

**Gottfried Wilhelm
Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik
Institut für Verteilte Systeme
Fachgebiet Data Science und Digital Libraries**

Analyse der Anwendung von Videos im Requirements Engineering mittels semantischer Struktur

**Analysis of Video Application in Requirements Engineering
Using Semantic Structures**

Bachelorarbeit

im Studiengang Informatik

von

Alex Hahn

**Prüfer: Prof. Dr. Sören Auer
Zweitprüfer: Dr. Oliver Karras
Betreuer: Dr. Oliver Karras**

Hannover, 14.09.2022

Erklärung der Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und 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 14.09.2022



Alex Hahn

Zusammenfassung

Die Anzahl der wissenschaftlichen Publikationen nimmt jedes Jahr zu, was die Suche nach bestimmten Daten und relevanten Information erschwert. Eine geringe Nachhaltigkeit der Datenbasis bei den aktuell verwendeten Analysemethoden stellt ein Problem für die Forscher dar, wenn ein aktueller Stand der Forschung in einem Forschungsfeld untersucht werden muss.

Systematische Literaturrecherche und systematische Mapping-Studie sind aktuell die meist benutzten Methoden, die das Problem lösen sollten. Jedoch bietet keiner von den beiden Methoden nachhaltige und langfristige Aufbereitung der Daten. Außerdem bieten die systematische Literaturrecherche und systematische Mapping-Studie keine Wiederverwendbarkeit. Um einen aktuellen Stand der Forschung zu erhalten nach dem neue Artikel publiziert wurden, müssen diese aufwändigen Methoden jedes Mal neu durchgeführt werden.

Als Alternative zu den bereits bekannten Methoden wird in dieser Arbeit Open Research Knowledge Graph (ORKG) untersucht. ORKG bietet sogenannte Templates, diese ermöglichen das Erzeugen einer semantischen Struktur der jeweiligen Publikation. Damit wiederum lassen sich alle Daten maschinenlesbar machen. Im Rahmen dieser Arbeit wird ein bereits bestehendes Template "Video Process" aufgebaut und an fünfzig wissenschaftlichen Publikationen zum Thema "Production and Use of Videos Requirements Engineering" angewendet. Außerdem werden die Daten der entsprechenden Publikationen mit Hilfe von SPARQL, Schnittstelle von ORKG, abgerufen und in einem Jupyter Notebook graphisch dargestellt. Damit werden die Zusammenhänge von bestimmten Eigenschaften des Video Prozesses analysiert.

Die durchgeführte Analyse hat gezeigt, dass es anhand der zur Verfügung gestellten Publikationen zwei Forschungslücken gibt. Das Ergebnis dieser Arbeit bezeichnet ORKG als eine mögliche Lösung um wissenschaftliches Wissen nachhaltig und langfristig zu akquirieren, kuratieren, veröffentlichen und zu verarbeiten.

Abstract

Analysis of Video Application in Requirements Engineering Using Semantic Structures.

The number of scientific publications is increasing every year, making it difficult to find specific data and relevant information. A low sustainability of the database with the currently used analysis methods poses a problem for the researchers when the current state of research in a research field has to be examined.

Systematic literature review and systematic mapping study are currently the most used methods that should solve the problem, but neither of the two methods offers more sustainable and long-term processing of the data. In addition, the systematic literature review and systematic mapping study do not offer any reusability. These time-consuming methods have to be repeated every time new articles are published in order to keep the research up-to-date.

As an alternative to the already known methods, Open Research Knowledge Graph (ORKG) is examined in this work. ORKG offers so-called templates that make it possible to create a semantic structure for the respective publication and also make the data machine-readable. As part of this work, an existing template “Video Process” is set up and applied to 50 scientific publications on the subject of “Production and Use of Videos Requirements Engineering”. In addition, the data from the corresponding publications are retrieved using the SPARQL interface of ORKG and displayed graphically in a Jupyter notebook in order to analyze the relationships between certain properties of the video process.

The made analysis, based on available publications, has shown that there are two research gaps. The results of this work also describe the ORKG as a possible solution to acquire, curate, publish and process scientific knowledge in a sustainable and long-term manner.

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation	1
1.2 Ziel der Arbeit	3
1.3 Struktur der Arbeit	5
2 Grundlagen	7
2.1 Etablierte Methoden zur Analyse von Literatur	7
2.1.1 Systematische Literaturrecherche	7
2.1.2 Systematische Mapping-Studie	10
2.2 ORKG als neue Methode zur Literaturanalyse	12
2.2.1 Was ist der ORKG?	12
2.2.2 Semantische Struktur	12
2.2.3 Vorteile vom ORKG gegenüber etablierten Methoden	13
2.3 Requirements Engineering	14
3 Ausbau und Anwendung des Templates	17
3.1 Ziel der semantischen Analyse	17
3.2 Vorteile und Nachteile des Ansatzes	17
3.3 Initial Template	18
3.4 Ausbau des Templates	22
3.5 Anwendung des Templates	23
4 Datenanalyse	25
4.1 Anforderungen an die Analyse	25
4.2 Funktion des Jupyter Notebooks	26
4.3 Ergebnis der Datenanalyse	27
4.4 Fazit der Datenanalyse	34
5 Auswertung	37
5.1 Threats to Validity	37
5.2 Auswertung des Templates	39
5.2.1 Zeitaufwand	39
5.2.2 Nachhaltigkeit	39
5.3 Auswertung der Datenanalyse	40

5.3.1 Zeitaufwand	40
5.3.2 Nachhaltigkeit	40
6 Verwandte Arbeiten	43
7 Fazit und Ausblick	45
7.1 Fazit	45
7.2 Ausblick	46
A Anhang	47

Abbildungsverzeichnis

2.1 Systematischer Literaturrecherche Prozess	8
2.2 Systematischer Mapping Prozess nach Petersen et al. [19]	10
2.3 Referenzmodell des Requirements Engineering [4]	15
3.1 Production Property des Video Process Templates	19
3.2 VideoObject Property	20
3.3 Application Property des Video Process Templates	21
3.4 Problem Template	22
3.5 Study Template	22
3.6 Costs Template	23
3.7 Fidelity Template	23
4.1 Abhängigkeit der Guidelines von der Zeit	28
4.2 Eigenschaften des Kontextes von Videos	29
4.3 Kombination der Eigenschaften des Kontextes von Videos	30
4.4 Phasen der Requirements Engineering	31
4.5 Synonyme von Video	32
4.6 Content und Dauer von Videos	32
4.7 Interaktionsmöglichkeiten in Videos	33
4.8 Videos die ein/e Problem/Solution darstellen	33
4.9 Hersteller von Videos	34
4.10 High-Cost Videos	35
4.11 High-Fidelity and Low-Fidelity Videos	35

Kapitel 1

Einleitung

1.1 Motivation

Die meisten Forschungen beginnen mit einer Art Literaturrecherche, damit die Forschungsergebnisse, die für eine bestimmte Forschungsfrage, ein bestimmtes Themengebiet oder ein interessierendes Phänomen relevant sind, identifiziert, bewertet und interpretiert werden können [1]. Mit dem Wachstum der wissenschaftlichen Gemeinschaft nimmt die Zahl der wissenschaftlichen Veröffentlichungen jedes Jahr zu. Allein im Bereich Science and Engineering beträgt das Wachstum der Publikationszahlen vier Prozent pro Jahr [16]. Hierbei besteht das Problem, die relevante Literatur in der Vielzahl von wissenschaftlichen Artikeln zu finden, deren Ergebnisse zu vergleichen und die wichtige Information aus den Artikeln zu extrahieren.

Ein grundlegendes Problem bei der Untersuchung des aktuellen Standes der Forschung in einem Forschungsfeld, ist eine geringe Nachhaltigkeit der Datenbasis bei den verwendeten Analysemethoden. Die Verbreitung wissenschaftlicher Erkenntnisse hat sich seit vielen hundert Jahren nicht sonderbar verändert. Diese erfolgt in der Regel dokumentenbasiert – früher als klassischer Aufsatz auf Papier gedruckt und heute online als PDF [2]. Jährlich werden rund 2.5 Millionen neue Forschungsbeiträge publiziert und die meisten davon nur in PDF Format aufbewahrt [2]. Das führt zu einem großen Problem für die Forscher, das sie mit pseudodigitalisierter PDF-Publikationen überfordert sind. Als Folge davon wird die Forschung stark geschwächt [2]. Außerdem werden die meisten Daten der Forschung nicht nachhaltig aufbewahrt. Laut einer Online-Umfrage zu ingenieurwissenschaftlichen Aspekten in Datenmanagementplänen gilt folgendes. Langfristig (min. zehn Jahre), nicht öffentlich geteilte aufbewahrte Forschungsdaten verbleiben nach dem Projekt meistens auf den jeweiligen PCs oder Laptops [1].

¹D. Hausen, NFDI4Ing Special Interest Group (SIG) Qualitätssicherung und Metriken für FAIR-Daten 4. Treffen, Online-Umfrage zu ingenieurwissenschaftlichen Aspekten in Datenmanagementplänen, 2022

Eine solche Ablage erschwert oder macht es zum Teil gar nicht mehr möglich nach Fristablauf auf die Forschungsdaten zu zugreifen.

Die systematische Literaturrecherche (SLR) soll helfen dieses Problem zu lösen, weil ihre Aufgabe darin besteht alle verfügbaren Beweise zu einer bestimmten Forschungsfrage zu identifizieren, zu analysieren und zu interpretieren [11]. Jedoch hat diese Methode auch einen großen Nachteil - Systematische Literaturrecherche erfordert wesentlich mehr Aufwand als übliche Literaturrecherche. Brings et al. gehen davon aus, dass für eine effiziente Literaturrecherche, eine Kombination von verschiedenen Suchmethoden verwendet werden muss [5], welches den Zeitaufwand für eine qualitative Literaturrecherche enorm erhöht.

Ein weiterer Ansatz ist die systematische Mapping-Studie, bekannt auch als "Scoping-Studie". Damit wird ein breiteres Feld von Forschungen durchgesucht um ein Überblick über den "state-of-art" oder "state-of-practice" zu einem Thema zu verschaffen [24]. Im Vergleich zur SLR hat die Mapping-Studie einen breiteren Scope, andere Kriterien für Ein-/Ausschlüsse [24] und bietet die Möglichkeit, Lücken für neue Forschungen zu finden. Da die Mapping-Studie dem gleichen Prozess wie der SLR unterliegt, hat sie den gleichen Nachteil - es wird viel mehr Aufwand betrieben, als bei der üblichen Literaturrecherche.

Das Auffinden der notwendigen Literatur und der Vergleich von Ergebnissen aus verschiedenen Quellen ist ein sehr aufwändiger Prozess, der viel Zeit in Anspruch nimmt. Jedoch ist dieses essentiell für die Forscher, um alle vorhandene Information über ein Phänomen sorgfältig und unvoreingenommen zusammenfassen, allgemeine Schlussfolgerungen zu einem Phänomen zu ziehen oder Informationen als Auftakt für weitere Forschungsaktivitäten festzulegen [11].

Eines der zentralen Probleme von den oben genannten Ansätzen ist die Wiederverwendbarkeit. Wie schon vorher erwähnt, werden jedes Jahr viele neue wissenschaftliche Artikel veröffentlicht. Diesbezüglich ist es notwendig, SLR und/oder Mapping-Studien erneut durchzuführen, um aktuelle Daten für Studien oder Forschungen zu erhalten. Die Wiederverwendbarkeit hat dabei eine sehr große Bedeutung.

Die geringe Nachhaltigkeit der Daten führt zu einer erschwerten Recherche. Zudem lässt sich dadurch der Datenbestand nicht einfach ohne weiteres aufstocken. Aktuell findet in der Regel nach Veröffentlichung einer SLR ein Abgleich der vorher publizierten SLR statt, dabei werden die Rohdaten der jeweiligen SLR nicht betrachtet und somit vernachlässigt. Wenn die Rohdaten betrachtet werden sollen, muss die jeweilige SLR oder Systematic Mapping Study erneut durchgeführt werden. Dies ist mit sehr viel Zeit und Aufwand verbunden.

Daraus ergeben sich folgende Probleme:

- **Aufwand** Sowohl für die SLR als auch für die Mapping-study wird viel Zeit und Mühe benötigt. Falls das Problem gelöst wird, können diese Ressourcen in andere Aktivitäten integriert werden.
- **Nachhaltigkeit** Die extrahierten und analysierten Daten aus wissenschaftlichen Artikeln werden oft nur in Datentabellen zusammengetragen [14], die nicht nachhaltig aufbereitet und langfristig zur Verfügung gestellt werden. Eine Erweiterung der bestehenden Untersuchung ist somit kaum möglich und erfordert stattdessen meist eine Wiederholung des gesamten Extraktions- und Analyseprozesses.
- **Wiederverwendbarkeit** Sollte eine Literaturrecherche für das gleiche Forschungsgebiet viele Jahre später ein weiteres mal durchgeführt werden, weil z.B. neue Technologien und Methoden entwickelt wurden, wäre es von Vorteil, den Prozess der Suche nur für die neuesten Daten zu wiederholen. Die alten, nachhaltig aufbereiteten Daten könnten im Anschluss hinzugezogen werden um einen tiefen Einblick und eine umfassende Analyse zu erhalten.

In dieser Arbeit wird die Nutzung des ORKG für eine Literaturrecherche untersucht, um eine nachhaltige und wiederverwendbare Datenbasis aufzubauen mit der wiederum nachhaltige und wiederverwendbare Analysen ermöglicht werden sollen.

1.2 Ziel der Arbeit

Das übergeordnete Ziel, das mit dieser Arbeit verfolgt wird, kann mit Goal Definition Template [3] wie folgt beschrieben werden:

Hauptziel

*Das zentrale Ziel dieser Arbeit...
 ...ist die **Analyse des ORKG**
 für den Zweck der Evaluation seines Potentials zur Organisation (Aquirierung, Kuration, Veröffentlichung und Verarbeitung) von wissenschaftlichem Wissen
 in Bezug auf die Anwendbarkeit für die Erstellung und Analyse (Durchführung) einer Literaturrecherche mit nachhaltiger und wiederverwendbarer Datenbasis
 aus der Sichtweise des Forschers als ORKG Nutzer
 im Kontext von einem realen Use Case zum Thema Video Produktion und Nutzung im Requirements Engineering mittels einer von Experten kuratierten Liste von Publikationen.*

Das Hauptziel lässt sich in zwei weitere Nebenziele aufteilen:

Nebenziel 1.1

Ziel dieser Arbeit...

