprüfen, muss zunächst eine Basis von „shared understanding“ zwischen den Probanden bestimmt werden. Ausgehend von dieser Basis lässt sich dann bestimmen, inwiefern Vision Videos einen positiven Mehrwert erzeugen können. Um diese Basis zu erzeugen wird den Probanden zunächst die textuelle Visionsbeschreibung des Projekts vorgelegt. Nachdem die Probanden diese Visionsbeschreibung gelesen haben,

wird das „shared understanding“ mittels der drei ausgewählten Methoden beurteilt. Die Ergebnisse dieser Beurteilung dienen im Folgenden als Basis. Nachdem bei den Probanden eine Basis von „shared understanding“ bestimmt wurde, wird den Probanden das zum Beispielprojekt passende Vision Video präsentiert. Nachdem die Probanden das Vision Video verinnerlicht haben, wird abermals das „shared understanding“ beurteilt. Die Ergebnisse dieser Beurteilung lassen sich dann mit der Basis vergleichen, die vor der Betrachtung des Vision Videos bestimmt wurde, um den Mehrwert der Vision Videos für das „shared understanding“ der Probanden zu bestimmen. Durch diesen Vergleich lassen sich dann Rückschlüsse auf die Nullhypothese $H_{0,2}$ ziehen.

Die dritte Hypothese, die sich auf den Mehrwert anderer Medien für das „shared understanding“ der Probanden bezieht, nachdem die Probanden bereits ein entsprechendes Vision Video gesehen haben, hat starke Ähnlichkeit mit der zweiten Hypothese. Das Vorgehen im Ablauf des Experiments ist entsprechend equivalent zu dem, der für die zweite Hypothese beschrieben wurde, mit dem einzigen Unterschied, dass die Reihenfolge, in der die Medien präsentiert werden, getauscht wird. Im Fall der dritten Hypothese dient also das Vision Video des jeweiligen Beispielprojekts als Basis des „shared understandings“ zwischen den Probanden, während die Visionsbeschreibung nach dem Vision Video betrachtet und auf die Erzeugung eines potentiellen Mehrwerts geprüft wird. Entsprechend lassen sich aus diesem Vorgehen Rückschlüsse auf die Nullhypothese $H_{0,3}$ ziehen.

5.8.2 Methoden

Die drei Methoden, die in Kapitel 5.4.1 für dieses Experimentdesign ausgewählt werden, sind die Pathfinder Methode, das Card Sorting und die linguistische Analyse aus Basis von Egozentrismus. Wenn der Rahmen, in dem dieses Experimentdesign umgesetzt wird, groß genug ist, sollte im Optimalfall für jedes Treatments jede der Methoden angewendet werden. Dadurch würde die größtmögliche Menge an Ergebnissen aus dem Experiment gewonnen werden, was wiederum die Verallgemeinerbarkeit der Ergebnisse verbessert. In einem kleineren Rahmen kann entweder die Menge der Methoden reduziert werden oder die Verwendung der Methoden auf verschiedene Probanden gleichmäßig verteilt werden. Im Folgenden wird die Umsetzung der jeweiligen Methoden beschrieben, nachdem der jeweilige Proband entweder ein Vision Video gesehen oder eine Visionsbeschreibung gelesen hat.

Für die Durchführung der Pathfinder Methode wird dem Probanden eine leere Tabelle vorgelegt, die symmetrisch in den Zeilen und Spalten die Konzepte enthält, die für das jeweilige Vision Video beziehungsweise für die jeweilige Visionsbeschreibung relevant sind. Ein Beispiel für eine solche Tabelle befindet sich in Anhang G. Die Aufgabe des Probanden besteht darin, die Konzeptpaare der Tabelle entsprechend ihres Zusammenhangs im Beispielprojekt zu bewerten. Die Bewertung wird dabei mit einem ganzzahligen Wert im Intervall $[0,10]$ vorgenommen, wobei ein größerer Wert einen stärkeren Zusammenhang repräsentiert. Diese Tabelle kann dann während der Analyse des Experiments entsprechend der Beschreibung der Pathfinder Methode in Kapitel 4.5.1 in ein Pathfindernetz übertragen werden, welches dann mit anderen Pathfindernetzen verglichen werden kann. Durch den Vergleich der Pathfindernetze lässt sich dann das „shared understanding“ zwischen zwei beliebigen Probanden bewerten, die diese Methode durchgeführt haben.

Für die Verwendung des Card Sortings wird den Probanden eine Menge an vordefinierten Karten gegeben, auf denen die relevanten Konzepte des Beispielprojekts geschrieben stehen. Die Probanden werden dann dazu aufgefordert, die ihrer Meinung nach zusammengehörigen Karten auf einen Stapel zu sortieren. Die daraus resultierenden Stapel werden vom

Experimentator dokumentiert und in der darauf folgenden Analyse entsprechend der Beschreibung der Card Sorting Methode in Kapitel 4.5.1 mit den Ergebnissen anderer Probanden verglichen, um das „shared understanding“ zwischen den jeweiligen Probanden zu bestimmen.

Die linguistische Analyse auf Basis von Egozentrismus ist im Vergleich zu den anderen hier beschriebenen Methoden etwas aufwändiger. Da die linguistische Analyse auf Basis eines Dialogs aufgebaut ist, der von zwei Probanden geführt wird, muss das Experiment mit einem Paar von Probanden anstelle eines einzelnen Probanden durchgeführt werden. Hierfür wird den beiden Probanden jeweils das gleiche Treatment präsentiert. Danach werden die Probanden dazu aufgefordert, durch einen Dialog ihr Verständnis bezüglich der Konzepte des Beispielprojekts mit dem jeweils anderen Probanden zu teilen und zu diskutieren. Der Dialog wird dabei durch den Experimentator aufgezeichnet. Die Konzepte, über welche die Probanden diskutieren sollen, sind dabei abhängig vom Beispielprojekt und werden durch den Experimentator vorgegeben. Abgesehen davon bleibt das Gespräch unmoderiert, um einen möglichen Einfluss des Experimentators auf die Ergebnisse auszuschließen. Während der Analyse wird dann die Aufzeichnung des Dialogs entsprechend der in Kapitel 4.5.2 beschriebenen Methode auf Egozentrismus überprüft. Entsprechend dieser Analyse lässt sich dann das „shared understanding“ der Probanden beurteilen, die den Dialog geführt haben. Ein Nachteil hierbei ist, dass im Gegensatz zu den anderen beiden Methoden nur das „shared understanding“ zwischen den beiden Probanden bestimmt werden kann, die zusammen an einem Dialog beteiligt waren.

Abhängig davon, wie komplex die Auswertung der Ergebnisse erfolgen und wie differenziert die Hypothesen verifiziert oder falsifiziert werden sollen, können die Ergebnisse der Probanden entweder innerhalb von Probandengruppen oder gruppenübergreifend analysiert werden. Darunter ist zu verstehen, dass beispielsweise entweder nur das „shared understanding“ zwischen Stakeholdern betrachtet wird oder gruppenübergreifend das „shared understanding“ zwischen Stakeholdern und Entwicklern betrachtet wird. Da in Softwareprojekten im Allgemeinen eine Vielzahl verschiedener Personengruppen beteiligt sind, ist es an dieser Stelle empfehlenswert, neben der gruppeninternen auch eine gruppenübergreifende Analyse der Ergebnisse vorzunehmen.

5.9 Threats to Validity

Wie bei der systematischen Literatursuche müssen auch bei einem Experiment die „Threats to Validity“ identifiziert werden, damit ein möglichst aussagekräftiges Ergebnis aus dem Experiment hervorgehen kann. Da das allgemeine Konzept der „Threats to Validity“ bereits in Kapitel 4.7 beschrieben ist, wird an dieser Stelle auf eine allgemeine Erläuterung der einzelnen Aspekte verzichtet. Stattdessen werden in den folgenden Unterkapiteln mögliche „Threats to Validity“ identifiziert, die in diesem konkreten Experimentdesign auftreten können. Wohil et al. [51] präsentieren hierfür eine Menge typischer „Threats to Validity“, anhand der sich die folgenden Unterkapitel orientieren.

5.9.1 Construct Validity

Für die Construct Validity ist es wichtig, dass das durch das Experiment betrachtete Konzept durch die Treatments ausreichend repräsentiert wird. Ein Problem, was in diesem Kontext auftreten kann, ist, dass zu wenig Treatments innerhalb des Experiment verwendet werden, um das Konzept zu repräsentieren. Dies kann zur Folge haben, dass das Experiment zu abhängig von einem einzelnen Treatment ist, wodurch das Ergebnis verfälscht wird. Aus

diesem Grund wird in diesem Experimentdesign empfohlen, entsprechend der Anzahl der verfügbaren Probanden, eine möglichst große Menge verschiedener Treatments zu betrachten. Ein ähnliches Problem tritt bei der Messung der Ergebnisse auf. Wird die Messung der Ergebnisse mit nur einer Methode vorgenommen, dann ist es möglich, dass durch die Methode das Ergebnis des Experiments beeinflusst wird. Aus diesem Grund wird in diesem Experimentdesign die Verwendung von drei verschiedenen Methoden für die Beurteilung des „shared understandings“ empfohlen, die jeweils unterschiedliche Ansätze verfolgen.

Eine weitere Quelle für verschiedene Probleme stellen die sozialen Aspekte innerhalb eines Experiments dar. Eines dieser Probleme ist das sogenannte „Hypothesis guessing“, welches die Angewohnheit beschreibt, dass Probanden versuchen, den Grund und die Absicht für das Experiment herauszufinden. Abhängig von ihren Vermutungen über die Hypothese kann sich dann ihr Verhalten während des Experiments sowohl positiv als auch negativ verbessern, abhängig davon, wie ihre Einstellung zu der vermuteten Hypothese ist. Ein Experimentator sollte daher während eines Experiments versuchen, den Fokus der Probanden auf das Experiment selbst zu richten und keine überflüssigen Informationen bezüglich der Absicht des Experiments an die Probanden weiterzuleiten.

Ein weiteres soziales Problem ist, dass einige Personen Angst davor haben, bewertet zu werden. Diese Angst drückt sich dann im Experiment so aus, dass sie versuchen ein Ergebnis zu erzielen, welches sie besser aussehen lässt, unabhängig davon, ob es ihren wahren Überzeugungen entspricht. Durch diese fehlerhaften Daten wird letztendlich das Ergebnis des Experiments verzerrt, was zu falschen Rückschlüssen über die Hypothese führen kann. Um dem entgegenzuwirken sollte der Experimentator den Probanden zu Beginn des Experiments vermitteln, dass durch das Experiment keine persönliche Bewertung der Probanden stattfindet. Darüber hinaus sollte den Probanden erklärt werden, dass jegliche durch die Probanden produzierten Daten hilfreich sind und die Ausprägung dieser Daten unerheblich für die Güte der Ergebnisse ist.