*...ist der **Ausbau und die Anwendung** eines Templates von ORKG für den Zweck zur Akquirierung von wissenschaftlichem Wissen über die Anwendung von Videos im Requirements Engineering aus wissenschaftlichen Publikationen*

in Bezug auf die Nachhaltigkeit der Datenbasis von wissenschaftlichen Publikationen

aus Sichtweise der Forscher

im Kontext von der Liste der Publikationen zum Thema „Production and Use of Videos Requirements Engineering“ im Open Research Knowledge Graph.

In dieser Arbeit wird demnach ein bestehendes Template (Video Process²) ausgebaut und an einer ORKG kuratierten Liste von Publikationen zum Thema “Production and Use of Videos Requirements Engineering” angewendet. Das Ziel dabei ist wissenschaftliches Wissen über die Anwendung von Videos im Requirements Engineering aus wissenschaftlichen Publikationen zu akquirieren, diese nachhaltig und langfristig mit dem ORKG zu kuratieren, veröffentlichen und zu verarbeiten. Das Template muss nur einmal erstellt werden und kann dann für mehrere wissenschaftlichen Artikel angewendet werden. Dadurch wird die gewünschte Nachhaltigkeit erreicht. Mit Ausbau des vorhandenen Templates sollten möglichst viel der wichtigen Aspekte des Video Prozesses zusammengefasst werden, indem der Video Prozess auf seine Bestandteile zerlegt wird. Somit sollte ein Gesamtbild von einem Video Prozess, das in einem Artikel stattgefunden hat, erschafft werden. Das bietet die Möglichkeit für die Forscher, nicht nur die benötigte Literatur mit wenigem Zeitaufwand zu finden, sondern auch die Ergebnisse von entsprechenden Video Prozessen aus mehreren Quellen schnell zu vergleichen. Außerdem ist dadurch die Möglichkeit gegeben, Publikationsinformation und die dazugehörigen Daten zusammenzuhalten und diese langfristig abrufbar zu machen.

Dadurch, dass mit dem oben genannten Template gewonnene Daten strukturiert sind, wird die Suche nach bestimmten Ergebnissen aus passenden Artikel deutlich erleichtert. Ein weiterer Vorteil besteht darin, dass die Daten abrufbar und maschinenlesbar werden. So können Skripte/Programme geschrieben werden, die immer wieder durchgeführt werden können und damit die bestehende und ggf. erweiterte Datenbasis zusammen abzurufen und zu analysieren. Dadurch wird eine kontinuierliche Analyse aller Daten über den gesamten Zeitraum erschafft. Somit geht die Datenbasis nicht mehr verloren und wird immer berücksichtigt, wenn ein neues Paper in der kuratierten Liste mit aufgenommen und beschrieben

²<https://orkg.org/template/R172526>

wird. Daraus ergibt sich ein weiteres Nebenziel der Arbeit:

Nebenziel 1.2

Ziel dieser Arbeit...

*...ist die **Analyse** der kuratierten Daten*

***für den Zweck** einen Überblick über den State-of-the-art zu erhalten*

***in Bezug auf** die Wiederverwendbarkeit der Daten durch eine entsprechende Analyseprozedur*

***aus der Sichtweise** des Forschers als ORKG Nutzer*

***im Kontext** aus der Liste der Publikationen zum Thema „Production and Use of Videos Requirements Engineering“ im Open Research Knowledge Graph.*

Des Weiteren können die erfassten Daten mit einem Jupyter Notebook graphisch dargestellt werden, um eine Analyse der Daten zu ermöglichen und zu vereinfachen. Dieses befreit die Forscher von unnötiger Suche nach bestimmten Parametern oder Einzelheiten. Es ist besonders hilfreich, wenn eine größere Anzahl der Daten überprüft und analysiert werden muss. Es kann auch einen Überblick auf die gesamte wissenschaftliche Literatur, die mit dem Template versorgt wurde, erschafft werden. Damit soll der Prozess der Suche nach bestimmter Literatur und deren Vergleich noch weiter vereinfacht werden. Dadurch, dass das Notebook immer wieder neu ausgeführt und erweitert werden kann, wird Wiederverwendbarkeit erreicht. Die Daten, die mit Hilfe von dem Template eingetragen werden, stellen die Nachhaltigkeit dar, weil sie in einer festen Struktur langfristig bereitgestellt werden.

1.3 Struktur der Arbeit

Die Arbeit ist in sieben Kapitel aufgeteilt und wie folgt strukturiert. Im Kapitel [2](#) werden grundlegende Definitionen, Konzepte und Ansätze erläutert. In Kapitel [3](#) wird das eingesetzte Template erklärt, welchen Zweck es dabei erfüllt, wie es strukturiert ist und wie es weiterentwickelt wurde. Außerdem werden die Vor- und Nachteile dieses Template-Ansatzes gelistet. Weiterhin wird in Kapitel [4](#) die erfolgte Analyse der Daten sowie ihre Umsetzung beschrieben. Am Ende erfolgt die Auswertung des Templates - Ansatzes und der Ergebnisse der durchgeführten Analyse in Bezug auf Nachhaltigkeit der Akquirierung, Kuratierung, Veröffentlichung, Verarbeitung und Zeitaufwand. Im Anschluss in Kapitel [6](#) werden verwandte Arbeiten vorgestellt und von dieser Arbeit abgegrenzt. Im letzten Kapitel [7](#) wird ein Fazit gezogen und ein Ausblick vorgestellt.

Kapitel 2

Grundlagen

2.1 Etablierte Methoden zur Analyse von Literatur

In dieser Arbeit wird aufgezeigt wie die Suche nach bestimmten Daten aus wissenschaftlichen Artikeln und deren Aufbewahrung im Bezug auf Nachhaltigkeit und Wiederverwendbarkeit verbessert werden kann. Dafür ist es notwendig einen Überblick über das zentrale Verfahren zur Analyse von Literatur zu bekommen, nämlich “systematische Literaturrecherche” und “systematische mapping Studie”. Außerdem wird im weiteren Verlauf aufgezeigt wie aufwändig diese etablierten Verfahren sind und welche Vorteile mit ORKG und den hier betrachteten Methoden geschafft werden können.

2.1.1 Systematische Literaturrecherche

Eine systematische Literaturrecherche (oft als systematisches Review bezeichnet) ist ein Mittel zur Identifizierung, Bewertung und Interpretation aller verfügbaren Forschungsergebnisse, die für eine bestimmte Forschungsfrage, ein bestimmtes Themengebiet oder ein interessierendes Phänomen relevant sind. Einzelne Studien, die zu einem systematischen Review beitragen, werden Primärstudien genannt; ein systematisches Review ist eine Art Sekundärstudie [11]. Um zu zeigen wie aufwändig die systematische Literaturrecherche sein kann, müssen die Schritte der Suche genauer betrachtet werden. Erst dann kann der Prozess komplett verstanden werden. Die Schritte sind in der Abbildung 2.1 dargestellt.

Laut Kitchenham und Charters besteht die systematische Literaturrecherche aus drei Schritten: *Planung*, *Durchführung* und *Berichterstattung*. Dabei ist jeder Schritt in weitere mehrere Schritte unterteilt [11].

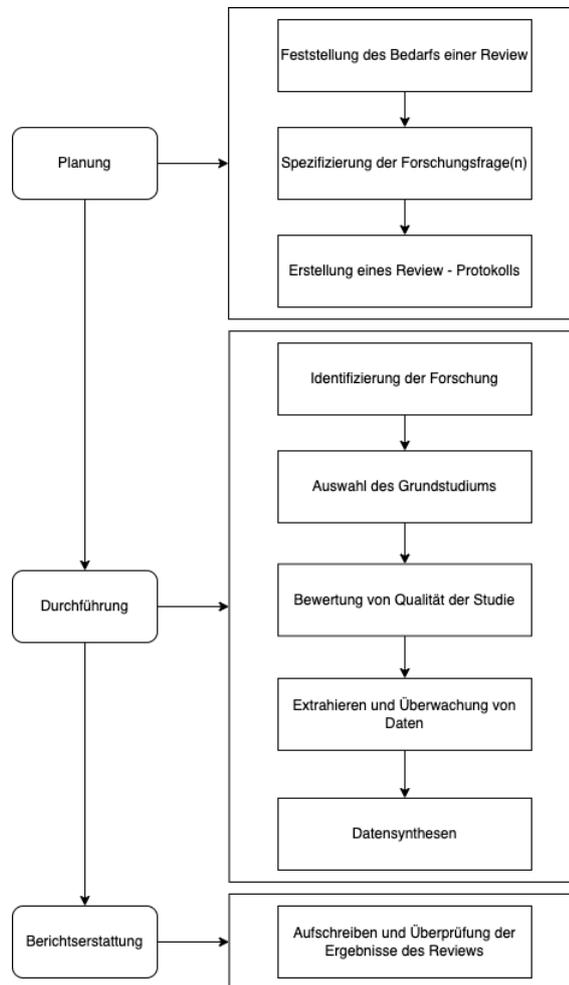


Abbildung 2.1: Systematischer Literaturrecherche Prozess

Die *Planung* der systematischen Literaturrecherche besteht aus drei Schritten:

- **Feststellung des Bedarfs einer Review.**

Der Bedarf eines systematischen Reviews entsteht von Forschern, die den “state-of-the-art” in einem Bereich verstehen möchte. Ebenso von Praktikern, die empirische Beweise für ihren strategischen Entscheidungsprozess oder ihre Verbesserungsaktivitäten verwenden möchten.

Wenn einige systematische Literaturrecherchen schon auf dem Gebiet verfügbar sind, sollten sogenannte tertiäre Studien durchgeführt werden. Dabei wird der Umfang und die Qualität der bereits vorhandenen Literaturrecherche bewertet. Das Ziel hierbei ist es herauszufinden, ob die herangezogen Informationen ausreichend sind, um den aktuellen

Bedarf an einer Review zu decken [24].

- **Spezifizierung der Forschungsfrage(n)**

Der Bereich des systematischen Reviews und die spezifischen Forschungsfragen setzen den Fokus für die Identifizierung der Primärstudien, die Datenextraktion aus den Studien und die Analyse. Dies setzt gut formulierte und durchdachte Forschungsfragen voraus [11].

- **Erstellung eines Review - Protokolls**

Das Reviewprotokoll definiert die Vorgehensweise bei der systematischen Literaturrecherche. Es dient auch als Protokoll für die Durchführung der Überprüfung [24].

Die *Durchführung* des Review ist ein sehr umfangreicher Schritt, welcher das Reviewprotokoll in die Praxis umsetzt. Dieser besteht auch aus mehreren Schritten:

- **Identifizierung der Forschung**

Dieser Schritt besteht darin, die sogenannten “Search strings” zu erstellen, die zu spezifizieren und sie auf Datenbanken anzuwenden. Die Suchergebnisse und ein Protokoll der durchgeführten Aktionen werden gespeichert [24].

- **Auswahl des Grundstudiums.**

Grundlage für die Auswahl der Primärstudien sind die Ein- und Ausschlusskriterien. Die Kriterien sollten vorher festgelegt werden. Die identifizierten Studienkandidaten werden in Bezug auf die Auswahlkriterien verarbeitet [24].

- **Bewertung von Qualität der Studie**

Mit dem Schritt wird die Qualität der Primärstudien bewertet. Es ist insbesondere dann wichtig, wenn die Studien widersprüchliche Ergebnisse berichten [24].

- **Extrahieren und Überwachung von Daten**

Sobald die Liste der Primärstudien festgelegt ist, werden die Daten aus den Primärstudien extrahiert. Es wird ein Formular für das Extrahieren der Daten entworfen, welches zum Sammeln der erforderlichen Informationen aus den primären Studienberichten dient [24].

- **Datensynthesen**

Bei der Datensynthese werden die Ergebnisse der eingeschlossenen Primärstudien zusammengetragen und zusammengefasst [11].

Nachdem *Planung* und *Durchführung* fertig sind, müssen alle Ergebnisse des Review aufgeschrieben, überprüft und den potentiellen Interessenten vermittelt werden. Damit befasst sich der letzte Schritt der systematischen Literatursuche - *Berichterstattung*.

2.1.2 Systematische Mapping-Studie

In diesem Kapitel werden Grundlagen zu dem Thema systematische Mapping-Studie erläutert und erklärt weshalb sie für diese Arbeit wichtig sind.

Die systematische Mapping-Studie beschäftigt sich mit der Suche nach bestimmten wissenschaftlichen Daten. Der Fokus für die Ablage der Rohdaten und das Suchergebnis wird dabei stark vernachlässigt. Da ein Ziel dieser Bachelorarbeit die Vereinfachung und Verbesserung der Suche sowie Aufbewahrung von Daten ist, ist es notwendig die Vorgehensweise der systematischen Mapping-Studie zu erklären.

Systematische Mapping-Studien (auch als Scoping-Studien bekannt) sollen einen breiten Überblick über ein Forschungsgebiet ermöglichen, feststellen ob Forschungsnachweise zu einem Thema vorliegen und einen Hinweis auf die Menge der Nachweise geben. Die Ergebnisse einer Mapping-Studie können Bereiche identifizieren, die sowohl für die Durchführung systematischer Literaturrecherchen geeignet sind, als auch die, für die Primärstudie durchgeführt werden können [11].

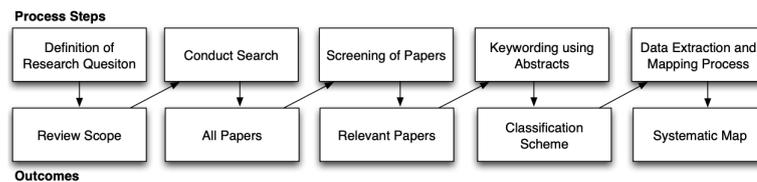


Abbildung 2.2: Systematischer Mapping Prozess nach Petersen et al. [19]

Petersen et al. [19] teilen den Prozess der systematischen Mapping-Studie in fünf Punkte auf, die auf der Abbildung 2.2 wie folgt, zu sehen sind: *Definition der Forschungsfrage*, *Durchführung der Suche*, *Screening der Paper*, *Keywording des Abstracts*, *Extrahieren der Daten und Mapping Prozess*.

- **Definition der Forschungsfrage**

Erst wird die Forschungsfrage definiert und anschließend ein Scope für das Review erstellt.

- **Durchführung der Suche**

In dem Schritt werden "Search Strings" erstellt und an Datenbanken angewendet. Als Ergebnis werden alle, für die gestellte Forschungsfrage relevanten, Papers zurückgeliefert. Search String spielt dabei eine sehr große Rolle, denn dieser beeinflusst direkt die zurückgelieferten Ergebnisse.

- **Screening der Paper**

Das Screening der Paper dient dazu, die Ein- und Ausschlusskriterien zu ermitteln und diese dann für die relevanten Paper anzuwenden.

- **Keywording des Abstracts**

Das *Keywording* ist eine Methode, die den Zeitaufwand für die Entwicklung des Klassifikationsschemas reduziert sowie das Schema der vorhandenen Studien berücksichtigt und sicherstellt. Keywording erfolgt in zwei Schritten. Zunächst lesen die Reviewer die Abstracts und suchen nach Schlüsselwörtern und Konzepten, die den Beitrag der Arbeit widerspiegeln. Dabei identifiziert der Reviewer den Kontext der Forschung. Wenn dies geschehen ist, werden die Schlüsselwörter aus verschiedenen Artikeln miteinander kombiniert, um ein umfassendes Verständnis über die Art und den Beitrag der Forschung zu entwickeln. Das Ergebnis vom *Keywording des Abstracts* ist ein Klassifikationsschema.

- **Extrahieren der Daten und Mapping Prozess**

Bei dem Schritt findet die eigentliche Extraktion der Daten statt. Das Klassifikationsschema wird während der Extraktion der Daten weiterentwickelt. Z.B werden neue Kategorien hinzugefügt, zusammengeführt oder bereits in bestehende Kategorien aufgeteilt.

Auf den Ersten Blick sehen die SLR und die systematische Mapping-Studie sehr ähnlich aus, jedoch sind sie in der Praxis sehr unterschiedlich. In der Arbeit von Kitchenham und Chartes [11] wurden wesentliche Unterschiede zwischen der SLR und der systematischen Mapping-Studie zusammengefasst, z.B.:

- Mapping-Studien haben im Allgemeinen umfassendere Scope und stellen oft mehrere Forschungsfragen im Vergleich zur SLR.
- Die Suchbegriffe für Mapping-Studien sind weniger stark fokussiert als bei der SLR und können eine sehr große Anzahl von Studien während der Suchphase zurückgeben. Das Ziel der Mapping-Studie ist, eine breite Abdeckung zu erreichen. Die SLR hingegen arbeitet auf einen engen Fokus zu.
- In der Analysephase einer systematischen Mapping-Studie geht es darum, die Daten zusammenzufassen, um die gestellten Forschungsfragen zu beantworten. Normalerweise enthält die Mapping-Studie keine Techniken der depth Analyse, Metaanalyse und narrative Synthese, sondern nur Zusammenfassungen.

2.2 ORKG als neue Methode zur Literaturanalyse

ORKG ist eine Alternative zu den bereits vorgestellten Methoden. Das ORKG macht wissenschaftliches Wissen sowohl Menschen als auch für Maschinen nutzbar und ermöglicht damit neue Wege der maschinellen Assistenz. Dies sollte den Forschern dabei helfen, relevante Beiträge zu ihrem Fachgebiet zu finden und aktuelle Vergleiche und Reviews zu erstellen [\[1\]](#).

2.2.1 Was ist der ORKG?

Research knowledge Graph wird wie folgt definiert: Ein Research Knowledge Graph repräsentiert wissenschaftliches Wissen semantisch, d. h. explizit und formal, indem es vorhandene (Meta-)Daten von wissenschaftlichen Artefakten (Publikationen, Datensätzen und Quellcodes) und Entitäten (Personen und Organisationen) auf nachvollziehbare Weise verknüpft [\[9\]](#). Dadurch, dass das wissenschaftliche Wissen semantisch definiert ist, ermöglicht es die Daten auch für Maschinen lesbar zu machen. Dieses bietet dadurch ein großes Spektrum der Möglichkeiten, z.B. eine automatisierte, wiederverwendbare Analyse der Daten. Außerdem bietet ORKG mit seinem Template Feature eine Lösung für die nachhaltige Aufbereitung und Akquirierung der Daten, da das Template genau die semantische Analyse der Daten darstellt. Zusammenfassend kann die Schlussfolgerung gezogen werden, dass das ORKG ein perfekter Kandidat für nachhaltige Aufbereitung der Daten und Akquirierung von wissenschaftlichem Wissen ist. Zu dem ermöglicht es die Analyse der Daten wiederwendbar gestalten.

2.2.2 Semantische Struktur

Um zu verstehen, was die semantische Struktur bedeutet, sollte erst der Begriff Semantik erläutert werden. Semantik kann als die Studie der Bedeutung definiert werden. Als Teilgebiet der Linguistik, untersucht die Semantik die Bedeutung lexikalischer Elemente [\[13\]](#).

Die allgemeine Definition der semantischen Struktur laut Lyons ist: “eine Aussage, die für alle Wörter gilt, ob sie sich auf Objekte oder Merkmale der physischen Welt beziehen oder nicht” [\[13\]](#). Aus Sicht der Psycholinguisten repräsentiert eine semantische Struktur das Wissen über die Bedeutung lexikalischer Elemente [\[7\]](#).

Die im Rahmen dieser Arbeit betrachteten Publikationen werden mit Hilfe von semantischer Struktur abgearbeitet, um einzelne Information daraus extrahieren zu können. Das dient auch als Grundlage für das Template, welches innerhalb dieser Arbeit ausgebaut und an wissenschaftlichen Publikationen angewendet wird.

¹<https://orkg.org/about/1/Overview>

Ein weiterer Punkt sind Ontologien, die im Zusammenhang mit semantischer Struktur auch für diese Arbeit und genau für das hier verwendete Template wichtig sind. Ontologien erlauben verschiedene Konzepte, ihre Beziehungen und die Regeln der Domäne in maschinenlesbaren Form darzustellen [6]. Außerdem ermöglichen sie Anwendungen zu erstellen, die nicht nur die Struktur der Daten erfassen und verarbeiten, sondern auch die Semantik von Informationen [6]. Dieses Ziel soll mit dem Template verfolgt und umgesetzt werden. Furrer [6] unterteilt die Entwicklung und Verwendung des Begriffes „Ontologie“ in drei Phasen:

- **Altertum**

Die Ontologie dient als ein Ausdrucksmittel der allgemeinen Metaphysik mit dem Zweck, das Grundgerüst zur Formulierung der philosophischen Theorien mit Grundstrukturen des Seienden als Hauptpunkt bereitzustellen [6].

- **18. Jahrhundert**

Der Zweck der Ontologie in diesem Fall ist die Wissensrepräsentation. Die Ontologie wird gemäß den Regeln einer wissenschaftlichen Methode gebildet. Die Begriffe werden klar und eindeutig definiert und sollten genügend Merkmale aufweisen, um als Ontologie erfasst zu werden [6].

- **Heutige Phase**

Ontologien können durch Computerprogramme bearbeitet werden. Zweck der Ontologien ist die Wissensrepräsentation mit Möglichkeit das Wissen zusätzlich zu verarbeiten, z.B. durch logische Schlussfolgerungen, Beweisführungen oder Konsistenzprüfungen. Um das zu ermöglichen, ist eine feste logische Basis sowie eine formale Beschreibungssprache vorausgesetzt [6].

Der letzte Schritt stellt ein mögliches Vorgehen, das innerhalb dieser Arbeit mit Hilfe von einem Template untersucht wird, dar.

2.2.3 Vorteile vom ORKG gegenüber den bisher etablierten Methoden

Die Unterschiede zwischen der systematischen Mapping-Studie und der systematischen Literaturrecherche wurden bereits im Kapitel 2.1.2 beschrieben. In diesem Kapitel wird angezeigt, welche Vorteile ORKG im Vergleich zu den bereits erwähnten Methoden anbietet.

- **Aufwand**

Die systematische Mapping-Studie und die systematische Literaturrecherche sind sehr aufwändig. In dem Kapitel 2.1.2 und 2.1.1 wurde bereits das Vorgehen der beiden Methoden dargestellt. Anhand deren Schritte kann die Schlussfolgerung gezogen werden, dass die beiden

Methoden sehr aufwändig sind. Im Gegenteil dazu bietet ORKG die Möglichkeit die sogenannten Templates zu erstellen und die Papers mit den entsprechenden Templates zu versorgen. Die Papers mit den Templates können dann mithilfe von “comparison” Feature verglichen werden. Dieses schafft einen Überblick über die ganzen Paper, die zum Vergleich hinzugefügt wurden.

- **Nachhaltigkeit**

Die wissenschaftlichen Daten werden meist in Form von PDF Dateien abgespeichert [2, 12]. Damit wird keine Nachhaltigkeit geboten, da das Format für Maschinen schwer zu verarbeiten ist [12]. Außerdem PDF-Dateien konzentrieren sich auf die *visuelle* Darstellung von Dokumenten anstelle einer *strukturierten* Darstellung, was die Zugänglichkeit erschwert [17]. Die Ergebnisse der systematischen Mapping-Studie sowie der systematischen Literaturrecherche sind nicht anders. Im Gegensatz zu den bereits vorgestellten Methoden, bietet ORKG sogenannte Templates. Wird eine Publikation mit einem Template versorgt - entsteht eine semantische Struktur der jeweiligen Publikation. Damit werden die Daten immer nachhaltig und strukturiert abgespeichert. Zusätzlich stehen die Daten jederzeit zur Verfügung, da sie online aufbewahrt werden.

- **Wiederverwendbarkeit**

Wie bereits erwähnt wurde, wird mit Hilfe von Templates eine semantische Struktur der Publikation erzeugt. Damit werden die Identifizierung, Nachvollziehbarkeit und Klarheit der Begriffe erreicht, sowie dessen Mehrdeutigkeit reduziert. Diese Verbesserungen wiederum führen zu einer einfacheren Wiederverwendung wissenschaftlicher Erkenntnisse. Somit wird Redundanz und Doppelarbeit reduziert. Auch wiederholende Inhalte von Publikationen, wie z.B verwandte Arbeiten, können im Laufe der Zeit immer wieder verwendet und erweitert werden. Zusätzlich wirken sie so der ständig wachsende Anzahl von Publikationen entgegen [9]. Dadurch, dass die Daten maschinenlesbar sind, wird die Gelegenheit erschafft die Daten mit wiederverwendbaren Programmen und/oder Skripten zu erfassen und zu analysieren. Mit der systematischen Mapping-Studie oder systematischen Literaturrecherche ist dies nicht möglich.

2.3 Requirements Engineering

In Rahmen dieser Arbeit werden Publikationen zum Thema “Production and Use of Videos Requirements Engineering” betrachtet. Daher ist es essentiell die Bedeutung von Requirements Engineering zu verstehen. In diesem Kapitel wird ein Überblick über Requirements Engineering geschaffen

und der Begriff erläutert.

Nach International Requirements Engineering Boards (IREB) wird das Requirements Engineering wie folgt definiert:

Requirements Engineerings nach IREB

Das Requirements Engineering ist ein systematischer und disziplinierter Ansatz zur Spezifikation und zum Management von Anforderungen mit dem Ziel, die Wünsche und Bedürfnisse der Stakeholder zu verstehen und die Gefahr zu minimieren, ein System auszuliefern, welches diese Wünsche nicht erfüllt [20].

Requirements Engineering kann in zwei Bereiche aufgeteilt werden. Der erste Teilbereich ist die Requirements Analysis, welche sich mit der Spezifikation der Anforderung befasst. Der zweite Teilbereich ist das Requirements Management. Dieser beschäftigt sich mit der Verwaltung der Anforderungen [21]. Börger et. al. [4] haben ein Referenzmodell erstellt, welches die Aufteilung des Requirements Engineering aufzeigt. (siehe Abbildung 2.3).

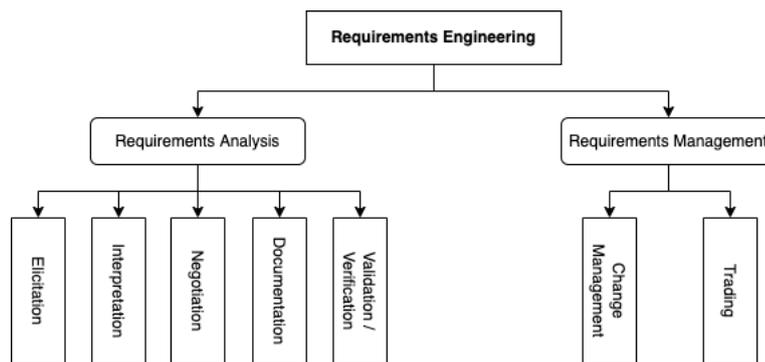


Abbildung 2.3: Referenzmodell des Requirements Engineering [4]

In diesem Fall ist Requirements Management nicht relevant und wird nicht weiter betrachtet. Die Requirements Analysis dagegen hat aufgrund seiner fünf Aktivitäten eine große Bedeutung für diese Arbeit. Diese Aktivitäten sind im verwendeten Template zu finden und müssen deswegen genauer beschrieben werden.

Nach dem Referenzmodell des Requirements Engineering besteht die Requirements Analysis aus fünf Aktivitäten: *Elicitation*, *Interpretation*, *Negotiation*, *Documentation* und *Validation/Verification*. Im folgenden Text werden die einzelnen Aktivitäten erklärt.