Neben den Probanden kann jedoch auch durch den Experimentator ein Problem für die „Construct Validity“ entstehen. Es ist möglich, dass der Experimentator die Ergebnisse des Experiment abhängig von seinen eigenen Erwartungen absichtlich oder unabsichtlich beeinflusst. Aus diesem Grund sollte das Experiment im Optimalfall von mindestens zwei verschiedenen Experimentatoren überwacht werden, die entweder gegensätzliche oder gar keine Erwartungen bezüglich des Ergebnisses besitzen.

5.9.2 Internal Validity

Der Thematik der „Internal Validity“ lassen sich verschiedene Probleme zuordnen, die durch den Verlauf des Experiments entstehen. Eines dieser Probleme ist, dass sich das Verhalten der Personen abhängig von der Dauer des Experiments verändern kann. Es kann dabei sowohl zu positiven als auch negativen Einflüssen auf das Ergebnis kommen, wie beispielsweise Lerneffekte und Erschöpfung. Um diesen Problemen entgegenzuwirken sollte der zeitliche Umfang des Experiments auf einen angemessenen Rahmen begrenzt werden, um die Konzentration der Probanden zu gewährleisten. Darüber hinaus kann ein möglicher Lerneffekt gleichmäßig auf alle Treatments verteilt werden, indem die Reihenfolge, in der die Treatments durchgeführt werden, für jeden Probanden verändert wird.

Ein anderes Problem ist die Varianz der menschlichen Leistung. Abhängig davon, welche Personen als Probanden ausgewählt werden, kann selbst bei gleicher Qualifikation der Personen ein Unterschied in ihrer Leistung während des Experiments vorhanden sein. Entsprechend kann abhängig von den ausgewählten Personen das Ergebnis variieren. Dieses Problem kann durch eine größere Stichprobe von Probanden minimiert werden.

Dementsprechend sollte bei der Umsetzung dieses Experimentsdesigns das Experiment mit einer möglichst großen Anzahl von Probanden durchgeführt werden.

Wie bei der „Construct Validity“ gibt es auch bei der „Internal Validity“ einige soziale Probleme, die im Rahmen eines Experiments auftreten können. Eines dieser Probleme das für dieses Experimentdesign zutrifft ist, dass Probanden andere Probanden mit gleichen Treatments eventuell imitieren wenn sie wissen, welche Ergebnisse die anderen Probanden produziert haben. Dieses Verhalten verfälscht die Daten eines Experiments und damit dessen Ergebnis immens. Aus diesem Grund muss der Experimentator verhindern, dass die Probanden untereinander Informationen austauschen. Darüber hinaus darf der Experimentator die von anderen Probanden produzierten Daten nicht preisgeben.

5.9.3 External Validity

Ein Problem der „External Validity“ für ein Experiment ist die Auswahl der Probanden. Bei der Auswahl der Probanden muss berücksichtigt werden, dass eine Gruppe von Probanden gewählt wird, die repräsentativ für die relevante Population ist. Falls dies nicht der Fall ist, dann ist das Ergebnis nicht repräsentativ für die Population und kann somit nicht verallgemeinert werden. Um diesem Problem entgegenzuwirken wird in diesem Experimentdesign vorgesehen, dass die Probanden entsprechend der Personengruppen ausgewählt werden, die an einem Softwareprojekt beteiligt sind. Darüber hinaus ist im Design vorgesehen, dass bei der Durchführung des Experiments mit Studenten ein gewisser Standard an Erfahrung vorausgesetzt wird, damit die Studenten als Probanden für das Experiment zulässig sind.

5.9.4 Conclusion Validity

Ein grundlegendes Problem bei der Durchführung von Experimenten ist die Größe der Stichprobe. Da die Stichprobe der Daten, die durch ein Experiment überprüft werden, im Allgemeinen repräsentativ für eine größere Population sein soll, muss die Stichprobe möglichst groß sein, um eine bessere Verallgemeinerbarkeit zu gewährleisten. Aus diesem Grund ist es für die Umsetzung dieses Experimentdesigns wichtig, dass eine möglichst große Menge an Probanden herangezogen wird, um die Hypothesen zu überprüfen.

Ein weiteres Problem, was unbedingt verhindert werden muss, ist das sogenannte „Fishing“. Das Fishing stellt ein Problem dar, bei dem ein Experimentator nach einem bestimmten Ergebnis sucht, wodurch absichtlich oder unabsichtlich Einfluss auf das Ergebnis des Experiments ausgeübt wird. Aus diesem Grund ist es für die Analyse der Ergebnisse des Experimentss wichtig, dass vor der Durchführung des Experiments eine oder mehrere der in der Experimentplanung vorgestellten alternativen Methoden für die Bewertung von „shared understanding“ ausgewählt werden.

Der letzte wichtige Aspekt bezüglich der Construct Validity ist, dass das Experiment für jeden Probanden unter den gleichen Bedingungen stattfindet, um mögliche unberücksichtigte Variablen ausschließen zu können. Hierfür muss gewährleistet werden, dass die Durchführung der Treatments für jeden Probanden gleich abläuft und keine Abweichungen vom geplanten Experimentdesign vorgenommen werden. Darüber hinaus sollte das Experiment in einer möglichst neutralen Umgebung durchgeführt werden, in der keine externen Reize einen Einfluss auf das Ergebnis des Experiments haben könnten. Beispiele für solche externen Reize sind hierbei ablenkende Geräusche außerhalb des Raums oder starke visuelle Reize im Raum, die den Fokus der Probanden auf sich ziehen.

6 Fazit und Ausblick

In diesem Kapitel wird abschließend das Ziel dieser Arbeit anhand der Ergebnisse, die im Verlauf dieser Arbeit entstanden sind, in einem Fazit geklärt. Darüber hinaus werden in einem Ausblick mögliche Ansätze für weiterführende Arbeiten beschrieben, die auf den Ergebnissen dieser Arbeit aufbauen könnten.

6.1 Fazit

Die initiale Motivation dieser Arbeit war, dass die Kommunikation zwischen Stakeholdern und Entwicklern einen der Kernaspekte darstellt, die für die erfolgreiche Durchführung eines Softwareprojekts gewährleistet sein müssen. Diese Kommunikation ist in diesem Zusammenhang entscheidend, um ein gemeinsames Verständnis zwischen Stakeholdern und Entwicklern zu erzeugen. Jedoch besteht oftmals das Problem, dass diese Kommunikation durch eine Sprachbarriere zwischen Stakeholdern und Entwicklern gestört wird, die durch das verschiedene Domänenwissen der jeweiligen Personengruppe entsteht. In der Literatur schlagen daher diverse Quellen die Nutzung von Vision Videos vor, um die Kommunikation zwischen Stakeholdern und Entwicklern zu unterstützen und dadurch ein gemeinsames Verständnis zu erzeugen. Jedoch wurden bisher noch keine allgemeingültigen Methoden definiert, mit denen das gemeinsame Verständnis gemessen oder beurteilt werden kann, das auf Basis von Vision Videos erzeugt wurde. In dieser Arbeit wird dieses Problem aufgegriffen und mittels einer systematischen Literatursuche werden Methoden, Metriken und Heuristiken für die Messung und Beurteilung von gemeinsamen Verständnis identifiziert, das auf Basis von Vision Videos erzeugt wurde.

Aus der auf drei verschiedenen Suchmethoden aufbauenden systematischen Literatursuche, die zu diesem Zweck durchgeführt wurde, konnte eine Menge von 101 Publikationen identifiziert werden, die potentielle Methoden, Metriken oder Heuristiken für die Messung oder Beurteilung von gemeinsamen Verständnis thematisieren. Im Zuge dessen konnten sowohl die Effektivität als auch die Effizienz der verwendeten Suchmethoden bestätigt werden. Darüber hinaus konnte anhand der Ergebnismenge der Zeitrahmen überprüft werden, der für die Thematik dieser Arbeit relevant ist. Dieser deckt sich dabei mit dem eingangs während der manuellen Suche angenommenen Zeitrahmen.

In Kapitel 4.5 konnten darauf hin einige allgemein verwendbare Methoden und Metriken identifiziert werden, die sich in Mapping Konzepte und numerische Konzepte kategorisieren lassen. Die Mapping Konzepte basieren dabei auf einer visuellen Darstellung, während die numerischen Konzepte auf Metriken basieren, die grundsätzlich keine visuelle Darstellung benötigen. Die entsprechenden Methoden und Metriken wurden dabei jeweils im Detail erläutert. Es konnten jedoch keine Heuristiken identifiziert werden, die in diesem Zusammenhang allgemein relevant sind. Dies könnte damit zu begründen sein, dass „shared understanding“ ein komplexes Konzept darstellt, das durch die Beteiligung verschiedenster Personen von diversen Faktoren abhängen kann, die nicht durch eine allgemeingültige Heuristik verallgemeinerbar sind. Anhand der verschiedenen Methoden und Metriken konnte geschlussfolgert werden, dass sich alle der identifizierten Methoden und Metriken auf den Kontext von Vision Videos übertragen lassen, da die Methoden und Metriken alle das „shared understanding“ unabhängig vom Medium bestimmen, durch das es erzeugt wurde.

Auf dieser Erkenntnis aufbauend wurde in Kapitel 5 ausführlich ein Experimentdesign definiert, das beispielhaft drei konzeptionell verschiedene Ansätze zur Beurteilung und Messung von „shared understanding“ verwendet, um drei verschiedene Hypothesen bezüglich der Wirkung von Vision Videos in Bezug auf „shared understanding“ zu überprüfen. Das

Experimentdesign zeigt dabei verschiedene Alternativen auf, die bei der Umsetzung des Experiments verwendet werden können.

Abschließend lässt sich zusammenfassen, dass diverse Methoden und Metriken identifiziert wurden, die allgemeingültig zur Messung und Beurteilung von „shared understanding“ verwendet werden können, das auf Basis von Vision Videos erzeugt wurde. Es konnten in diesem Zusammenhang verschiedene experimentelle Ansätze definiert werden, mit denen eine solche Messung und Beurteilung von „shared understanding“ möglich ist.

6.2 Ausblick

Die in dieser Arbeit identifizierten Methoden und Metriken können verwendet werden, um in anderen Arbeiten im Kontext der Nutzung von Vision Videos das „shared understanding“ der beteiligten Personen zu messen und zu beurteilen. Da jedoch aus dieser Arbeit hervorgegangen ist, dass die identifizierten Methoden und Metriken unabhängig vom Medium verwendbar sind, mit dem das „shared understanding“ erzeugt wurde, ist auch die Nutzung in einem Kontext außerhalb von Vision Videos denkbar. Die Ausführungen dieser Arbeit können dabei als Referenz für der Nutzung dieser Methoden und Metriken verwendet werden, die jedoch abhängig vom jeweiligen Kontext gegebenenfalls angepasst werden müssen.