- In der der Aktivität *Elicitation* werden Anforderungen an das System von verschiedenen Stakeholdern gesammelt. Außerdem werden in dieser Phase Grenzen, Ziele und Aufgaben ermittelt sowie Stakeholder identifiziert [15].

- In *Interpretation* Phase werden, die während der Elicitation gesammelte Daten, bearbeitet. Da, die in der Elicitation erfassten Anforderungen nur grob dargestellt sind, müssen sie genauer definiert und beschrieben werden. Als erster Schritt werden die “grob-erfassten-Anforderungen” identifiziert. Mit dem zweiten Schritt wird entschieden, welche davon wichtig und welche nur nebensächlich sind. Abschließend werden sie strukturiert und nach Ähnlichkeit gruppiert, damit sich ergänzende Anforderungen verschmelzen lassen können [21].
- Bei der *Negotiation* verhandeln die Requirements Engineers mit den Stakeholdern alle Anforderungen um die Definitionen und die Dokumentation der Anforderungen zu vereinbaren [15].
- Die Aktivität *Documentation* hat das Ziel alle Anforderungen nieder zu schreiben. Die Anforderungen werden in Einzelanforderungen zerlegt und mit passenden Attributen beschrieben. Außerdem werden die Anforderungen mit der Dokumentation verbunden und mit gewissen Einzelheiten wie z.B. Absichten und Rollen versehen [21].
- Bei der *Validation/Verification* wird geprüft, ob die Anforderungen den Bedarf des Systems richtig und genau repräsentieren und ob sie vollständig, korrekt und konsistent sind. Die Anforderungen werden außerdem von Quality Assurern oder technischen Experten überprüft [15].

Kapitel 3

Ausbau und Anwendung des Video Process Templates

In diesem Kapitel wird ein sehr wichtiges Feature der ORKG, nämlich das Template, näher und detailliert betrachtet. Zudem werden Vor- und Nachteile des Ansatzes erläutert. Außerdem wird beschrieben, wie das bereits entwickelte Template “Video process”, im Rahmen dieser Arbeit angewendet und ausgebaut wurde.

3.1 Ziel der semantischen Analyse

Das Hauptziel des Template-Ansatzes ist eine wiederverwendbare und nachhaltige Akquirierung und Aufbereitung der wissenschaftlichen Daten zu erreichen. Mit dem Template wird eine semantische Struktur der wissenschaftlichen Publikation geschaffen. Dabei spielen die Ontologien eine signifikante Rolle. Wie bereits im Kapitel [2.2.2](#) erwähnt wurde, erlauben Ontologien verschiedene Konzepte, ihre Beziehungen und die Regeln der Domäne in maschinenlesbare Formen darzustellen [\[6\]](#). Des Weiteren ermöglichen die Ontologien solche Anwendungen zu erstellen, die nicht nur die Struktur der Daten erfassen und verarbeiten, sondern auch die Semantik von Informationen [\[6\]](#). Dadurch, dass die Daten maschinenlesbar werden, wird das oben genannte Ziel erreicht.

3.2 Vorteile und Nachteile des Ansatzes

Anhand des Bedarfs von Wissenschaftler/-in und/oder des Forschungsfelds wird ein bestimmtes Template entwickelt. Das Template kann sowie vordefinierte als auch benutzerdefinierte Properties enthalten. Die Properties werden anschließend mit Daten aus den dazugehörigen wissenschaftlichen Publikationen ausgefüllt, wie z.B “Equipment”. Die Properties können nicht nur andere Properties oder bestimmte Datentypen, wie Boolean oder String,

sondern auch andere Templates enthalten. Das macht die ganze Struktur sehr flexibel.

Mit der semantischen Struktur der Publikation wird nicht nur die Identifizierung, Nachvollziehbarkeit und Klarheit der Begriffe erreicht, sondern auch Mehrdeutigkeit reduziert [9]. Dadurch, dass das Template für bestimmte Zwecke und/oder Forschungsgebiete nur ein Mal erstellt wird, und dann auf alle gewünschten wissenschaftlichen Publikationen angewendet werden kann, wird auch Doppelarbeit reduziert.

In dieser Arbeit wurde bereits erwähnt, dass mit Hilfe von Templates und semantischer Struktur, die das Template anbietet, die Daten maschinenlesbar werden. Diese bietet viele weitere Möglichkeiten, wie z.B. die Analyse der Daten mit bestimmten Skripts und Programmen.

Jedoch hat dieser Ansatz nicht nur Vorteile. Ein Hauptnachteil besteht darin, dass das Template erst entwickelt werden muss. Die Entwickler des Templates sollten nicht nur viel Zeit in die Entwicklung des Templates investieren, sondern sich sehr gut mit dem Einsatzgebiet des Templates auskennen.

Ein weiterer Nachteil ist, dass die Daten aus wissenschaftlichen Publikationen (bisher) manuell in das Template eingetragen werden. Diese bringt sehr viel Zeitaufwand mit.

3.3 Initial Template

Im diesen Kapitel wird das Ausgangs-Template mit all seinen Properties beschrieben, welches bereits von Herrn Dr. rer. nat. Oliver Karras und Forschern des Fachgebietes Requirements Engineering entwickelt wurde.

Das “Video Process” Template [1] besteht aus mehreren Properties. Viele von den Properties stellen wiederum selbst ein Template dar. Die Hauptproperties sind: *Production*, *Application*, *Research problem* und *has Study*. Folgend werden alle Properties näher betrachtet.

Production

Die Production Property ist wie folgt definiert: “Der Prozess der Produktion von Rohstoffen (= Öl, Metalle, Pflanzen usw.) oder der Herstellung von Gütern und Produkten” [2]. Die Production Property ist selbst ein Template, welches aus mehreren Properties *Duration*, *Producer*, *guideline* und *equipment*, wie auf der Abbildung 3.1 dargestellt ist, besteht. Die Rechtecke mit abgerundeten Ecken beschreiben Ressourcen und die anderen atomare Daten (wie z.B String oder Boolean).

Duration ist als “Länge der Zeit, die etwas dauern kann” definiert [3]. Diese besteht aus *Number* - Anzahl und *Unit* - Name der Unit (z.B Hours, Minutes,

¹<https://orkg.org/template/R172526>

²<https://orkg.org/property/P45119>

³<https://orkg.org/property/P23149>

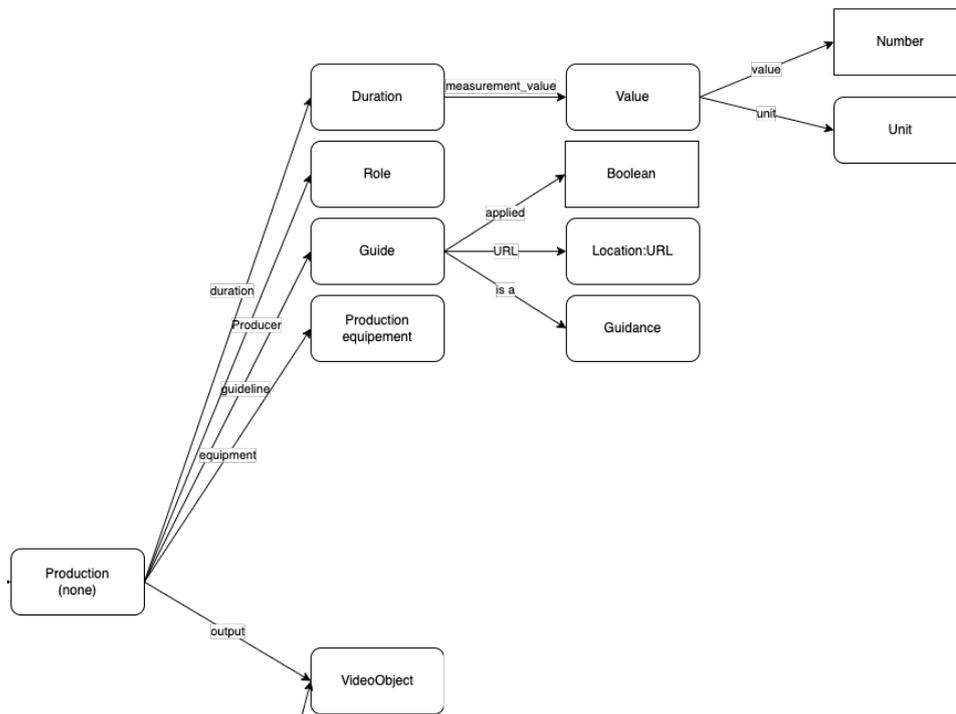


Abbildung 3.1: Production Property des Video Process Templates

Seconds) und stellt einen messbaren Wert dar.

Producer ist definiert als “Die Person oder Organisation, die das Werk produziert hat (z. B. Musikalbum, Film, Fernseh-/Radioserie usw.)”⁴. *Guideline* wird als “Informationen, die dazu bestimmt sind, Menschen zu beraten, wie etwas getan werden sollte oder was etwas sein sollte” beschrieben⁵.

Diese Property besteht aus einem Template, welches als *Guide* benannt wurde. *Guide* wiederum besteht aus den Properties *url*, eine *applied*-boolean Wert und *Is a*. Die *url* ist die Adresse zu *Guideline*, falls eine vorhanden ist. *applied*-boolean bestimmt ob die *guideline* angewendet wurden oder nicht. *Is a* beschreibt die *Guideline* selbst.

Eine weitere Property des *Production* Templates ist das *Production equipment*. Dieses existiert um alle, während der Herstellung des Videos, notwendigen Ausrüstungen aufzulisten. Die letzte Property von *Production* ist das *video* selbst, die im Template als “*VideoObject*” erscheint.

Das *Video Object* Template hat mehrere Eigenschaften, die auf der Abbildung 3.2 zu sehen sind. *Duration* wurde bereits mit der Beschreibung des *Production* Templates erklärt. *Presentation* - beschreibt wie/in welcher Form das Video dargestellt wird (z.B als eine Bildschirmaufnahme,

⁴<https://orkg.org/property/P45077>

⁵<https://orkg.org/property/P45079>

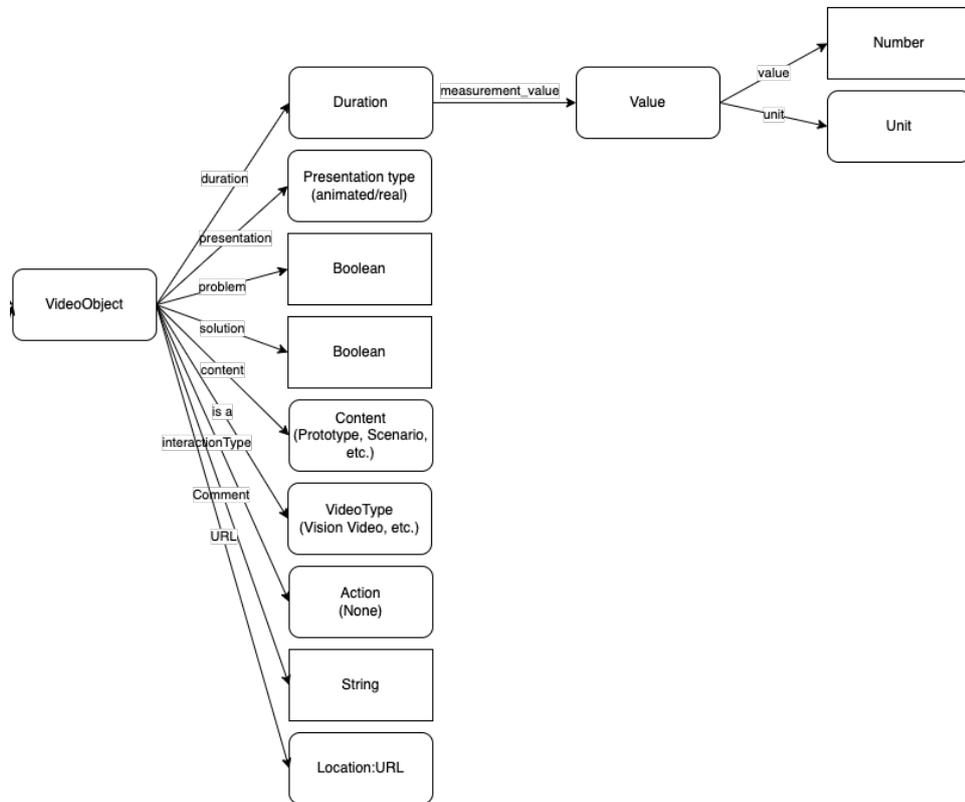


Abbildung 3.2: VideoObject Property

Animation oder Realität). *Problem* und *Solution* können entweder True oder False sein und zeigen, ob das Video ein Problem oder Lösung eines Problems darstellt. Das *Is a* enthält ein VideoType Class, der entsprechend Type des Videos repräsentiert (z.B Vision Video, Scenario Film usw.). Der Block *Content* stellt den Inhalt des Videos dar (z.B Vision, Scenario usw.). *InteractionType* zeigt wie mit dem Video interagiert wird. *url* enthält ein URL auf das Video als String. *Comment* ist ein Kommentar zu dem Video.

Application

Die Application Property besteht ebenfalls aus mehreren Komponenten, wie auf der Abbildung [3.3](#) dargestellt. Das *VideoObject* ist genau dasselbe VideoObject wie bei **Production**. *Audience* beschreibt die Gruppe, die für das Video geschaffen wurde. Das *Equipment* enthält die Ausrüstung, die für die Anwendung des Videos nötig ist. Das *context* stellt den Kontext von dem Video dar. Die *context*-Property ist ein eigenes Template und wird im weiteren Verlauf detaillierter beschrieben.

Das *Context* Template besteht aus acht Eigenschaften. Die erste Eigenschaft ist *phase*. Die *phase* besteht aus Phasen der Requirements Engineering und zeigt in welche der Phasen das Video angewendet wird.

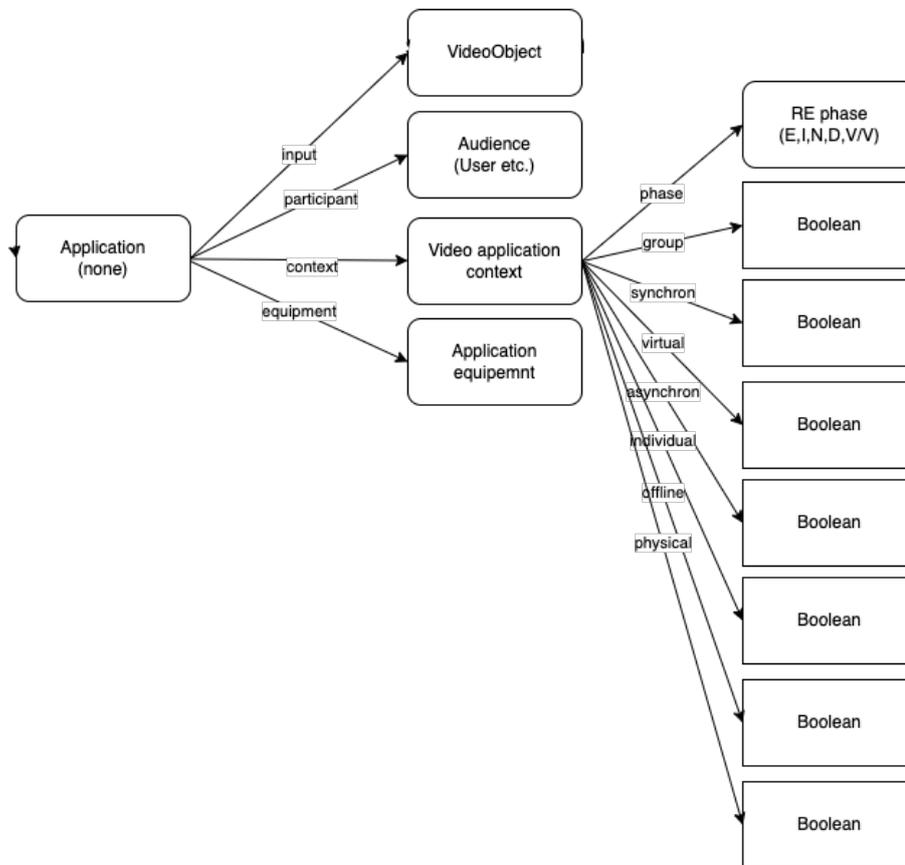


Abbildung 3.3: Application Property des Video Process Templates

Alle andere Eigenschaften: *group*, *individual*, *synchronous*, *asynchronous*, *virtual* und *physical* können entweder True oder False sein und zeigen ob die Eigenschaft vorhanden ist oder nicht. *Group* oder *individual* zeigt ob die Video in Gruppen oder Individuell angewendet werden. *Synchronous* und *asynchronous* beschreiben ob für das Video gleichzeitige Interaktion mehrerer Teilnehmer erforderlich ist. *Virtual* ist wie folgt definiert: “Wird mithilfe von Computertechnologie über das Internet durchgeführt, ohne das Personen physisch irgendwohin gehen”⁶. Dagegen ist mit *physical* folgendes gemeint: “In Bezug auf Dinge, die Sie sehen oder berühren können, oder in Bezug auf die Naturgesetze, in denen Menschen physisch irgendwohin gehen”⁷.

Research problem

Research problem beschreibt das Forschungsproblem, welches in der Publikation angesprochen wurde. *Research problem* besteht aus dem “*Problem*”

⁶<https://orkg.org/property/P45103>

⁷<https://orkg.org/property/P45104>

Template (Abbildung 3.4), das folgende Eigenschaften hat: *description* - Beschreibung von Problem, *same as* - Forschungsproblem das dem Problem ähnlich ist und *Sub problem* - ein Sub-Problem des Hauptforschungsproblems.

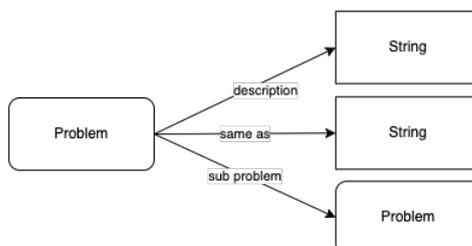


Abbildung 3.4: Problem Template

Has Study

Has Study Property des Video Process Templates, zeigt ob eine Studie im Rahmen der wissenschaftlichen Publikation durchgeführt wurde. Sie besteht aus *study* Template (Abbildung 3.5), welches wiederum aus zwei weiteren Eigenschaften besteht: *Study type* - die Art der durchgeführten Studie und *Number of participants* - Anzahl der Teilnehmer.

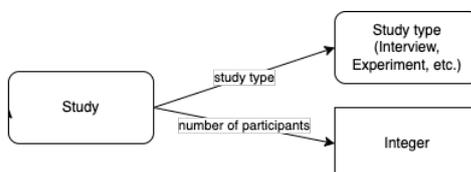


Abbildung 3.5: Study Template

3.4 Ausbau des Templates

Nachdem die, vom Betreuer dieser Arbeit, zur Verfügung gestellten Publikationen gelesen wurden, hat sich gezeigt, dass das Template bereits fast alle Aspekte des Video Prozesses beschreibt. Jedoch wurde in acht Publikationen (Anhang A: [9, 14, 32, 37, 41, 42, 47, 49]), die Kosten der Production von den Videos erwähnt. In sieben Publikationen (Anhang A [5, 9, 13, 14, 30, 37, 46]) wurde auch das Fidelity der Videos beschrieben. Daher wurden vier neue folgenden Properties erstellt. *Production costs* wurde mithilfe von *Low-Costs* und *High-Costs* dargestellt (Abbildung 3.6). Und *Video fidelity* wurde mithilfe von *High-Fidelity* und *Low-Fidelity* Properties beschrieben (Abbildung 3.7).

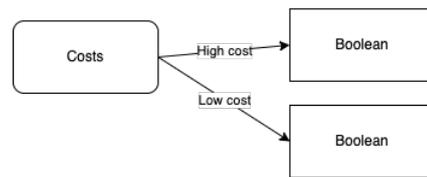


Abbildung 3.6: Costs Template

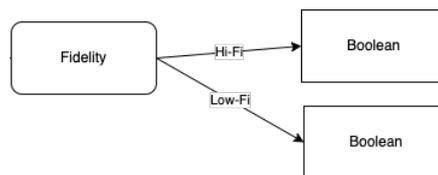


Abbildung 3.7: Fidelity Template

Video Production wurde mit *Low-Costs* und *High-Costs* Properties erweitert, die jeweils boolesche Werte annehmen (entweder True oder False sein können) und zeigen ob die Kosten für die Produktion eines Videos hoch oder niedrig sind.

Video Object wurde mit *High-Fidelity* und *Low-Fidelity* erweitert. Die Properties sind genauso wie im vorherigen Fall durch boolesche Werte repräsentiert und Zeigen damit ob die Fidelity eines Videos “High” oder “Low” ist.

3.5 Anwendung des Templates

Um den, in dieser Arbeit vorgestellten, Ansatz zu prüfen, wurden von dem Betreuer dieser Arbeit fünfzig Publikationen zum Thema “Production and Use of Videos in Requirements Engineering” ⁸ auf ORKG zur Verfügung gestellt. Die Liste der Publikationen befindet sich im Anhang [A](#). Jeder Artikel wurde gelesen, passende Artikel wurden mit dem Video Process Template, durch das Editig-Tool der ORKG, versorgt und die Properties des Templates anhand der Daten von jeweiliger Publikation ausgefüllt. In dem Fall, dass eine Property des Templates nicht in der Publikation vorkommt, wurde die Property nicht ausgefüllt.

⁸<https://orkg.org/list/R182413>

Kapitel 4

Datenanalyse

4.1 Anforderungen an die Analyse

Die Arbeit beinhaltet die Erstellung der Analyse der erfassten Daten mit Python (Jupyter Notebook) und einem Endpunkt *SPARQL Protocol And RDF Query Language* (SPARQL), auf Basis, der mit dem Template beschriebenen Publikationen. Diese Analysen basieren auf Kompetenzfragen zum beschriebenen wissenschaftlichen Wissen, die von den Forschern des Fachgebietes Requirements Engineering und vom Betreuer dieser Arbeit Herrn Dr. rer. nat. Karras erstellt wurden.

Anhand der erfassten Daten werden in Jupyter Notebook entsprechende Graphen erzeugt, um die Ergebnisse graphisch zu präsentieren. Die Graphen sollen dabei helfen die folgenden Fragen zu beantworten:

- Wie haben sich die Guidelines mit der Zeit geändert?
- Wie sehen die Eigenschaften des Kontextes (group, individual, virtual, physical, synchronous, asynchronous) von Videos aus? Welche Eigenschaft kommt am meisten vor?
- Welche Kombinationen von Eigenschaften des Kontextes der Videos (group, individual, virtual, physical, synchronous, asynchronous) treten am meisten auf?
- In Welchen Phasen der Requirements Engineering werden die Videos angewendet? Welche Phase der Requirements Engineering kommt am meisten vor?
- Welche Synonyme von Videos wurden in Artikeln benutzt? Welche Synonyme kommen am meisten vor?
- Gibt es eine Abhängigkeit zwischen dem Type von Video und seiner Dauer?

- Welche Interaktionsmöglichkeiten wurden in Publikationen erwähnt?
- Wie oft kommt das Video um ein “Problem” / “Solution” zu kommunizieren, in Phasen von Requirements Engineering vor?
- Von wem werden die Videos am meisten produziert?
- Welche Typen von Videos sind in der Regel mit hohen Kosten verbunden?
- Welche Typen von Videos haben am meisten Low-Fidelity und welche High-Fidelity?

4.2 Funktion des Jupyter Notebooks

Um auf die Fragen, die in [4.1](#) gestellt wurden, Antworten geben zu können, wurde ein Jupyter Notebook¹ mithilfe von JupyterLab erstellt. JupyterLab ermöglicht es Workflows im Bereich Data Science zu konfigurieren und zu arrangieren. Jupyter Notebook ist eine Webanwendung zum Erstellen und Freigeben von Computerdokumenten. Es bietet eine einfache, optimierte, dokumentenzentrierte Erfahrung. Es ermöglicht den Text sowie den Code und seine Ergebnisse gleich zu sehen, was “Look and Feel” eines ganzheitlichen Dokumentes anbietet, das bequem zu lesen ist.

Um die Daten aus den wissenschaftlichen Publikationen, die mit einem Template versorgt wurden, zu extrahieren, wurde auf SPARQL Schnittstelle von ORKG zugegriffen. SPARQL ist eine Graph-Matching-Abfragesprache, die aus drei Teilen besteht:

- *pattern matching part*
Pattern matching part enthält mehrere Merkmale des pattern matching von Graphen, wie zum Beispiel: optionale Teile, Vereinigung von Mustern, Verschachtelung, Filterung von Werten und die Möglichkeit, die Datenquelle auszuwählen, die mit einem Muster verglichen werden soll [\[18\]](#).
- *solution modifiers*
Sobald die Ausgabe des Musters berechnet wurde, können diese Werte mithilfe klassischer Operatoren wie Projektion, Distinct, Order und Limit geändert werden [\[18\]](#).
- *output*
Die Ausgabe einer SPARQL-Abfrage können sowohl einfache Ja/Nein-Abfragen als auch die Auswahl von Werten der Variablen, die mit den Mustern übereinstimmen sein [\[18\]](#).

¹<https://jupyter.org/>

²<https://github.com/AlexHahnGF/bachelor/blob/main/Untitled.ipynb>

Die oben genannten Teile von SPARQL sind für den Abruf der Daten sehr wichtig. Mit *pattern matching* werden Publikationen, die “Video Process” Templates beinhalten, abgefragt und mit *output* können die Werte der bestimmten Properties zurückgegeben werden. Um die Daten aus den Publikationen zu bekommen, wurden die SPARQL-Anfragen geschrieben. Die Ausgabe der SPARQL-Anfragen dient als Basis für weitere Datenanalyse.

Um die Graphen anhand der mit SPARQL-Anfragen erhaltenen Daten zu erzeugen, wurde auf Programmiersprache *Python*³ und ihren Bibliotheken *sparql-dataframe*⁴ und *matplotlib*⁵ zugegriffen. *Sparql-dataframe* erlaubt es Daten, die mit SPARQL-Anfragen erhalten wurden, in *pandas Dataframes*⁶ zu konvertieren. *Matplotlib* ist eine umfassende Bibliothek zum Erstellen statischer, animierter und interaktiver Visualisierungen in Python. *Matplotlib* ist in JupyterLab und dessen grafische Benutzeroberfläche eingebettet. Es erlaubt Plots in Publikationsqualität zu erzeugen und mit Pandas Dataframes zu arbeiten.

4.3 Ergebnis der Datenanalyse

In diesem Kapitel wird erklärt, in welcher Form die Ergebnisse der Analyse dargestellt werden, und welche Schlussfolgerungen aus den gewonnenen Daten gezogen werden können.

Um die Fragen aus Kapitel 4.1 zu beantworten wurden mehrere Diagramme erzeugt. Die Art von Graphen wurde anhand der Daten, die zu analysieren sind, ausgewählt. Somit hat sich folgendes ergeben:

- *Streudiagramm*
Ein Streudiagramm kann verschiedene Arten von Korrelationen zwischen Variablen mit einem bestimmten Konfidenzintervall vorschlagen. Aus diesem Grund wurde der Graph für die Analyse der folgenden Fragen angewendet: *Wie haben sich die Guidelines mit der Zeit geändert? Gibt es eine Abhängigkeit zwischen dem Typ vom Video und seiner Dauer?*
- *Säulendiagramm*
Mit Säulendiagrammen können die Häufigkeitsverteilungen in bestimmten Ausprägungen ermittelt werden. Außerdem ist das Säulendiagramm gut für Vergleiche von unabhängigen Messwerten geeignet [1], deswegen passt das Diagramm gut für z.B folgende Fragen: *Wie sehen die Eigenschaften des Kontextes von Videos aus? (group, individual, virtual, physical, synchronous, asynchronous) / Welche*

³<https://www.python.org/>

⁴<https://github.com/lawlesst/sparql-dataframe>

⁵<https://matplotlib.org/>

⁶<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

Eigenschaft kommt am meisten vor? Welche Synonyme von Videos wurden in Artikeln benutzt? Welche Synonyme kommen am meisten vor? Welche Interaktionsmöglichkeiten wurden in Publikationen erwähnt?