Einen anderer Ansatzpunkt für auf dieser Arbeit aufbauende Arbeiten stellt das Experimentdesign dar, was in Kapitel 5 dieser Arbeit definiert wurde. Die Durchführung des Experiments kann dabei verschiedene Ziele verfolgen. Zum einen können durch die Durchführung die definierten Hypothesen bezüglich des Mehrwerts, den Vision Videos auf das „shared understanding“ zwischen Personen ausüben, überprüft werden. Zum anderen kann durch die Durchführung des Experiments in einem entsprechend großen Rahmen statistisch nachgewiesen werden, dass sich die in dieser Arbeit identifizierten Methoden und Metriken zu Messung und Beurteilung von „shared understanding“ auf den Kontext von Vision Videos übertragen lassen. Darüber hinaus könnte durch die Durchführung des Experiments mit den verschiedenen Methoden und Metriken, die im Rahmen dieser Arbeit identifiziert wurden, untersucht werden, wie effektiv die Umsetzungen der jeweiligen Methoden und Metriken in der Realität sind.

Im Rahmen dieser Experimente könnte es dabei auch interessant sein, die Experimente in realen Softwareprojekten umzusetzen. Dies hätte den Vorteil, dass sowohl bezüglich der Hypothesen im Zusammenhang mit dem Mehrwert von Vision Videos als auch bezüglich der jeweils verwendeten Methode oder Metrik zur Beurteilung von „shared understanding“ die Ergebnisse von professionellen Arbeitskräften gesammelt werden könnten. Diese Ergebnisse könnten dann mit den Ergebnissen von Experimenten verglichen werden, bei denen Studenten als Probanden fungiert haben. Aus diesem Vergleich könnten mögliche Unterschiede identifiziert werden, die im Rahmen der realen Nutzung von Vision Videos auftreten können.

Anhang

A) Liste des Quasi-Goldstandards und der hilfreichen Publikationen

Quasi-Goldstandard:

1. Jia, Jingdong, and Luiz Fernando Capretz. "Direct and mediating influences of user-developer perception gaps in requirements understanding on user participation." *Requirements Engineering* 23.2 (2018): 277-290.
2. Bjarnason, Elizabeth, and Helen Sharp. "The role of distances in requirements communication: a case study." *Requirements Engineering* 22.1 (2017): 1-26.
3. Fricker, Samuel A., et al. "Workshop videos for requirements communication." *Requirements Engineering* 21.4 (2016): 521-552.
4. Aranda, Gabriela N., Aurora Vizcaíno, and Mario Piattini. "A framework to improve communication during the requirements elicitation process in GSD projects." *Requirements engineering* 15.4 (2010): 397-417.
5. Knauss, Eric, et al. "Patterns of continuous requirements clarification." *Requirements Engineering* 20.4 (2015): 383-403.
6. Bjarnason, Elizabeth, Helen Sharp, and Björn Regnell. "Improving requirements-test alignment by prescribing practices that mitigate communication gaps." *Empirical Software Engineering* 24.4 (2019): 2364-2409.
7. Yu, Xiaodan, and Stacie Petter. "Understanding agile software development practices using shared mental models theory." *Information and software technology* 56.8 (2014): 911-921.
8. Di Ruscio, Davide, et al. "Collaborative model-driven software engineering: a classification framework and a research map." *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*. IEEE, 2018.

Hilfreiche Publikationen:

1. Figl, Kathrin, and Jan Recker. "Exploring cognitive style and task-specific preferences for process representations." *Requirements Engineering* 21.1 (2016): 63-85.
2. Hadar, Irit, Pnina Soffer, and Keren Kenzi. "The role of domain knowledge in requirements elicitation via interviews: an exploratory study." *Requirements Engineering* 19.2 (2014): 143-159.
3. Salay, Rick, et al. "Managing requirements uncertainty with partial models." *Requirements Engineering* 18.2 (2013): 107-128.
4. Sutcliffe, Alistair, Sarah Thew, and Paul Jarvis. "Experience with user-centred requirements engineering." *Requirements Engineering* 16.4 (2011): 267-280.
5. Liaskos, Sotirios, et al. "Representing and reasoning about preferences in requirements engineering." *Requirements Engineering* 16.3 (2011): 227.
6. Khatwani, Charu, et al. "Advancing viewpoint merging in requirements engineering: a theoretical replication and explanatory study." *Requirements Engineering* 22.3 (2017): 317-338.
7. Calefato, Fabio, Daniela Damian, and Filippo Lanubile. "Computer-mediated communication to support distributed requirements elicitation and negotiations tasks." *Empirical Software Engineering* 17.6 (2012): 640-674.
8. de la Vara, Jose Luis, et al. "An Empirical Evaluation of the Use of Models to Improve the Understanding of Safety Compliance Needs." *Information and Software Technology* (2020): 106351.

9. Trkman, Marina, et al. "Impact of the conceptual model's representation format on identifying and understanding user stories." *Information and software technology* 116 (2019): 106169.
10. Siqueira, Fábio Levy. "Comparing the comprehensibility of requirements models: An experiment replication." *Information and Software Technology* 96 (2018): 1-13.
11. Dikici, Ahmet, Oktay Turetken, and Onur Demirors. "Factors influencing the understandability of process models: A systematic literature review." *Information and Software Technology* 93 (2018): 112-129.
12. Moe, Nils Brede, Aybüke Aurum, and Tore Dybå. "Challenges of shared decision-making: A multiple case study of agile software development." *Information and Software Technology* 54.8 (2012): 853-865.
13. Bidlake, Leah, Eric Aubanel, and Daniel Voyer. "Systematic literature review of empirical studies on mental representations of programs." *Journal of Systems and Software* (2020): 110565.
14. Bernardi, Mario Luca, et al. "The relation between developers' communication and fix-Inducing changes: An empirical study." *Journal of Systems and Software* 140 (2018): 111-125.
15. Barbour, Joshua B., Rebecca Gill, and J. Kevin Barge. "Organizational communication design logics: A theory of communicative intervention and collective communication design." *Communication Theory* 28.3 (2018): 332-353.
16. Chen, Di, et al. "Applications of psychological science for actionable analytics." *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 2018.

B) Kombinationen der Suchbegriffe für SpringerLink

Tabelle 9: Kombination der Suchbegriffe für die Titelsuche mit SpringerLink

shared understanding	32	perception gap	9
mutual understanding	24	comprehension gap	0
common understanding	15	communication gap	0
differing understanding	0	cognitive gap	4
varying understanding	0	mental gap	1
lack of understanding	2		
poor understanding	4	perception distance	1
		comprehension distance	0
shared comprehension	0	communication distance	6
mutual comprehension	0	cognitive distance	5
common comprehension	0	mental distance	1
differing comprehension	0		
varying comprehension	0	shared model	4
lack of comprehension	0	mutual model	0
poor comprehension	1	mental model	94
		cognitive model	161
shared viewpoint	0	internal model	126
mutual viewpoint	0	partial model	16
common viewpoint	1		
differing viewpoint	0	shared representation	7
varying viewpoint	0	mutual representation	0
lack of viewpoint	0	mental representation	77
poor viewpoint	0	cognitive representation	15
		internal representation	43
		partial representation	4
		Gesamt	653

C) Angepasste Suchstrings für die jeweilige Syntax der Datenbanken

Science Direct:

Aufgrund von Operatorenbegrenzung aufgeteilt in sieben einzelne Suchstrings

- 1)
(shared OR mutual OR common)AND (understanding OR comprehension OR viewpoint)AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 2)
(differing OR varying) AND (understanding OR comprehension OR viewpoint) AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 3)
(„lack of“ OR poor) AND (understanding OR comprehension OR viewpoint) AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 4)
(perception OR comprehension OR communication) AND (gap OR distance) AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 5)
(cognitive OR mental) AND (gap OR distance) AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 6)
(shared OR mutual OR mental) AND (model OR representation)AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)
- 7)
(cognitive OR internal OR partial) AND (model OR representation) AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)

ACM Digital Library:

((shared OR mutual OR common OR differing OR varying OR „lack of“ OR poor) AND (understanding OR comprehension OR viewpoint))
OR ((perception OR comprehension OR communication OR cognitive OR mental) AND (gap OR distance))
OR ((shared OR mutual OR mental OR cognitive OR internal OR partial) AND (model OR representation)))
AND („Requirements Engineering“ OR „Software Engineering“ OR „Human Computer Interaction“)

IEEE Xplore:

((("Document Title": "shared" OR "Document Title": "mutual" OR "Document Title": "common" OR "Document Title": "differing" OR "Document Title": "varying" OR "Document Title": "lack of" OR "Document Title": "poor") AND ("Document Title": "understanding" OR "Document Title": "comprehension" OR "Document Title": "viewpoint"))
OR
(("Document Title": "perception" OR "Document Title": "comprehension" OR "Document Title": "communication" OR "Document Title": "cognitive" OR "Document Title": "mental") AND ("Document Title": "gap" OR "Document Title": "distance"))
OR
(("Document Title": "shared" OR "Document Title": "mutual" OR "Document Title": "mental" OR "Document Title": "cognitive" OR "Document Title": "internal" OR "Document Title": "partial") AND ("Document Title": "model" OR "Document Title": "representation") AND ("Full Text .AND. Metadata": "Requirements Engineering" OR "Full Text .AND. Metadata": "Software Engineering" OR "Full Text .AND. Metadata": "Human Computer Interaction"))

Web of Science:

TI=(((shared OR mutual OR common OR differing OR varying OR "lack of" OR poor) AND (understanding OR comprehension OR viewpoint))
OR ((perception OR comprehension OR communication OR cognitive OR mental) AND (gap OR distance))
OR ((shared OR mutual OR mental OR cognitive OR internal OR partial) AND (model OR representation)))

AND
(TI=("Requirements Engineering" OR "Software Engineering" OR "Human Computer Interaction") OR AB=("Requirements Engineering" OR "Software Engineering" OR "Human Computer Interaction") OR WC=("Requirements Engineering" OR "Software Engineering" OR "Human Computer Interaction") OR AK=("Requirements Engineering" OR "Software Engineering" OR "Human Computer Interaction"))

Springer Link:

Siehe Tabelle 9 Anhang B

D) Liste relevanter Publikationen aus der systematischen Literatursuche

1. Schaffernicht, Martin, and Stefan N. Groesser. "A comprehensive method for comparing mental models of dynamic systems." *European Journal of Operational Research* 210.1 (2011): 57-67.
2. Grenier, Robin S., and Dana Dudzinska-Przesmitzki. "A conceptual model for eliciting mental models using a composite methodology." *Human Resource Development Review* 14.2 (2015): 163-184.
3. Markíczy, Livia, and Jeff Goldberg. "A method for eliciting and comparing causal maps." *Journal of management* 21.2 (1995): 305-333.
4. Brooks, Andrew, and Louise Scott. "A Methodology from Software Engineering Inspection which Supports Replicable Mental Models Research." 6th IEEE International Conference on Cognitive Informatics. IEEE, 2007.
5. Bjarnason, Elizabeth, et al. "A theory of distances in software engineering." *Information and Software Technology* 70 (2016): 204-219.
6. Cooke, Nancy J., et al. "Advances in measuring team cognition." (2004).
7. Rauterberg, Matthias. "AMME: an Automatic Mental Model Evaluation to analyse user behaviour traced in a finite, discrete state space." *Ergonomics* 36.11 (1993): 1369-1380.
8. Memon, Tasneem, Jie Lu, and Farookh Khadeer Hussain. "An enhanced mental model elicitation technique to improve mental model accuracy." *International Conference on Neural Information Processing*. Springer, Berlin, Heidelberg, 2013.
9. Fox, Grace, et al. "Are we on the same page? Exploring stakeholders' shared mental models of mobile health in rural Nigeria." *Health informatics journal* (2020): 1460458220909715.
10. Diesner, Jana, and Kathleen M. Carley. *AutoMap 1.2: extract, analyze, represent, and compare mental models from texts*. Pittsburgh: Carnegie Mellon University, School of Computer Science, Institute for Software Research International, 2004.
11. Kiekel, Preston A., et al. "Automating measurement of team cognition through analysis of communication data." *Usability evaluation and interface design* (2001): 1382-1386.
12. Seeber, Isabella, et al. "Brainstorming is just the beginning: Effects of convergence techniques on satisfaction, perceived usefulness of moderation, and shared understanding in teams." 2015 48th Hawaii International Conference on System Sciences. IEEE, 2015.
13. Bravo-Lillo, Cristian, et al. "Bridging the gap in computer security warnings: A mental model approach." *IEEE Security & Privacy* 9.2 (2010): 18-26.
14. Matteson, Miriam. "Capturing shared mental models: An approach for bona fide groups." *Journal of Librarianship and information Science* 47.1 (2015): 56-70.
15. Hodgkinson, Gerard P., A. John Maule, and Nicola J. Bown. "Causal cognitive mapping in the organizational strategy field: A comparison of alternative elicitation procedures." *Organizational Research Methods* 7.1 (2004): 3-26.
16. Biegel, Benjamin, et al. "Code Basket: Making Developers' Mental Model Visible and Explorable." 2015 IEEE/ACM 2nd International Workshop on Context for Software Development. IEEE, 2015.
17. Levesque, Laurie L., Jeanne M. Wilson, and Douglas R. Wholey. "Cognitive divergence and shared mental models in software development project teams." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 135-144.
18. Tversky, Barbara. "Cognitive maps, cognitive collages, and spatial mental models." *European conference on spatial information theory*. Springer, Berlin, Heidelberg, 1993.

19. Nacheva, Radka, et al. "Concept Map Mining Approach Based on the Mental Models Retrieval." *TEM Journal* 8.4 (2019): 1484.
20. Rieh, Soo Young, et al. "Conceptualizing institutional repositories: Using co-discovery to uncover mental models." *Proceedings of the third symposium on Information interaction in context*. 2010.
21. Green, Thomas R. G., Simon P. Davies, and David J. Gilmore. "Delivering cognitive psychology to HCI: the problems of common language and of knowledge transfer." *Interacting with Computers* 8.1 (1996): 89-111.
22. Loeffler, Diana, et al. "Developing intuitive user interfaces by integrating users' mental models into requirements engineering." *27th International BCS Human Computer Interaction Conference (HCI 2013)* 27. 2013.
23. Sim, Wee Wee, and Peggy Brouse. "Developing ontologies and persona to support and enhance requirements engineering activities—a case study." *Procedia computer science* 44.1 (2015): 275-284.
24. Georgiou, Andrea M., Glenn E. Littlepage, and Jennifer A. Henslee. "Development of criterion measures to assess interpositional knowledge and task mental models." *16th International Symposium on Aviation Psychology*. 2011.
25. Balijepally, VenuGopal, Sridhar Nerur, and RadhaKanta Mahapatra. "Effect of task mental models on software developer's performance: An experimental investigation." *2012 45th Hawaii International Conference on System Sciences*. IEEE, 2012.
26. Maass, Wolfgang, Veda C. Storey, and Tobias Kowatsch. "Effects of external conceptual models and verbal explanations on shared understanding in small groups." *International Conference on Conceptual Modeling*. Springer, Berlin, Heidelberg, 2011.
27. Jones, Natalie A., et al. "Eliciting mental models: a comparison of interview procedures in the context of natural resource management." *Ecology and Society* 19.1 (2014).
28. Webber, Sheila Simsarian, et al. "Enhancing team mental model measurement with performance appraisal practices." *Organizational Research Methods* 3.4 (2000): 307-322.
29. Zade, Himanshu. *Evolution of Mental Models of Interactive Machines: A Formal Approach*. Diss. International Institute of Information Technology Hyderabad, 2015.
30. Ross, Sarah, and Natalie Allen. "Examining the convergent validity of shared mental model measures." *Behavior research methods* 44.4 (2012): 1052-1062.
31. Van De Kieft, Iris, Catholijn M. Jonker, and M. Birna Van Riemsdijk. "Explaining negotiation: Obtaining a shared mental model of preferences." *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, Berlin, Heidelberg, 2011.
32. Langfield-Smith, Kim. "Exploring the need for a shared cognitive map." *Journal of management studies* 29.3 (1992): 349-368.
33. Carley, Kathleen M. "Extracting team mental models through textual analysis." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 18.S1 (1997): 533-558.
34. Carley, Kathleen, and Michael Palmquist. "Extracting, representing, and analyzing mental models." *Social forces* 70.3 (1992): 601-636.
35. Suzuki, Nobuo, Yoshikatsu Fujita, and Kazuhiko Tsuda. "Extraction method of the mutual understanding gap based on egocentrism in short dialogues." *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, Berlin, Heidelberg, 2011.
36. Hammarbäck, Jimmy. "Finding paths or getting lost?: Examining the mental model construct and mental model methodology." (2017).

37. Xie, Bingjun, Jia Zhou, and Huilin Wang. "How influential are mental models on interaction performance? exploring the gap between users' and designers' mental models through a new quantitative method." *Advances in Human-Computer Interaction 2017* (2017).
38. McKay, Judy, Peter Marshall, and Donald McDermid. "How shared is shared understanding in information requirements determination?." *Systems Development Methods for the Next Century*. Springer, Boston, MA, 1997. 155-168.
39. Rauterberg, M., S. Schluep, and M. Fjeld. "How to model behavioural and cognitive complexity in human-computer interaction with Petri nets." *Proceedings 6th IEEE International Workshop on Robot and Human Communication. RO-MAN'97 SENDAI*. IEEE, 1997.
40. Siau, Keng, and Xin Tan. "Improving the quality of conceptual modeling using cognitive mapping techniques." *Data & Knowledge Engineering* 55.3 (2005): 343-365.
41. Siau, Keng, and Xin Tan. "Information systems requirements determination and analysis: A mental modeling approach." *AMCIS 2003 Proceedings* (2003): 170.
42. Steiger, David M., and Natalie M. Steiger. "Instance-based cognitive mapping: a process for discovering a knowledge worker's tacit mental model." *Knowledge Management Research & Practice* 6.4 (2008): 312-321.
43. Humayun, Mamoon, and Cui Gang. "Investigating the role of organizational structure in developing shared understanding of requirements within GSD." *2012 15th International Multitopic Conference (INMIC)*. IEEE, 2012.
44. Reverdy, Justine, and Carl Vogel. "Linguistic repetitions, task-based experience and a proxy measure of mutual understanding." *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2017.
45. Mayr, Eva, et al. "Looking at the representations in our mind: Measuring mental models of information visualizations." *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 2016.
46. Doyle, James K., Michael J. Radzicki, and W. Scott Trees. "Measuring change in mental models of complex dynamic systems." *Complex decision making*. Springer, Berlin, Heidelberg, 2008. 269-294.
47. Wildman, Jessica L., Eduardo Salas, and Charles PR Scott. "Measuring cognition in teams: A cross-domain review." *Human factors* 56.5 (2014): 911-941.
48. Zin, Abdullah Mohd, and N. C. H. E. Pa. "Measuring communication gap in software requirements elicitation process." *Proceedings of the 8th WSEAS International Conference on Software engineering, parallel and distributed systems*. World Scientific and Engineering Academy and Society (WSEAS), 2009.
49. Gisick, Logan M., et al. "Measuring shared mental models in healthcare." *Journal of Patient Safety and Risk Management* 23.5 (2018): 207-219.
50. DeChurch, Leslie A., and Jessica R. Mesmer-Magnus. "Measuring shared team mental models: A meta-analysis." *Group dynamics: Theory, research, and practice* 14.1 (2010): 1.
51. Braunschweig, Brandt Benedict. *Measuring shared understanding in software design teams*. University of Maryland, Baltimore County, 2016.
52. Braunschweig, Brandt, and Carolyn Seaman. "Measuring shared understanding in software project teams using pathfinder networks." *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 2014.
53. Harbers, Maaïke, MB van Riemsdijk, and C. M. Jonker. "Measuring sharedness of mental models and its relation to team performance." *Proceedings 14th International Workshop on Coordination, Organisations, Institutions and Norms*. 2012.
54. Espinosa, J. Alberto, and Kathleen M. Carley. "Measuring team mental models." *Academy of Management Proceedings*. 2001.

55. Johnson, Tristan E., and Debra L. O'Connor. "Measuring team shared understanding using the analysis-constructed shared mental model methodology." *Performance Improvement Quarterly* 21.3 (2008): 113-134.
56. Schvaneveldt, Roger W., et al. "Measuring the structure of expertise." *International journal of man-machine studies* 23.6 (1985): 699-728.
57. Wu, Peng, and Jiang Song. "Measuring Website Users' Mental Model in nformation seeking with the PathFinder Network Approach." WHICEB. 2014.
58. Resick, Christian J., et al. "Mental model metrics and team adaptability: A multi-facet multi-method examination." *Group Dynamics: Theory, Research, and Practice* 14.4 (2010): 332.
59. Jones, Natalie A., et al. "Mental models: an interdisciplinary synthesis of theory and methods." *Ecology and Society* 16.1 (2011).
60. Diesner, Jana, Ponnurangam Kumaraguru, and Kathleen M. Carley. "Mental models of data privacy and security extracted from interviews with indians." 55th Annual Conference of the International Communication Association (ICA), New York, NY. 2005.
61. Coopamootoo, Kovila PL, and Thomas Groß. "Mental models of online privacy: Structural properties with cognitive maps." *Proceedings of the 28th International BCS Human Computer Interaction Conference (HCI 2014)* 28. 2014.
62. Langan-Fox, Janice, Jeromy Anglim, and John R. Wilson. "Mental models, team mental models, and performance: Process, development, and future directions." *Human Factors and Ergonomics in Manufacturing & Service Industries* 14.4 (2004): 331-352.
63. Mohammed, Susan, Lori Ferzandi, and Katherine Hamilton. "Metaphor no more: A 15-year review of the team mental model construct." *Journal of management* 36.4 (2010): 876-910.
64. Rushby, John. "Modeling the human in human factors." *International Conference on Computer Safety, Reliability, and Security*. Springer, Berlin, Heidelberg, 2001.
65. Rouse, William B., and Nancy M. Morris. "On looking into the black box: Prospects and limits in the search for mental models." *Psychological bulletin* 100.3 (1986): 349.
66. Jonassen, David H. "Operationalizing mental models: strategies for assessing mental models to support meaningful learning and design-supportive learning environments." (1995).
67. Ross, Susi, Magnus Ramage, and Yvonne Rogers. "PETRA: Participatory evaluation through redesign and analysis." *Interacting with Computers* 7.4 (1995): 335-360.
68. Porubän, Jaroslav. "Program comprehension with four-layered mental model." 2015 13th International Conference on Engineering of Modern Electric Systems (EMES). IEEE, 2015.
69. Biemann, Torsten, Thomas Ellwart, and Oliver Rack. "Quantifying similarity of team mental models: An introduction of the rRG index." *Group processes & intergroup relations* 17.1 (2014): 125-140.
70. Cannon-Bowers, Janis A., and Eduardo Salas. "Reflections on shared cognition." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 195-202.
71. Bjarnason, Elizabeth, Krzysztof Wnuk, and Björn Regnell. "Requirements are slipping through the gaps—A case study on causes & effects of communication gaps in large-scale software development." 2011 IEEE 19th international requirements engineering conference. IEEE, 2011.
72. Blanco, Eric, Franck Pourroy, and Serap Arikoglu. "Role of personas and scenarios in creating shared understanding of functional requirements: An empirical study." *Design Computing and Cognition'12*. Springer, Dordrecht, 2014. 61-78.