- **Kreisdiagramm**

Das Kreisdiagramm ist sehr gut für die Fälle geeignet, bei denen Anteile von Gesamtmengen dargestellt werden sollen [1]. Das ist z.B. bei der Frage *In Welchen Phasen des Requirements Engineering werden die Videos angewendet / Welche Phase des Requirements Engineering kommt am meisten vor?* der Fall.

Um eine Schlussfolgerung zu ziehen, werden hier auch die erzeugten Graphen präsentiert.

Abhängigkeit der Guidelines von der Zeit

Um die Kompetenzfrage “Wie haben sich die Guidelines mit der Zeit

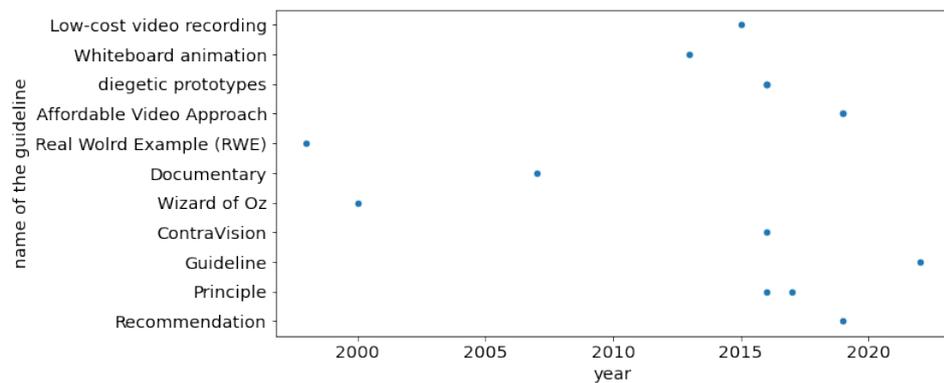


Abbildung 4.1: Abhängigkeit der Guidelines von der Zeit

geändert?” zu beantworten wurde der Graph *Abhängigkeit der Guidelines von der Zeit* [4.1] erzeugt. Aus der Abbildung [4.1] ergibt sich, dass die Anwendung der Guidelines mit der Zeit zunimmt und dass sich die Anzahl der unterschiedlichen Guidelines mit der Zeit vermehrt.

Eigenschaften des Kontextes von Videos

Das Diagramm [4.2] soll die Antwort auf die Kompetenzfrage “Wie sehen die Eigenschaften des Kontextes von Videos aus? / Welche Eigenschaft kommt am meisten vor?” liefern. Mit dem entsprechenden Säulendiagramm [4.2] wird klar, dass zu dem Zeitpunkt, als diese Arbeit geschrieben wurde, die *synchrone* Eigenschaft des Video Kontextes am meisten vorkommt, gefolgt von *physical* und *group*. Es kann davon ausgegangen werden, dass die Videos am meisten “synchron”, nicht “virtual” sondern “physical” - in einem echten Platz, und in einer “Gruppe” angewendet werden. Das kann so interpretiert werden, dass mehrere Teilnehmer z.B. in einem Raum zusammen sitzen und

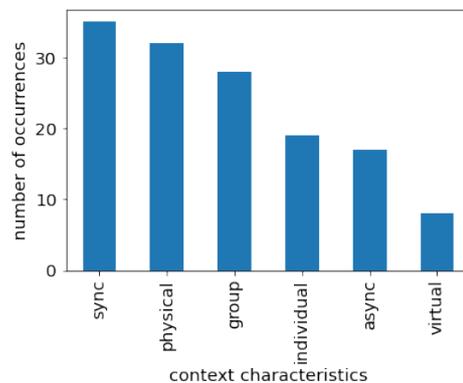


Abbildung 4.2: Eigenschaften des Kontextes von Videos

sich das Video zusammen anschauen. Weiter wird auf die Kombinationen der Eigenschaften eingegangen.

Kombination der Eigenschaften des Kontexts von Videos

Das Diagramm [4.3](#) wurde mit dem Ziel erstellt, die Kompetenzfrage “Welche Kombinationen von Eigenschaften des Kontextes der Videos treten am meisten auf?” zu beantworten. Auf dem Säulendiagramm [4.3](#) sieht man, dass die Kombination der *group* und *sync* Eigenschaften am meisten vorkommt, welches auch das Diagramm [4.2](#) widerspiegelt. Dies kann so interpretiert werden, dass die Videos am meisten synchron und innerhalb einer Gruppe angewendet werden.

Phasen der Requirements Engineering

Um die Kompetenzfrage “In welche Phasen der Requirements Engineering werden die Videos angewendet / Welche Phase der Requirements Engineering kommt am meisten vor?” zu beantworten, wurde das Kreisdiagramm *Phasen der Requirements Engineering* [4.4](#) erstellt. Aus den *Phasen der Requirements Engineering* Kreisdiagramm [4.4](#) kann folgende Schlussfolgerung gezogen werden: Die Videos werden am meisten während der *Elicitaion* angewendet. Außerdem gab es bisher kein Video, das während der *Interpretation* angewendet wurde, welches auch sehr interessant in Bezug auf neue Forschungsgebiete sein könnte. Ein weiterer Punkt ist, dass *Vision Videos* zur Diskussion und zur Klärung dienen soll, was wiederum ein Beitrag zur Negotiation und Interpretation leisten darf. Aus dem Graph folgt aber, dass die Nutzung von *Vision Videos* und Videos generell, in diesen Phasen bisher kaum berichtet bzw. untersucht wurde, obwohl die *Vision Videos* am häufigsten in den Publikationen vorkommen, was hier auch später gezeigt wird.

Synonyme von Video

Das Säulendiagramm *Synonyme von Video* [4.5](#), hat das Ziel auf die Kompetenzfrage “Welche Synonyme von Videos wurden in Artikel benutzt? Welche Synonyme kommen am meisten vor?” Antwort zu geben. Laut dem

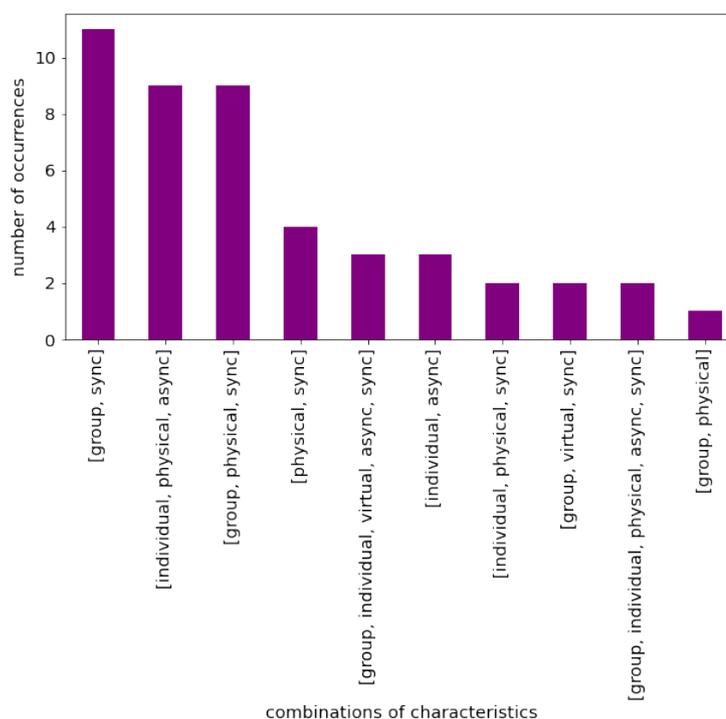


Abbildung 4.3: Kombination der Eigenschaften des Kontextes von Videos

Synonyme von Video Säulendiagramm [4.5](#), kommen “Vision videos” am häufigsten vor, jedoch gibt es immer noch genügend andere Typen bzw. Synonyme von Videos.

Abhängigkeit zwischen den Content von Video und seiner Dauer

Auf dem Streudiagramm *Content und Dauer von Videos* [4.6](#) sieht man, dass sich zwei Gruppen in Bezug auf Zeit bilden. Eine Gruppe beinhaltet alle Videos, bzw. Context der Videos, die kürzer als 3,63 Minuten sind und die andere Gruppe hat Videos (bzw. Context der Videos) die mehr als 10 Minuten dauern. Es bildet sich eine Lücke zwischen den zwei Gruppen. Laut dem Diagramm sind *Workshop Videos* - die längsten, jedoch gibt es auch *Meetings Videos*, die sehr lang sein können (10 Minuten). Interessant an dieser Stelle ist, dass *Vision Videos* fast jede Länge haben können und der Fakt, dass *Scenario Video* sowohl sehr kurz (1.07 Minuten) als auch sehr lang (10 Minuten) sein können. Somit ist die Kompetenzfrage “Gibt es eine Abhängigkeit zwischen dem Typ von Videos und seiner Dauer?” beantwortet.

Interaktionsmöglichkeiten in Videos

Anhand der erhaltenen Daten und *Interaktionsmöglichkeiten in Videos* Diagramm [4.7](#) kann folgendes beschlossen werden: die Interaktion mit Videos kommt noch nicht häufig vor, jedoch gibt es mindestens drei Publikationen

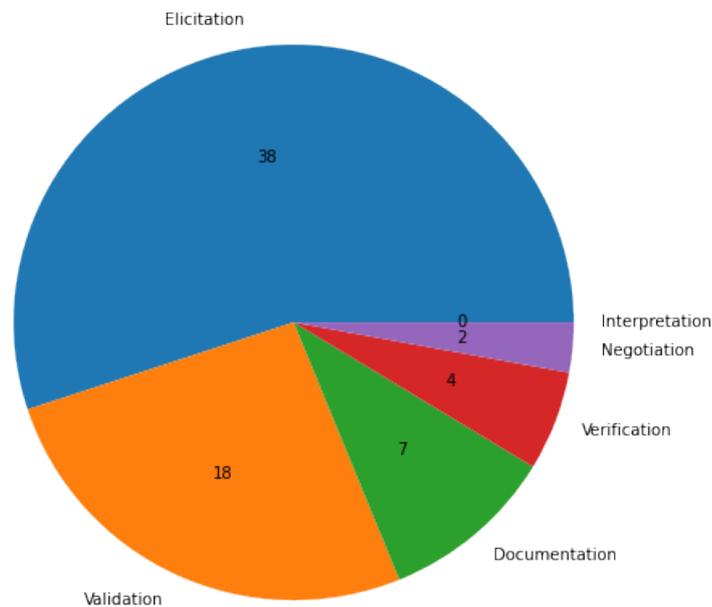


Abbildung 4.4: Phasen der Requirements Engineering

die darüber berichten und zwei unterschiedliche Interaktionsansätze, *Poll* und *Tagging*. Dadurch, dass es nicht genügend Daten gibt, kann die Kompetenzfrage “Welche Interaktionsmöglichkeiten wurden in Publikationen erwähnt?” nur teilweise beantwortet werden.

Anzahl der Nutzungen von Videos um ein Problem/Solution zu kommunizieren in Phasen der Requirements Engineering

Das Säulendiagramm *Anzahl der Nutzungen von Videos um ein Problem/Solution zu kommunizieren in Phasen der Requirements Engineering* 4.8 sollte eine Antwort auf die Kompetenzfrage “Wie oft kommt das Video um ein “Problem” / “Solution” zu kommunizieren, in Phasen von requirements Engineering vor?” geben. Aus *Anzahl der Nutzungen von Videos um ein Problem/Solution zu kommunizieren wird in Phasen der Requirements Engineering* Säulendiagramm 4.8 deutlich das Videos, die ein “Problem” oder “Solution” kommunizieren, am meisten in der Elicitation Phase des Requirements Engineering vorkommen. Das “Problem” und die “Solution” sind in dieser Phase relativ gleich verteilt. Aus dem Diagramm folgt, das Videos, die ein “Problem” kommunizieren, deutlich seltener in der Dokumentationphase vorkommen. Dies stellt eine mögliche Forschungslücke dar. Außerdem gibt es auch mehrere Fälle, bei dem ein Video gleichzeitig ein “Problem” als auch eine “Solution” anspricht.

Hersteller von Videos

Um die Kompetenzfrage “Von wem werden die Videos am meisten produziert?” zu beantworten, wurde das Säulendiagramm *Hersteller von Videos*

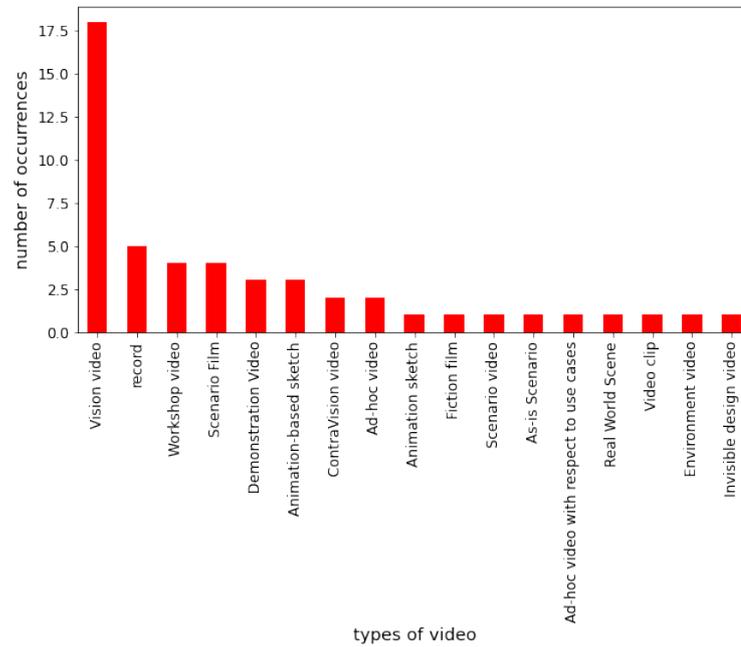


Abbildung 4.5: Synonyme von Video

4.9 erstellt. Laut dem Diagramm werden die Videos überwiegend vom *Requirements Engineer* erstellt. Dadurch, dass am häufigsten Vision Videos in den Publikationen vorkommen, war das Ergebnis zu erwarten und unterstützt somit das *Synonyme von Videos* 4.5 Diagramm.

“High-Cost Videos”

Das Säulendiagramm 4.10 sollte zeigen, ob es bestimmte Typen von Videos gibt, deren Produktion in der Regel mit hohen Kosten verbunden ist und somit die Kompetenzfrage “Welche Typen von Videos sind in der Regel mit hohen Kosten verbunden?? ” zu beantworten. Es wurde das Diagramm 4.10 erstellt mit dem Ziel, Abhängigkeiten des Typs des Videos

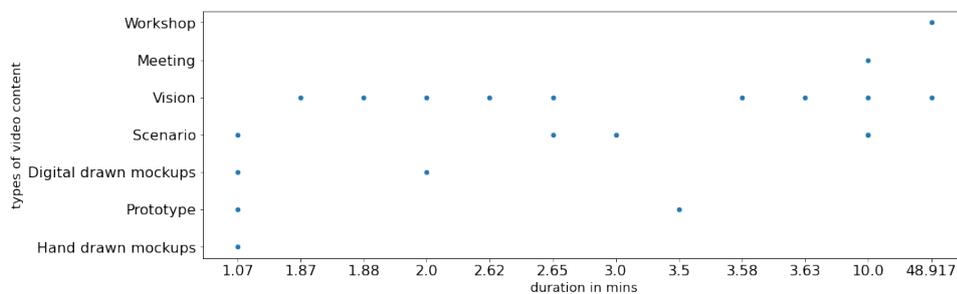


Abbildung 4.6: Content und Dauer von Videos

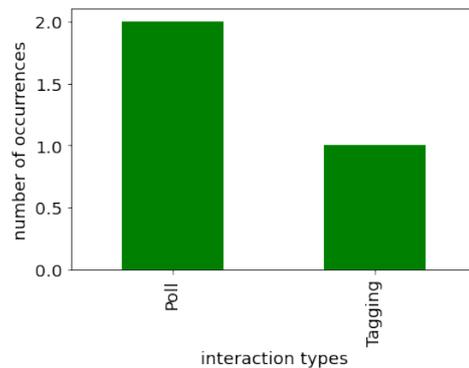


Abbildung 4.7: Interaktionsmöglichkeiten in Videos

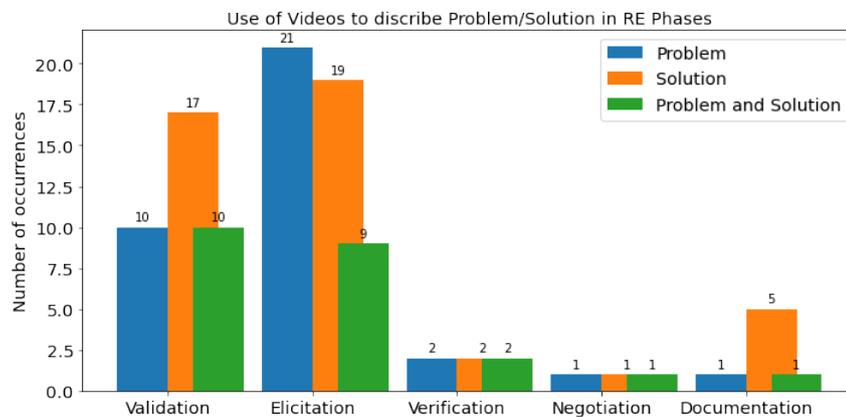


Abbildung 4.8: Anzahl der Nutzungen von Videos um ein Problem/Solution zu kommunizieren in Phasen der Requirements Engineering

von seinen Produktionskosten, zu ermitteln. Aus dem Diagramm folgt, dass die *Vision Videos* am häufigsten hohe Produktionskosten haben, jedoch reichen die Daten nicht aus um dies mit Sicherheit zu bestätigen. Dadurch das *Vision Videos* am meisten und mit großem Abstand in wissenschaftlichen Publikationen vorkommen, kommt es zu unzureichenden Daten, um eine Aussage bezüglich der Produktionskosten von anderen Videotypen zu treffen. Somit kann die Kompetenzfrage nicht ganz beantwortet werden.

“High-Fidelity and Low-Fidelity Videos”

Um Typen von Videos, die am meisten “Low”, bzw “High Fidelity” haben, zu ermitteln, wurde das Diagramm [4.11](#) erstellt. Somit sollte die Kompetenzfrage “Welche Typen von Videos haben am meisten Low-Fidelity und welche High-Fidelity?” beantwortet werden. Aufgrund fehlender Datenmenge, ist es schwer eine Schlussfolgerung zu treffen. *Animation-based-sketch* kommt am meisten als “Low-Fidelity” Video vor, jedoch kann, laut dem Diagramm,

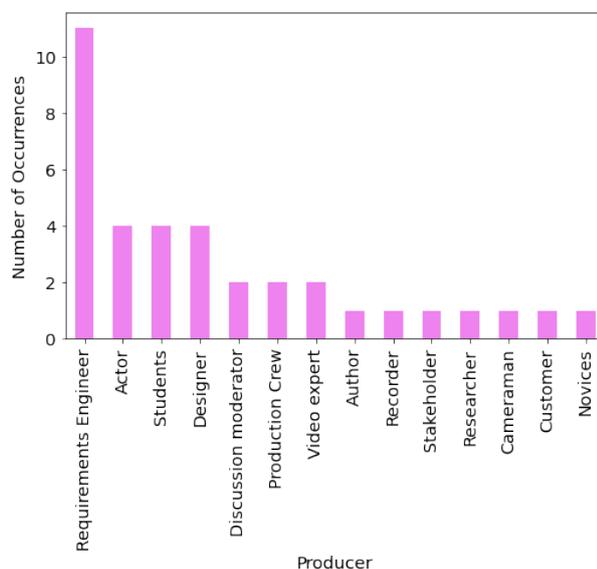


Abbildung 4.9: Hersteller von Videos

das Video auch als “High-fidelity” dargestellt werden. Es ist aber deutlich bemerkbar, dass es zurzeit mehr “Low-Fidelity” Videos im Requirements Engineering gibt, was ebenfalls auf eine mögliche Forschungslücke hinweist. Abschließend kann gesagt werden, dass es in diesem Fall nicht genügend Daten gibt, um mit Sicherheit die Kompetenzfrage “Welche Typen von Videos haben am meisten Low-Fidelity und welche High-Fidelity?” zu beantworten.

4.4 Fazit der Datenanalyse

Anhand der durchgeführten Analyse können 9 aus 11 Fragen beantwortet werden, jedoch zeigten sich große Unterschiede in der Anzahl der Daten. Zum Beispiel “Guidelines” kommen nur 12 Mal in den, im Rahmen dieser Arbeit, betrachteten Publikationen vor. Dafür wurden Phasen des Requirements Engineering 69 Mal erwähnt. Bei den Interaktionsmöglichkeiten zeigte sich eine Lücke, was nicht unbedingt schlecht sein muss, sondern als ein potenziell wenig erforschtes Forschungsfeld betrachtet werden kann. Dadurch öffnet sich die Möglichkeit für neue Forschungen. Es hat sich ergeben, dass die Videos am meisten in *Elicitation* Phase des Requirements Engineering, synchron und in einer Gruppe angewendet werden. Interessant ist, dass obwohl Vision Videos zur Diskussion und zur Klärung dienen sollten, welches während den Negotiation und Interpretation Phasen des Requirements Engineering passieren sollte, kaum berichtet bzw. untersucht wurde. Es wurden mehrere Synonyme/Typen von Videos entdeckt, welche den Forschern helfen sollte,

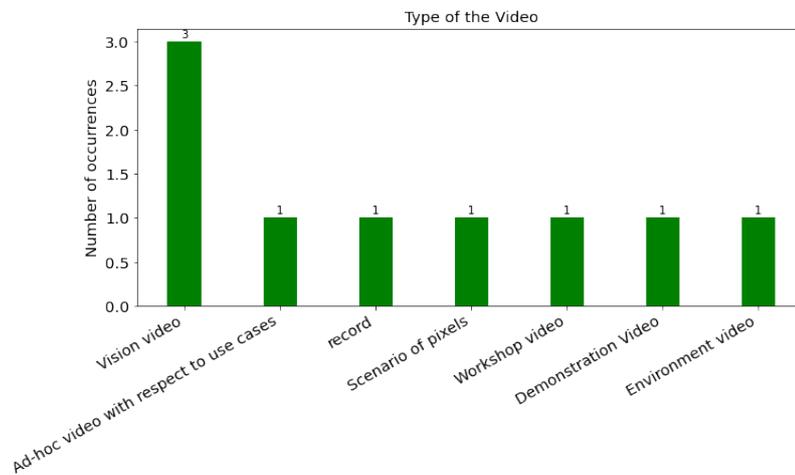


Abbildung 4.10: High-Cost Videos

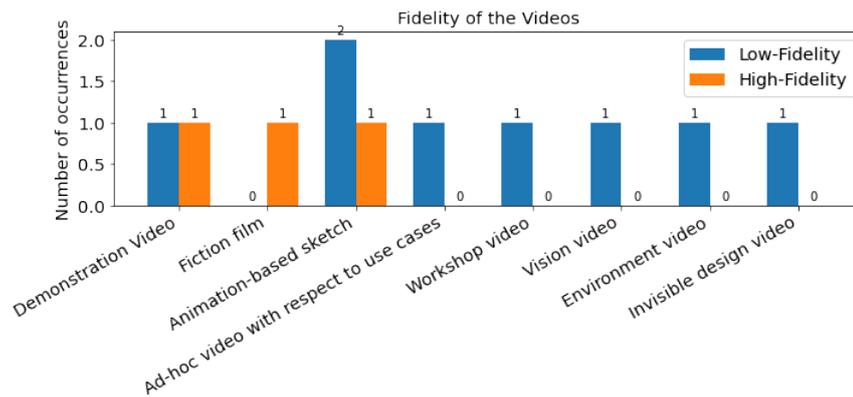


Abbildung 4.11: High-Fidelity and Low-Fidelity Videos

den “Suchstring” entsprechend zu erstellen. Es zeigte sich, dass die Länge von Videos zwei Gruppen bildet, eine mit der Länge kleiner als 4 Minuten und die zweite mit einer Länge größer gleich 10 Minuten. Es wurden keine Videos mit der Länge von 4 bis 9 Minuten gefunden. Außerdem wurden einige mögliche Forschungslücken entdeckt, z.B dass gar keine Videos in der *Interpretation* Phase vom Requirements Engineering angewendet wurden. Weiterhin wurde entdeckt, dass Videos die ein “Problem” kommunizieren, deutlich seltener in der Dokumentation Phase vorkommen, welche eine weitere mögliche Forschungslücke darstellt. Außerdem gibt es zurzeit deutlich mehr “Low-Fidelity” als “High-Fidelity” Videos in Reuirements Engineering. Diese stellt ebenfalls einen möglichen Forschungsbereich dar.

Kapitel 5

Auswertung

Bei jeder Art von Forschung existieren sogenannte “Threats to validity” [24]. Bevor das Template sowie die durchgeführte Datenanalyse ausgewertet werden, werden in diesem Kapitel erst die möglichen Gefahren für diese Arbeit bzw. Forschung angedeutet.

5.1 Threats to Validity

Um das Template und die durchgeführte Analyse der Daten auswerten zu können, ist es wichtig die möglichen Gefahren für die Arbeit zu erwähnen. Es ist möglich, dass Forschung ohne den Gefährdungen nicht durchgeführt werden kann und daher müssen diese akzeptiert und bei der Interpretation der Ergebnisse berücksichtigt werden [24]. Um mögliche Gefahren zu erkennen wurde “Checklist of Threats” von Wohlin et al. [24] als Basis genommen.

Nach Wohlin können die “Threats to validity” in vier Gruppen aufgeteilt werden: *Conclusion Validity*, *Internal validity*, *External validity* und *Construct validity*. Die jeweiligen Bedrohungen von jeder Gruppe werden hier weiter näher betrachtet.

Conclusion Validity.

Conclusion Validity Threats beziehen sich auf Probleme, die die Fähigkeit beeinträchtigen, die richtigen Schlussfolgerungen über die Beziehungen zwischen der Behandlung und dem Ergebnis eines Experiments zu ziehen [24]. Um die, im Rahmen dieser Arbeit, vorgestellte Methode für die nachhaltige und langfristige Akquirierung, Kuration, Veröffentlichung und Verarbeitung von wissenschaftlichem Wissen mit dem ORKG zu ermöglichen, wurden die wissenschaftlichen Publikationen nur von einer Person gelesen und die gefundenen Daten manuell in das Template übertragen. Dies kann dazu führen, dass die Daten falsch interpretiert wurden und in weitere Analysen übernommen wurden. Dies wiederum kann die Analyse der Daten negativ beeinflussen. *Random irrelevancies in experimental setting* stellt eine Gefahr dar, da alle Publikationen sorgfältig durchgelesen werden sollten um die

Daten in den Publikationen richtig zu interpretieren und in das Template übertragen zu können. Dabei könnte zum Beispiel Lärm oder andere Faktoren von dem Lesen oder Übertragen der Daten ablenken, was dazu führen könnte, dass die Daten falsch interpretiert/übertragen oder einfach überlesen werden. Um diese Risiken gering zu halten, wurde versucht alle Teile der Arbeit unter gleichen Bedingungen durchzuführen um alle Einflüsse von Außen zu vermeiden.

Internal Validity

Internal Validity sind Einflüsse, die ohne Wissen des Forschers unabhängige Variablen kausal beeinflussen können [24]. Die Auswahl der Publikationen ist wahrscheinlich die größte Bedrohung für die *Internal Validity*, da jede Voreingenommenheit bei der Auswahl die Genauigkeit und Qualität der Endergebnisse beeinflusst [10]. Um diese Gefahr zu senken, wurden nur wissenschaftliche Publikationen zum Thema “Production and Use of Videos Requirements Engineering” betrachtet, die vom Betreuer dieser Arbeit und von Mitarbeitern des Software Engineering Fachgebietes der Leibniz Universität Hannover sorgfältig ausgewählt und zur Verfügung gestellt wurden. Aus diesem Grund, kann die ausgewählte Literatur als geeignet bezeichnet werden. Allerdings ist die Anzahl der analysierten Publikationen relativ gering und schränkt somit die Aussagekraft der Analyse ein.