73. Johnson, Tristan E., et al. "Selection of team interventions based on mental model sharedness levels measured by the team assessment and diagnostic instrument (TADI)." *Computer-based diagnostics and systematic analysis of knowledge*. Springer, Boston, MA, 2010. 335-354.
74. Ensley, Michael D., and Craig L. Pearce. "Shared cognition in top management teams: Implications for new venture performance." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 145-160.
75. Salim, Siti Salwah. "Shared Mental Model Processing in Visualization Technologies: A Review of Fundamental Concepts and a Guide to Future Research in Human-Computer Interaction." *International Conference on Human-Computer Interaction*. Springer, Cham, 2020.
76. Espinosa, Alberto. "Shared mental models: Accuracy and visual representation." *AMCIS 2001 Proceedings* (2001): 404.
77. Short, Lawson. *Shared Understanding Within Large Information Systems Projects*. Diss. The Open University, 2016.
78. Barney, Sebastian, et al. "Software quality across borders: Three case studies on company internal alignment." *Information and Software Technology* 56.1 (2014): 20-38.
79. Yamaoka, Takayuki, et al. "Supporting mutual understanding in collaborative design project." *Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No. 98EX110)*. IEEE, 1998.
80. Oppl, Stefan. "Supporting the collaborative construction of a shared understanding about work with a guided conceptual modeling technique." *Group Decision and Negotiation* 26.2 (2017): 247-283.
81. Resick, Christian J., et al. "Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy." *Group Dynamics: Theory, Research, and Practice* 14.2 (2010): 174.
82. Van den Bossche, Piet, et al. "Team learning: building shared mental models." *Instructional Science* 39.3 (2011): 283-301.
83. Lim, Beng-Chong, and Katherine J. Klein. "Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 27.4 (2006): 403-418.
84. Mohammed, Susan, and Brad C. Dumville. "Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 89-106.
85. Langan-Fox, Janice, Sharon Code, and Kim Langfield-Smith. "Team mental models: Techniques, methods, and analytic approaches." *Human Factors* 42.2 (2000): 242-271.
86. Suzuki, Nobuo, and Kazuhiko Tsuda. "The Effective Extraction Method for the Gap of the Mutual Understanding Based on the Egocentrism in Business Communications." *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, Berlin, Heidelberg, 2009.
87. Hoffmann, Axel, Eva Alice Christiane Bittner, and Jan Marco Leimeister. "The emergence of mutual and shared understanding in the system development process." *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, Berlin, Heidelberg, 2013.
88. Ayoko, Oluremi B., and Eunice L. Chua. "The importance of transformational leadership behaviors in team mental model similarity, team efficacy, and intra-team conflict." *Group & Organization Management* 39.5 (2014): 504-531.

89. Mathieu, John E., et al. "The influence of shared mental models on team process and performance." *Journal of applied psychology* 85.2 (2000): 273.
90. Saikayasit, Rose, and Sarah Sharples. "The Influence of Shared-Representation on Shared Mental Models in Virtual Teams." *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, Berlin, Heidelberg, 2009.
91. Mohammed, Susan, Richard Klimoski, and Joan R. Rentsch. "The measurement of team mental models: We have no shared schema." *Organizational Research Methods* 3.2 (2000): 123-165.
92. Freeman, Lee A., and Leonard M. Jessup. "The power and benefits of concept mapping: measuring use, usefulness, ease of use, and satisfaction." *International Journal of Science Education* 26.2 (2004): 151-169.
93. Crowston, Kevin, and Barbara Scozzi. "The role of mental models in FLOSS development work practices." *IFIP International Conference on Open Source Systems*. Springer, Boston, MA, 2006.
94. Pascual, R. G. "Tools for capturing and training shared understanding in teams." (1999): 57-63.
95. Kearney, Anne R., and Stephen Kaplan. "Toward a methodology for the measurement of knowledge structures of ordinary people: the conceptual content cognitive map (3CM)." *Environment and behavior* 29.5 (1997): 579-617.
96. Sim, Wee Wee, and Peggy Brouse. "Towards an ontology-based persona-driven requirements and knowledge engineering." *Procedia Computer Science* 36 (2014): 314-321.
97. Lee, Mi Young. "Understanding changes in team-related and task-related mental models and their effects on team and individual performance." (2007).
98. Rushby, John. "Using model checking to help discover mode confusions and other automation surprises." *Reliability Engineering & System Safety* 75.2 (2002): 167-177.
99. Ohnishi, Kensuke, Tetsuya Yoshida, and Shogo Nishida. "Utilizing the correlation between decision trees to facilitate mutual understanding." *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*. Vol. 5. IEEE, 1999.
100. Daniels, Kevin, Leslie De Chernatony, and Gerry Johnson. "Validating a method for mapping managers' mental models of competitive industry structures." *Human Relations* 48.9 (1995): 975-991.
101. Mancuso, Vincent, et al. "What's on "their" mind: Evaluating collaborative systems using team mental models." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 55. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 2011.

E) Patterns für Egozentrismus nach Suzuki und Tsuda [42]

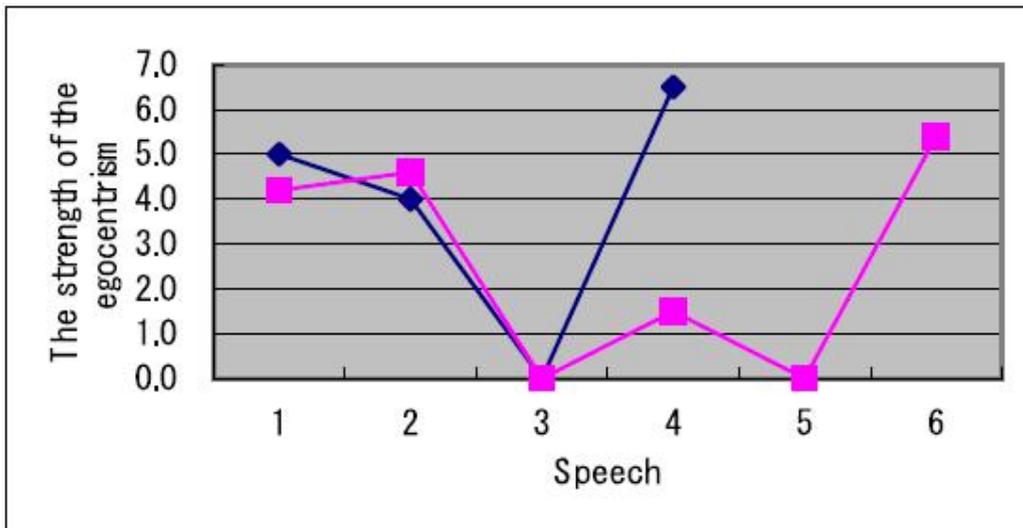


Abbildung 24: Beispielhafte Darstellung des zweiten Patterns von Suzuki und Tsuda [42]

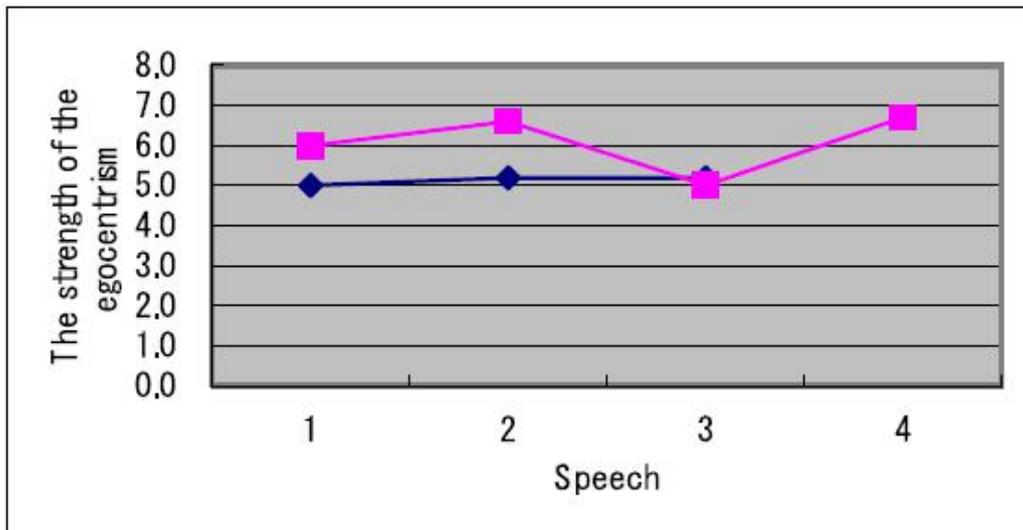


Abbildung 25: Beispielhafte Darstellung des dritten Patterns von Suzuki und Tsuda [42]

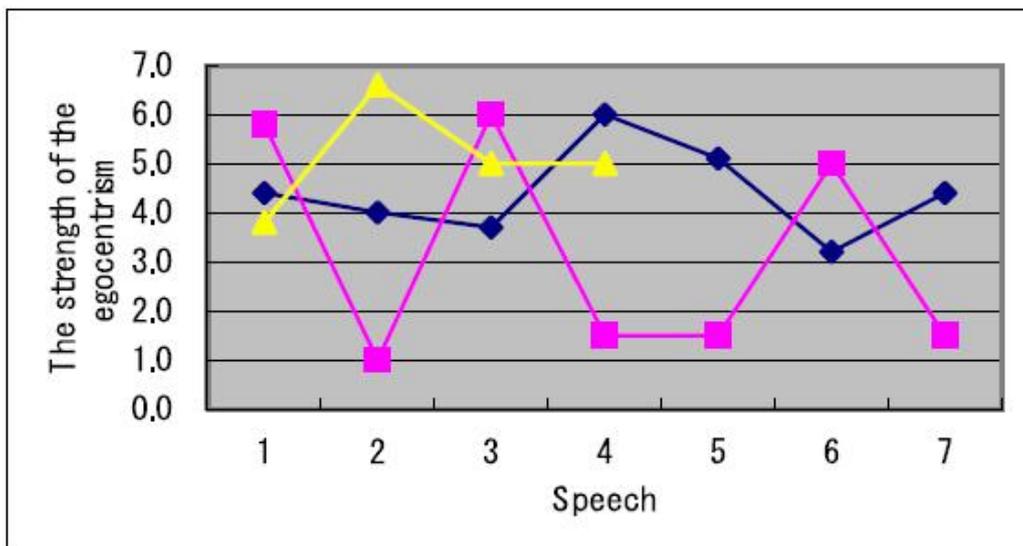


Abbildung 26: Beispielhafte Darstellung des vierten Patterns von Suzuki und Tsuda [42]

F) Tabellarische Übersicht der relevanten Publikationen

Die Referenzen der folgenden tabellarischen Übersicht beziehen sich auf die Liste der relevanten Publikationen in Anhang D. Eine „1“ bedeutet, dass das jeweilige Konzept in der Publikation thematisiert wurde.

Tabelle 10: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 1

Referenz	Jahr	Mental model	Shared Understanding	Interview	Case Study	Drawing/Image	Mapping	Pathfinder	Concept Maps
1.	2010	1							
2.	2015	1		1	1	1	1	1	1
3.	1995						1		
4.	2007	1			1				
5.	2015								
6.	2004								
7.	2007	1					1		
8.	2013	1				1	1		
9.	2020	1		1	1				
10.	2004	1					1		
11.	2001						1		
12.	2015	1		1					
13.	2010	1				1	1		
14.	2015	1		1	1	1			
15.	2004	1					1		
16.	2015	1				1	1		
17.	2001	1		1					
18.	1993	1					1		
19.	2019	1					1		1
20.	2010	1			1	1	1		
21.	1996	1					1		
22.	2013	1				1	1		
23.	2015			1		1	1		
24.	2011	1							
25.	2012	1							
26.	2011			1	1		1		
27.	2014	1		1		1			
28.	2000	1							
29.	2015	1							
30.	2012	1		1	1				1

Tabelle 11: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 2

Referenz	Fragebogen	Cognitive Distance	Common Understanding	Distance Ratio	Causal Map	Conceptual Model	Textuel	Cognitive Map	Persona
1.				1					
2.									
3.						1			
4.	1								
5.			1	1					
6.									
7.									
8.								1	
9.						1			
10.							1		
11.									
12.	1								
13.									
14.									
15.						1			1
16.									
17.	1		1						
18.									1
19.	1						1	1	
20.									
21.							1		
22.									
23.									1
24.	1								
25.	1								
26.							1		
27.									
28.									
29.									
30.									1

Tabelle 12: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 3

Referenz	pair rating	sorting	Repertory grid	Notizen
1.				Übersicht existierender Studien zu Messen & Vergleichen
2.				Multi-method Mental Model Elicitation, verbal vs graphisch & author vs computer
3.				Sehr auf mapping ausgelegt
4.				Interessante Visualisierungstechnik für Fragebogen
5.				
6.				Sehr lang und umfangreich
7.				Endlicher Automat genutzt
8.				Fokus auf Elicitation Techniken, direkt vs indirekt
9.				Healthcare! Unterscheidung verschiedener mental models
10.				Map analyse anhand eines Tools
11.				
12.				Selbsifter, FastFocus & TreasureHunt
13.				"Diagram of relationships"
14.				"degree of similarity" anhand von ähnlichen Beschreibungen
15.				zwei vorgeschlagene cognitive mapping techniken
16.				Tool für visuelles Mapping eines Mentalen models
17.				Gegenbeispiel: kein shared understanding erreicht; likert skala als measure
18.				Alternativen zum cognitive mapping: cognitive collage & spatial mental model
19.				Grundlagen von Concept Mapping erklärt
20.				Co-discovery study, gute quantitative Beschreibung
21.				Nutzt ERM um conceptual model darzustellen
22.				Image schemas: Mapping zwischen bedeutung & darstellung (oben=mehr etc)
23.				Definieren von ontologien mit persona technik um zusammenhänge klar zu definieren
24.				
25.				
26.				UML Diagramm, erklären von model im team; t-test
27.				
28.				Performance appraisal practice
29.				Endlicher Automat genutzt, sehr umfangreiche Arbeit
30.				Guter Überblick

Tabelle 13: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 1

Referenz	Jahr	Mental model	Shared Understanding	Interview	Case Study	Drawing/Image	Mapping	Pathfinder	Concept Maps
31.	2011		1						
32.	1992				1	1		1	
33.	1997		1					1	
34.	1992		1					1	
35.	2011			1	1	1			
36.	2017		1		1			1	1
37.	2017		1						
38.	1997			1				1	
39.	1997		1					1	
40.	2005		1					1	1
41.	2003		1					1	1
42.	2008		1					1	
43.	2012		1	1				1	1
44.	2017			1					
45.	2016		1		1		1	1	1
46.	2008		1						
47.	2014		1	1	1		1	1	1
48.	2009				1	1			
49.	2018		1		1			1	1
50.	2010		1			1		1	1
51.	2016		1	1	1	1		1	1
52.	2014		1	1	1	1			1
53.	2012		1	1					
54.	2001		1						
55.	2008		1	1					
56.	1985								1
57.	2014		1			1			1
58.	2010		1			1			
59.	2011		1	1	1		1	1	1
60.	2005		1		1			1	

Tabelle 14: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 2

Referenz	Fragebogen	Cognitive Distance	Common Understanding	Distance Ratio	Causal Map	Conceptual Model	Textuel	Cognitive Map	Persona
31.		1							
32.					1				1
33.							1		1
34.							1		1
35.									
36.									1
37.									
38.									1
39.									
40.					1		1		1
41.									1
42.									1
43.					1				
44.									
45.									
46.									
47.									
48.									
49.					1				
50.	1				1		1		1
51.	1								
52.	1								
53.									
54.									
55.									
56.									
57.									
58.									
59.							1		1
60.							1		

Tabelle 15: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 3

Referenz	pair rating	sorting	Repertory grid	Notizen
31.				SMM Definition enthalten; relativ mathematisch
32.				Relativ alt, Mental model nicht genannt
33.				Construction of Team cognitive map
34.				
35.				Egocentrism in Dialouges
36.				Highlight mental model methodological issues; definiert grundlagen & überblick diverse techniken
37.				Pfaddiagramm
38.				
39.				Petrinetze betrachtet
40.				Grundlagen zu Cognitive Mapping; causal, semantic & concept mapping
41.				Anderer Kontext, aber mit Erklärung zu versch. mapping
42.				Instance-based cognitive mapping; präsentiert diverse cognitive mapping techniken
43.	1	1		Erwähnt: card sorting
44.				Linguistische analyse von konversationen; stellt metrik anhand von signifikanz auf (proxy measure)
45.			1	1 Übersicht von versch. measures: inferential flow technique, non-reactive, think aloud, sorting
46.				Kontext dynamische systeme, vermutlich nicht nutzbar
47.			1	"All of ex. approaches for measuring team cognition", diverse techniken
48.				Measuring communication gap; Iterative Triangulation Method; Metrik anhand von Skala & Gap
49.	1	1		Überblick im kontext von healthcare
50.	1	1		Meta-analyse -> bietet Überblick über techniken
51.	1			Sehr umfangreiche Arbeit, auf eine metrik bezogen
52.				Aussage gegen mapping techniken, gegen surveys -> pathfinder
53.				Definiert eigene metrik
54.				Network analysis methods, sehr umfangreich
55.				Analysis constructed shared mental modek methodology; elicitation vs representation techniken
56.				Multi-dim scaling & link-weighted networks (pathfinder), relativ alt, mental model nicht erwähnt
57.				Kompletter fokus pathfinder
58.				Structural networks, priority ranking, importance ranking
59.				Überblick der elicitation techniken; direkt vs indirekt
60.				

Tabelle 16: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 1

Referenz	Jahr	Mental model	Shared Understanding	Interview	Case Study	Drawing/Image	Mapping	Pathfinder	Concept Maps
61.	2014	1			1		1		
62.	2004	1		1				1	
63.	2010	1		1				1	1
64.	2001	1							
65.	1986	1							
66.	1995	1					1	1	1
67.	1995			1					
68.	2015	1			1		1		
69.	2014	1		1					
70.	2001	1							
71.	2011			1	1	1			
72.	2014	1		1	1	1	1	1	
73.	2010	1							
74.	2001	1		1		1		1	
75.	2020	1			1	1		1	1
76.	2001	1							
77.	2016	1		1					
78.	2014			1		1			
79.	1998						1		
80.	2017	1		1				1	
81.	2010	1		1		1			
82.	2011	1						1	
83.	2006	1		1					1
84.	2001	1		1				1	1
85.	2000	1		1	1	1	1	1	1
86.	2009								
87.	2013			1					
88.	2014	1		1					
89.	2000	1							
90.	2009	1							

Tabelle 17: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 2

Referenz	Fragebogen	Cognitive Distance	Common Understanding	Distance Ratio	Causal Map	Conceptual Model	Textuel	Cognitive Map	Persona
61.	1							1	1
62.									
63.							1		
64.									
65.									
66.							1		1
67.								1	
68.									
69.	1								
70.									
71.	1		1						
72.									1
73.									
74.	1								1
75.	1						1	1	1
76.									
77.									
78.	1								
79.									
80.							1		
81.	1								
82.	1								1
83.									
84.									
85.					1	1			1
86.			1						
87.	1								
88.	1								
89.									
90.	1								

Tabelle 18: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 3

Referenz	pair rating	sorting	Repertory grid	Notizen
61.				Sprachliche Analyse von text
62.				Kritik an techniken;
63.	1	1		Große übersicht, tabellen sind sehr hilfreich
64.				Zustandsautomat, model checking
65.				Sehr grundlegende Arbeit zu mental models
66.				Representation von mental model als nodes & relationships
67.				characterising the interactions & conversations between participants; tool
68.				Bilden von mental model,beispiel code verstehen
69.				r RG index
70.				2 ways: assessing structure vs measure content
71.				fokus auf Communication gaps
72.				Persona studie; speamans rho
73.				tool (TADi)
74.				coefficient of variation of the STROBE scale; Buisness Kontext
75.	1	1		Review of fundamental concepts(übersicht); 2 ways: structure vs content; UCINET
76.				Network analysis method, hauptsächlich accuracy
77.				
78.				SAAM-SQ to quantify levels of alignment
79.				
80.				Examples for mismatch-> modelle greifen nicht ineinander
81.				
82.				
83.				taskwork und teamwork mental model
84.				interdisziplinär;
85.	1	1	1	Überblick versch. methoden, S.4 tabelle
86.				Egocentrism in Dialouges, mutual understanding
87.				definition shared understanding; Kategorisierung von fragen -> exploration:shared unders.
88.				
89.	1			task- & teambased mental models; typen von SMM erklärt
90.				shared representation facilities; paired-sample t-test

Tabelle 19: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 1

Referenz	Jahr	Mental model	Shared Understanding	Interview	Case Study	Drawing/Image	Mapping	Pathfinder	Concept Maps
91.	2000		1	1	1			1	1
92.	2004		1	1				1	1
93.	2006		1	1	1				
94.	1999		1	1	1				
95.	1997		1				1		1
96.	2014			1					
97.	2007		1	1				1	
98.	2002		1						
99.	1999								
100.	1993		1				1		
101.	2011		1			1	1		1

Tabelle 20: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 2

Referenz	Fragebogen	Cognitive Distance	Common Understanding	Distance Ratio	Causal Map	Conceptual Model	Textuel	Cognitive Map	Persona
91.		1						1	1
92.									
93.				1					1
94.	1								
95.									1
96.									1
97.		1							
98.									
99.			1						
100.									1
101.		1							

Tabelle 21: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 3

Referenz	pair rating	sorting	Repertory grid	Notizen
91.				Kritik & vergleich von techniken
92.				Fokus concept mapping
93.				
94.	1			Experimental&dynamic vs teamwork&taskwork mental model; tools
95.				3CM Method;
96.				Definieren von ontologien mit persona technik um zusammenhänge klar zu definieren
97.				sehr umfangreich; team&task mental models; großer theorieteil, vernünftiges inhaltsverzeichnis
98.				Model checking paper; endlicher automat; finde automatisch unterschiede
99.				Decision Tree, mutual understanding -> gleiches konzept versch. bezeichnung
100.		1	1	
101.	1			

G) Beispiel einer Konzepttabelle für den Pathfinder Algorithmus

Tabelle 22: Beispiel einer Konzepttabelle für einen Fahrkartenautomaten

	Fahrkarte	Bezahlung	Zonenanzahl	Personenanzahl
Fahrkarte				
Bezahlung				
Zonenanzahl				
Personenanzahl				

Literaturverzeichnis

- [1] Keele, Staffs. *Guidelines for performing systematic literature reviews in software engineering*. Vol. 5. Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.
- [2] Brings, Jennifer, et al. "On Different Search Methods for Systematic Literature Reviews and Maps: Experiences from a Literature Search on Validation and Verification of Emergent Behavior." *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. 2018.
- [3] Feldt, Robert. „ISI SE Journals (Ranked)“ (2016), Letzter Zugriff: 29.10.2020, http://www.robertfeldt.net/advice/se_venues/
- [4] Google Scholar. „Top-Publikationen; Engineering & Computer Science, Software Systems“, Letzter Zugriff: 29.10.2020, https://scholar.google.de/citations?view_op=top_venues&hl=de&vq=eng_softwaresystems
- [5] Zhang, He, and Muhammad Ali Babar. "On searching relevant studies in software engineering." *14th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. 2010.
- [6] Jones, Natalie A., et al. "Mental models: an interdisciplinary synthesis of theory and methods." *Ecology and Society* 16.1 (2011).
- [7] Hammarbäck, Jimmy. "Finding paths or getting lost?: Examining the mental model construct and mental model methodology." (2017).
- [8] Mohammed, Susan, Richard Klimoski, and Joan R. Rentsch. "The measurement of team mental models: We have no shared schema." *Organizational Research Methods* 3.2 (2000): 123-165.
- [9] Siau, Keng, and Xin Tan. "Improving the quality of conceptual modeling using cognitive mapping techniques." *Data & Knowledge Engineering* 55.3 (2005): 343-365.
- [10] Steiger, David M., and Natalie M. Steiger. "Instance-based cognitive mapping: a process for discovering a knowledge worker's tacit mental model." *Knowledge Management Research & Practice* 6.4 (2008): 312-321.
- [11] Siau, Keng, and Xin Tan. "Information systems requirements determination and analysis: A mental modeling approach." *AMCIS 2003 Proceedings* (2003): 170.
- [12] Carley, Kathleen, and Michael Palmquist. "Extracting, representing, and analyzing mental models." *Social forces* 70.3 (1992): 601-636.
- [13] Hodgkinson, Gerard P., A. John Maule, and Nicola J. Bown. "Causal cognitive mapping in the organizational strategy field: A comparison of alternative elicitation procedures." *Organizational Research Methods* 7.1 (2004): 3-26.
- [14] Langan-Fox, Janice, Sharon Code, and Kim Langfield-Smith. "Team mental models: Techniques, methods, and analytic approaches." *Human Factors* 42.2 (2000): 242-271.
- [15] Markíczy, Livia, and Jeff Goldberg. "A method for eliciting and comparing causal maps." *Journal of management* 21.2 (1995): 305-333.
- [16] Ross, Sarah, and Natalie Allen. "Examining the convergent validity of shared mental model measures." *Behavior research methods* 44.4 (2012): 1052-1062.
- [17] Humayun, Mamoona, and Cui Gang. "Investigating the role of organizational structure in developing shared understanding of requirements within GSD." *2012 15th International Multitopic Conference (INMIC)*. IEEE, 2012.
- [18] Novak, Joseph D. "How Do We Learn Our Lesson?." *Science Teacher* 60.3 (1993): 50-55.
- [19] Mayr, Eva, et al. "Looking at the representations in our mind: Measuring mental models of information visualizations." *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 2016.

- [20] Grenier, Robin S., and Dana Dudzinska-Przesmitzki. "A conceptual model for eliciting mental models using a composite methodology." *Human Resource Development Review* 14.2 (2015): 163-184.
- [21] Wildman, Jessica L., Eduardo Salas, and Charles PR Scott. "Measuring cognition in teams: A cross-domain review." *Human factors* 56.5 (2014): 911-941.
- [22] Gisick, Logan M., et al. "Measuring shared mental models in healthcare." *Journal of Patient Safety and Risk Management* 23.5 (2018): 207-219.
- [23] Mohammed, Susan, Lori Ferzandi, and Katherine Hamilton. "Metaphor no more: A 15-year review of the team mental model construct." *Journal of management* 36.4 (2010): 876-910.
- [24] Freeman, Lee A., and Leonard M. Jessup. "The power and benefits of concept mapping: measuring use, usefulness, ease of use, and satisfaction." *International Journal of Science Education* 26.2 (2004): 151-169.
- [25] Rauterberg, Matthias. "AMME: an Automatic Mental Model Evaluation to analyse user behaviour traced in a finite, discrete state space." *Ergonomics* 36.11 (1993): 1369-1380.
- [26] Rushby, John. "Using model checking to help discover mode confusions and other automation surprises." *Reliability Engineering & System Safety* 75.2 (2002): 167-177.
- [27] Zade, Himanshu. *Evolution of Mental Models of Interactive Machines: A Formal Approach*. Diss. International Institute of Information Technology Hyderabad, 2015.
- [28] Rauterberg, M., S. Schlupe, and M. Fjeld. "How to model behavioural and cognitive complexity in human-computer interaction with Petri nets." *Proceedings 6th IEEE International Workshop on Robot and Human Communication. RO-MAN'97 SENDAI*. IEEE, 1997.
- [29] Sim, Wee Wee, and Peggy Brouse. "Towards an ontology-based persona-driven requirements and knowledge engineering." *Procedia Computer Science* 36 (2014): 314-321.
- [30] Sim, Wee Wee, and Peggy Brouse. "Developing ontologies and persona to support and enhance requirements engineering activities—a case study." *Procedia computer science* 44.1 (2015): 275-284.
- [31] Ohnishi, Kensuke, Tetsuya Yoshida, and Shogo Nishida. "Utilizing the correlation between decision trees to facilitate mutual understanding." *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*. Vol. 5. IEEE, 1999.
- [32] Green, Thomas R. G., Simon P. Davies, and David J. Gilmore. "Delivering cognitive psychology to HCI: the problems of common language and of knowledge transfer." *Interacting with Computers* 8.1 (1996): 89-111.
- [33] Lee, Mi Young. "Understanding changes in team-related and task-related mental models and their effects on team and individual performance." (2007).
- [34] Ensley, Michael D., and Craig L. Pearce. "Shared cognition in top management teams: Implications for new venture performance." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 145-160.
- [35] Braunschweig, Brandt, and Carolyn Seaman. "Measuring shared understanding in software project teams using pathfinder networks." *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 2014.
- [36] Paullin, Cheryl, et al. *Development of criterion measures to assess retention and decay of aerospace physiology knowledge and skills*. No. NPSI-255. PERSONNEL DECISIONS RESEARCH INST MINNEAPOLIS MN, 2002.

- [37] Levesque, Laurie L., Jeanne M. Wilson, and Douglas R. Wholey. "Cognitive divergence and shared mental models in software development project teams." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22.2 (2001): 135-144.
- [38] Seeber, Isabella, et al. "Brainstorming is just the beginning: Effects of convergence techniques on satisfaction, perceived usefulness of moderation, and shared understanding in teams." *2015 48th Hawaii International Conference on System Sciences*. IEEE, 2015.
- [39] Resick, Christian J., et al. "Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy." *Group Dynamics: Theory, Research, and Practice* 14.2 (2010): 174.
- [40] Biemann, Torsten, Thomas Ellwart, and Oliver Rack. "Quantifying similarity of team mental models: An introduction of the rRG index." *Group processes & intergroup relations* 17.1 (2014): 125-140.
- [41] Reverdy, Justine, and Carl Vogel. "Linguistic repetitions, task-based experience and a proxy measure of mutual understanding." *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2017.
- [42] Suzuki, Nobuo, and Kazuhiko Tsuda. "The Effective Extraction Method for the Gap of the Mutual Understanding Based on the Egocentrism in Business Communications." *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, Berlin, Heidelberg, 2009.
- [43] Schvaneveldt, Roger W., et al. "Measuring the structure of expertise." *International journal of man-machine studies* 23.6 (1985): 699-728.
- [44] Wu, Peng, and Jiang Song. "Measuring Website Users' Mental Model in information seeking with the PathFinder Network Approach." *WHICEB*. 2014.
- [45] Braunschweig, Brandt Benedict. *Measuring shared understanding in software design teams*. University of Maryland, Baltimore County, 2016.
- [46] Lim, Beng-Chong, and Katherine J. Klein. "Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 27.4 (2006): 403-418.
- [47] Zin, Abdullah Mohd, and N. C. H. E. Pa. "Measuring communication gap in software requirements elicitation process." *Proceedings of the 8th WSEAS International Conference on Software engineering, parallel and distributed systems*. World Scientific and Engineering Academy and Society (WSEAS), 2009.
- [48] Daniels, Kevin, Leslie De Chernatony, and Gerry Johnson. "Validating a method for mapping managers' mental models of competitive industry structures." *Human Relations* 48.9 (1995): 975-991.
- [49] Carley, Kathleen M. "Extracting team mental models through textual analysis." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 18.S1 (1997): 533-558.
- [50] Zhou, Xin, et al. "A map of threats to validity of systematic literature reviews in software engineering." *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, 2016.
- [51] Wohlin, Claes, et al. *Experimentation in software engineering*. Springer Science & Business Media, 2012.
- [52] Basili, Victor R., and H. Dieter Rombach. "The TAME project: Towards improvement-oriented software environments." *IEEE Transactions on software engineering* 14.6 (1988): 758-773.

- [53] Easterbrook, Steve. "Coordination breakdowns: why groupware is so difficult to design." (1995): 8-8.
- [54] Aranda, Jorge. *A theory of shared understanding for software organizations*. University of Toronto, 2010.
- [55] Glinz, Martin, and Samuel Fricker. *On shared understanding in software engineering*. Gesellschaft für Informatik eV, 2013.
- [56] Pham, Raphael, et al. "Interactive multimedia storyboard for facilitating stakeholder interaction: supporting continuous improvement in IT-ecosystems." *2012 Eighth International Conference on the Quality of Information and Communications Technology*. IEEE, 2012.
- [57] Fricker, Samuel A., Rainer Grau, and Adrian Zwingli. "Requirements engineering: best practice." *Requirements Engineering for Digital Health*. Springer, Cham, 2015. 25-46.
- [58] Karras, Oliver, Kurt Schneider, and Samuel A. Fricker. "Representing software project vision by means of video: A quality model for vision videos." *Journal of Systems and Software* 162 (2020): 110479.
- [59] Karras, O. "Communicating stakeholders' needs-vision videos to disclose, discuss, and align mental models for shared understanding." *IEEE Softw. Blog* (2019).
- [60] Salim, Siti Salwah. "Shared Mental Model Processing in Visualization Technologies: A Review of Fundamental Concepts and a Guide to Future Research in Human-Computer Interaction." *International Conference on Human-Computer Interaction*. Springer, Cham, 2020.
- [61] Schneider, Kurt, et al. "Refining vision videos." *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, Cham, 2019.
- [62] Karras, Oliver, Jil Klünder, and Kurt Schneider. "Enrichment of requirements specifications with videos-enhancing the comprehensibility of textual requirements." (2016).
- [63] McKay, Judy, Peter Marshall, and Donald McDermid. "How shared is shared understanding in information requirements determination?." *Systems Development Methods for the Next Century*. Springer, Boston, MA, 1997. 155-168.
- [64] Craik, Kenneth James Williams. *The nature of explanation*. Vol. 445. CUP Archive, 1952.
- [65] Johnson-Laird, Philip N. "Mental models in cognitive science." *Cognitive science* 4.1 (1980): 71-115.
- [66] Johnson-Laird, Philip Nicholas. *Mental models: Towards a cognitive science of language, inference, and consciousness*. No. 6. Harvard University Press, 1983.
- [67] Norman, Donald A. "Some observations on mental models." *Mental models* 7.112 (1983): 7-14.
- [68] Carroll, John M., and Judith Reitman Olson. "Mental models in human-computer interaction." *Handbook of human-computer interaction*. North-Holland, 1988. 45-65.
- [69] Chermack, Thomas J. "Mental models in decision making and implications for human resource development." *Advances in developing human resources* 5.4 (2003): 408-422.
- [70] Rook, Laura. "Mental models: A robust definition." *The Learning Organization* (2013).
- [71] Gulliksen, Jan, and Ann Lantz. "Design versus design-from the shaping of products to the creation of user experiences." *International Journal of Human-Computer Interaction* 15.1 (2003): 5-20.
- [72] Pohl, Klaus, and Chris Rupp. *Basiswissen Requirements Engineering: Aus-und Weiterbildung nach IREB-Standard zum Certified Professional for Requirements Engineering Foundation Level*. dpunkt. verlag, 2015.
- [73] Rupp, Christine. *Requirements-Engineering und-Management: Aus der Praxis von klassisch bis agil*. Carl Hanser Verlag GmbH Co KG, 2014.

- [74] Fischer, Gerhard. "Symmetry of ignorance, social creativity, and meta-design." *Knowledge-Based Systems* 13.7-8 (2000): 527-537.
- [75] Broll, Gregor, et al. "Using Video Clips to Support Requirements Elicitation in Focus Groups-An Experience Report." *SE 2007 Workshop on Multimedia Requirements Engineering*. 2007.
- [76] Creighton, Oliver, Martin Ott, and Bernd Bruegge. "Software cinema-video-based requirements engineering." *14th IEEE International Requirements Engineering Conference (RE'06)*. IEEE, 2006.
- [77] Brill, Olesia, Kurt Schneider, and Eric Knauss. "Videos vs. use cases: Can videos capture more requirements under time pressure?." *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, Berlin, Heidelberg, 2010.
- [78] Papaioannou, Diana, et al. "Literature searching for social science systematic reviews: consideration of a range of search techniques." *Health Information & Libraries Journal* 27.2 (2010): 114-122.
- [79] Greenhalgh, Trisha, and Richard Peacock. "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources." *Bmj* 331.7524 (2005): 1064-1065.
- [80] Zhang, He, Muhammad Ali Babar, and Paolo Tell. "Identifying relevant studies in software engineering." *Information and Software Technology* 53.6 (2011): 625-637.
- [81] Halevi, Gali, Henk Moed, and Judit Bar-Ilan. "Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature." *Journal of informetrics* 11.3 (2017): 823-834.
- [82] Mathieu, John E., et al. "The influence of shared mental models on team process and performance." *Journal of applied psychology* 85.2 (2000): 273.
- [83] Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. *Einführung in die Automatentheorie, formale Sprachen und Berechenbarkeit*. Pearson Deutschland GmbH, 2011.
- [84] Bjarnason, Elizabeth, et al. "A theory of distances in software engineering." *Information and Software Technology* 70 (2016): 204-219.
- [85] Trkman, Marina, et al. "Impact of the conceptual model's representation format on identifying and understanding user stories." *Information and software technology* 116 (2019): 106169.
- [86] de la Vara, Jose Luis, et al. "An Empirical Evaluation of the Use of Models to Improve the Understanding of Safety Compliance Needs." *Information and Software Technology* (2020): 106351.
- [87] Billestrup, Jane, et al. "Persona usage in software development: advantages and obstacles." *Proc. of ACHI* (2014): 359-364.
- [88] Sutcliffe, Allistair G. "Convergence or competition between software engineering and human computer interaction." *Human-Centered Software Engineering—Integrating Usability in the Software Development Lifecycle*. Springer, Dordrecht, 2005. 71-84.
- [89] Ebling, Thaís, Jorge Luis Nicolas Audy, and Rafael Prikladnicki. "A Systematic Literature Review of Requirements Engineering in Distributed Software Development Environments." *ICEIS* (3). 2009.

Tabellenverzeichnis

Tabelle 1: Verteilung der Anzahl der Publikationen in den Schritten der manuellen Suche....	18
Tabelle 2: Rangliste der Häufigkeit der Nutzung von Datenbanken in systematischen Literatursuchen nach Zhang und Babar[5].....	22
Tabelle 3: Verteilung der Anzahl der Ergebnisse in den Schritten der Datenbanksuche.....	24
Tabelle 4: Gewichte für die Bestimmung von Egozentrismus nach Suzuki und Tsuda [42]...	49
Tabelle 5: Hypothese zum Ziel 1: Erzeugung von „shared understanding“	60
Tabelle 6: Erste Hypothese zum Ziel 2: Mehrwert bezüglich des „shared understandings“....	60
Tabelle 7: Zweite Hypothese zum Ziel 2: Mehrwert bezüglich des „shared understandings“.	61
Tabelle 8: Beispielhafte Aufteilung der Treatments.....	66
Tabelle 9: Kombination der Suchbegriffe für die Titelsuche mit SpringerLink.....	76
Tabelle 10: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 1.....	86
Tabelle 11: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 2.....	86
Tabelle 12: Übersicht der relevanten Publikationen, Referenz 1-30 Teil 3.....	87
Tabelle 13: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 1.....	87
Tabelle 14: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 2.....	88
Tabelle 15: Übersicht der relevanten Publikationen, Referenz 31-60 Teil 3.....	88
Tabelle 16: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 1.....	89
Tabelle 17: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 2.....	89
Tabelle 18: Übersicht der relevanten Publikationen, Referenz 61-90 Teil 3.....	90
Tabelle 19: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 1.....	90
Tabelle 20: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 2.....	90
Tabelle 21: Übersicht der relevanten Publikationen, Referenz 91-101 Teil 3.....	90
Tabelle 22: Beispiel einer Konzepttabelle für einen Fahrkartenautomaten.....	91

Abbildungsverzeichnis

Abbildung 1: Videonutzung im Software-Lebenszyklus nach Creighton et al. [76].....	8
Abbildung 2: Effektivität der verschiedenen Suchmethoden nach Brings et al. [2].....	14
Abbildung 3: Effizienz der verschiedenen Suchmethoden nach Brings et al. [2].....	15
Abbildung 4: Anzahl relevanter Publikationen im Zeitraum von 2010 bis 2020.....	19
Abbildung 5: Gruppierung der Schlüsselwörter anhand von farblicher Codierung.....	20
Abbildung 6: Aus der manuellen Suche resultierender Suchstring.....	20
Abbildung 7: Angepasster Suchstring anhand der Einschränkungen der Datenbanksuche.....	23
Abbildung 8: Effizienz der Datenbanksuchen in Prozent.....	25
Abbildung 9: Ablauf und Ergebnisse der einzelnen Schritte des Snowballings.....	27
Abbildung 10: Effizienz der einzelnen Schritte des Snowballings in Prozent.....	28
Abbildung 11: Überblick über den Ablauf der gesamten systematischen Literatursuche.....	30
Abbildung 12: Vergleich der Effektivität der Suchmethoden mit Brings et al. [2].....	32
Abbildung 13: Vergleich der Effizienz der Suchmethoden mit Brings et al. [2].....	32
Abbildung 14: Verteilung der relevanten Publikationen anhand der Publikationsjahre.....	34
Abbildung 15: Vergleich der prozentualen Anteile der relevanten Publikationen mit den Themen „shared understanding“ beziehungsweise mentales Modell.....	35
Abbildung 16: Beispiel einer „causal map“ von Siau und Tan [9].....	37
Abbildung 17: Beispielhafte Adjazenzmatrizen für einen Ausschnitt aus Abbildung 16.....	37
Abbildung 18: Beispiel einer „concept map“ von Siau und Tan [9].....	39
Abbildung 19: Beispiel einer Reduktion eines vollständig verbundenen Graphen zu einem Pathfinder Netzwerk von Braunschweig [45].....	41
Abbildung 20: Vergleich zweier Pathfinder Netzwerke von Lim und Klein [46].....	41
Abbildung 21: Vergleichs zweier Entscheidungsbäume von Ohnishi et al. [31].....	46
Abbildung 22: Beispielhafte Darstellung des ersten Patterns von Suzuki und Tsuda [42].....	50
Abbildung 23: Ablauf der Schritte einer Experimentplanung nach Wohlin et al. [51].....	56
Abbildung 24: Beispielhafte Darstellung des zweiten Patterns von Suzuki und Tsuda [42]....	85
Abbildung 25: Beispielhafte Darstellung des dritten Patterns von Suzuki und Tsuda [42].....	85
Abbildung 26: Beispielhafte Darstellung des vierten Patterns von Suzuki und Tsuda [42].....	85

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbstständig ohne fremde Hilfe verfasst und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keinem anderen Prüfungsamt vorgelegen.

Hannover, den 03.11.2020

Robert Völkner