Construct Validity

Construct Validity betrifft die Verallgemeinerung des Ergebnisses des Experiments auf das Konzept oder die Theorie hinter dem Experiment [24]. Die Methode für die nachhaltige und langfristige Akquirierung, Kuration, Veröffentlichung und Verarbeitung von wissenschaftlichem Wissen mit dem ORKG ist relativ neu, deswegen gibt es zurzeit keine Guidelines für die Anwendung dieser Methode. Mehr Erfahrung in diesem Bereich könnte helfen die Daten aus den Publikationen einfacher zu extrahieren, sowie den Ausbau des Templates zu verbessern indem neue Properties hinzugefügt werden, die für weitere Analysen eine große Rolle spielen könnten. Die Verwendung einer einzigen Methode führt zu einem Mono-Methoden-Bias, welches sich weiter in der Analyse der Daten widerspiegeln kann. Jedoch ist das Ziel an dieser Stelle nicht die möglichst präzise Analyse und Darstellung der Daten anzubieten, sondern eine alternative Methode zu bereits bekannter semantischer Literatursuche und semantischer Mapping-Studie zu präsentieren. Der Grund dafür ist, die unzureichend nachhaltige Aufbereitung und langfristige Zurverfügungstellung der Daten mit den vorher erwähnten Methoden. Eine Erweiterung der bestehenden Untersuchung ist somit kaum möglich und erfordert stattdessen meist eine Wiederholung des gesamten Extraktions- und Analyseprozesses.

External Validity

External Validity sind Bedingungen, die die Fähigkeit einschränken, die Ergebnisse von Experimenten auf die industrielle Praxis zu übertragen [24].

Die Methode, die in dieser Arbeit vorgestellt wurde, ist relativ neu, weshalb nicht garantiert werden kann, dass die Methode für alle Fälle und alle Forschungsgebiete gleich gut angewendet werden kann. Allerdings wurden in dieser Arbeit auch Vor- und Nachteile der Methode erläutert, was als eine Basis für weitere Arbeiten dienen kann.

5.2 Auswertung des Templates

In den weiteren Unterkapiteln [5.2.1](#) und [5.2.2](#) wird der, in dieser Arbeit vorgestellte Template-Ansatz, im Bezug auf nachhaltige Akquirierung und Verarbeitung der Daten und Zeitaufwands diskutiert.

5.2.1 Zeitaufwand

Der Template-Ansatz ermöglicht es die wissenschaftlichen Daten langfristig mit dem ORKG zu akquirieren, Kurieren, veröffentlichen und zu verarbeiten. Das Erstellen des Templates erfordert gute Kenntnisse im Bereich, in denen das Template angewendet wird. Außerdem soll die Person, die das Template erstellt, mit ORKG vertraut gemacht werden. Falls die oben genannten Punkte fehlen, sollte mehr Zeit in Anspruch genommen werden. Des Weiteren, müssen alle Daten manuell in das Template eingetragen werden. Das bedeutet, dass alle relevanten Publikationen erst gelesen werden müssen und danach die benötigten Daten aus den Artikeln in das Template übertragen werden. Je größer die Anzahl von relevanten Artikeln, desto mehr Zeit wird benötigt. Das Übertragen der Daten passiert schnell, da die jeweiligen Properties mit entsprechenden Daten befüllt werden. Die ganzen Schritte sind relativ aufwändig und hängen von der Anzahl der Publikationen ab. Der Zeitaufwand, der in dieser Arbeit vorgestellten Methode, ist vergleichbar mit dem Aufwand von systematischer Literaturrecherche [2.1.1](#) (Abbildung [2.1](#)) und systematischer Mapping-Studie [2.1.2](#) (Abbildung [2.2](#)). Bei den beiden Methoden müssen viele zeitaufwändige Schritte (z.B. Extrahieren der Daten und Mapping Prozess bei systematischer Mapping-Studie und Datensynthesen bei systematischer Literaturrecherche) durchgeführt werden. Beim Template-Ansatz werden diese Schritte auf einer anderen Art umgesetzt. Der Zeitaufwand im Vergleich zu den oben genannten Methoden bei gleichem Ziel ist dabei nicht nennenswert unterschiedlich. Abschließend kann gesagt werden, dass der Template-Ansatz kein Gewinn in Bezug auf den Zeitaufwand darstellt, oder der Gewinn ist so niedrig, dass es außer Acht gelassen werden kann.

5.2.2 Nachhaltigkeit

Im Gegenteil zu systematischer Literaturrecherche und systematischer Mapping-Studie bietet das Template nachhaltige und langfristige

Aufbereitung der Daten. Es wird nicht nur eine Tabelle oder eine einzelne PDF Datei erzeugt, sondern eine ganze semantische Struktur der Publikation, was wiederum erlaubt die Daten in maschinenlesbarer Form darzustellen [6]. Dadurch, dass die Daten maschinenlesbar werden, wird eine nachhaltige und langfristige Akquirierung, Kuration und Verarbeitung der Daten ermöglicht. Zudem sind die Daten immer online und können, mittels Schnittstellen wie z.B. SPARQL und Hilfe von Maschinen (z.B. PC), abgearbeitet und dargestellt werden. Das ist besonders dann vorteilhaft, wenn eine große Menge von Daten verarbeitet werden soll.

5.3 Auswertung der Datenanalyse

In den weiteren Unterkapiteln 5.3.1 und 5.3.2 wird das, im Rahmen dieser Arbeit erstellte Notebook und mit dessen Hilfe durchgeführter Analyse, in Bezug auf nachhaltige Akquirierung und Verarbeitung der Daten und des Zeitaufwands ausgewertet.

5.3.1 Zeitaufwand

Dadurch, dass die Daten mit dem Template maschinenlesbar gemacht werden, kann auf die Daten mit Hilfe von unterschiedlichen Schnittstellen zugegriffen werden. In dieser Arbeit wurde dafür eine SPARQL Schnittstelle des ORKG benutzt. Die Daten werden mit Anfragen im SPARQL Format empfangen und mit Python und *matplotlib* graphisch dargestellt. Der Aufwand für den Aufbau des Notebooks ist relativ gering und hängt von der Größe und Komplexität der SPARQL Abfragen ab. Nachdem das Jupyter Notebook erstellt ist, kann das Notebook innerhalb von wenigen Sekunden durchgespielt werden um die entsprechende Graphen zu generieren. Die Analyse der erzeugten Graphen muss aber erst manuell stattfinden, da der Forscher selbst erst eine Schlussfolgerung aus den erhaltenen Graphen ziehen muss. Dadurch, dass die Daten in Form von Graphen präsentiert werden, kann das Fazit in wenigen Minuten gemacht werden. Diese bedingt einen geringen Zeitaufwand. Im Vergleich zu bereits bekannten Methoden ist der Zeitaufwand hier viel geringer, da alle Daten in wenigen Sekunden graphisch dargestellt werden. Dies vereinfacht die letztendliche Analyse, da nur Schlussfolgerungen gezogen werden müssen.

5.3.2 Nachhaltigkeit

Im vorherigen Kapitel wurde bereits verdeutlicht, dass mit Hilfe von Templates die Daten in maschinenlesbarer Form dargestellt werden. Dies ist genau die Grundlage für eine nachhaltige Analyse der Daten, weil die Rohdaten, auf der die Analyse basiert, immer für jeden online zugänglich sind. Die Daten werden mit SPARQL Anfragen empfangen und anhand

der erhaltenen Daten werden die Graphen erzeugt. Das bedeutet, dass die Graphen zu jedem Zeitpunkt ohne Aufwand neu generiert werden können. Selbst wenn neue Publikationen hinzugefügt werden, werden die bisherigen Daten aus den Graphen zu finden sein und berücksichtigt. Außerdem kann das Notebook immer erweitert werden, falls neue Graphen oder Daten generell für jeweilige Forscher ein Interesse darstellen. Im Gegenteil dazu, werden die Daten von der systematischen Literaturrecherche und systematischen Mapping-Studie meist nicht nachhaltig aufbewahrt¹[2]. Wenn neue Artikel veröffentlicht werden, die für das jeweilige Forschungsgebiet relevant sein können und nach einigen Jahren eine neue Literaturrecherche oder Mapping-Studie durchgeführt wird, dann ist die einzige Option, die gesamte Studie zu wiederholen, falls die Rohdaten berücksichtigt werden sollen.

¹D. Hausen, NFDI4Ing Special Interest Group (SIG) Qualitätssicherung und Metriken für FAIR-Daten 4. Treffen, Online-Umfrage zu ingenieurwissenschaftlichen Aspekten in Datenmanagementplänen, 2022

Kapitel 6

Verwandte Arbeiten

Der Einsatz des Knowledge Graphs für die nachhaltige und wiederverwendbare Auffindung und Aufbewahrung der Forschungsdaten ist bereits bekannt in der Forschung. Es gibt bereits einige Arbeiten, die dieses Thema ebenfalls untersucht haben. Dieses Kapitel befasst sich mit einer Auswahl verwandter Arbeiten, die hohe Relevanz in Bezug auf diese Arbeit vorweisen, weil sie ähnliche Ziele und Problematiken verfolgen. Die der Arbeit zugrundeliegenden Grundkonzepte sowie Ziele werden in diesem Kapitel erläutert, um Unterschiede zwischen dieser Arbeit und den bereits vorhandenen Publikationen hervorzuheben.

Die hier präsentierten verwandten Arbeiten wurden, durch die Nutzung verschiedener Quellen, gefunden. Eine von diesen Quellen beinhaltet eine Menge von Publikationen auf ORKG zum Thema “Production and Use of Videos Requirements Engineering”. Eine weitere Quelle besteht aus Publikationen, die sich mit dem Thema ORKG oder “Knowledge graph” befassen. Beide Quellen wurden vom Betreuer dieser Arbeit als Einführung in die Thematik bereitgestellt. Eine weitere Quelle verwandter Arbeiten sind die Ergebnisse der Suchmaschine Google Scholar. Hierfür wurde nach diversen Kombinationen der Begriffe “knowledge graph”, “ORKG”, und “nachhaltige Aufbewahrung der Daten” gesucht.

Außer ORKG gibt es auch andere Knowledge-Graph-Projekte, wie z.B. Microsoft Academic Graph [23], Springer Nature SciGraph [8], Papers with Code¹, Semantic Scholar Literature Graph², OpenAIRE Research Graph³ und ResearchGraph⁴.

All diese RKGs gehören zu den generic RKGs, weil diese den Focus auf bibliographische Metadaten von wissenschaftlichen Artefakten und Entitäten haben. Sie haben die Gemeinsamkeit bibliographische Metadaten zu verwenden um wissenschaftliche Artefakte, Entitäten und Ihre Beziehungen

¹<https://paperswithcode.com>

²<https://www.semanticscholar.org>

³<https://graph.openaire.eu>

⁴<https://researchgraph.org/>

zu organisieren damit z.B. ihre Suche, Visualisierung und Verarbeitung ermöglicht werden [9].

Im Gegensatz zu allen bisher genannten RKGs ist ORKG domänen- und themenübergreifend. Auf ORKG werden wissenschaftliche Publikationen jeder Art und zu jedem Thema publiziert. Ein weiterer Vorteil von ORKG besteht darin, dass Papers mit eingebauter Funktion verglichen werden können.

Stocker et al [22] haben bereits Scholarly Knowledge Graphs for European Open Science Cloud (SKG4EOSC) in ihrer Arbeit beschrieben um zu zeigen, wie die wissenschaftliche Produktivität durch die Entwicklung und Implementierung von Diensten für die Text- und Datenkonvertierung sowie die Produktion, Kartierung und Wiederverwendung von FAIR-wissenschaftlichen Informationen verbessert werden können.

In seiner Arbeit haben Auer et al. [2] Funktionen der ORKG benutzt um wissenschaftliche Beiträge in strukturierter und semantischer Form als Wissensgraph zu repräsentieren. Sie haben Hauptfunktionen der ORKG im Bezug auf nachhaltige Aufbewahrung der Daten erläutert, wie z.B. strukturierte Beschreibung der Forschungsbeiträge, Templates, Graph Visualization usw. Des Weiteren wurden Use cases der Knowledge Graph, nämlich ORKG, in solchen Forschungsgebieten wie Computer Science, Epidemiology and Covid-19 und Materials Science beschrieben.

Gemeinsamkeiten der oben genannten Arbeiten sind, dass sie sich mit dem Thema nachhaltige Aufbewahrung von wissenschaftlichen Daten auseinandersetzen. Sie zeigen Orte auf, an denen aktuell Schwachstellen bei der Aufbewahrung der Daten sind und wie dieses Problem gelöst werden kann. Es wird erwähnt, dass die Forschungsdaten meist im PDF Format gespeichert werden. Dies hat zur Folge, dass die Daten für Maschinen nicht lesbar sind und damit die automatisierte Datenextraktion erschwert oder gar nicht möglich macht.

Der grundlegende Unterschied in dieser Arbeit zu den oben genannten Arbeiten ist, dass automatisierte Abliefern langfristiger und wiederverwendbarer graphischer Darstellung von Daten. Das Ziel dabei ist das einfachere Gestalten der Analyse von Ergebnissen eines bestimmten Forschungsgebiets. Außerdem ist der Scope der Arbeit von Stocker et al. [22] viel breiter. Hierdurch werden nur wissenschaftliche Publikationen zum Thema “Production and Use of Videos Requirements Engineering” betrachtet.

Im Vergleich zu dieser Arbeit bietet die Arbeit von Auer et al [2] einen umfassenden Überblick über ORKG und deren Funktionen. Weiterhin dient die Arbeit von Auer et al als Review der Knowledge Graphs im Allgemeinen und geht dabei besonders auf ORKG ein. Mit dieser Arbeit hingegen wird zusätzlich die Wiederverwendbarkeit der Analysen von Daten mit Hilfe von Funktionen bzw. Services der ORKG erzielt.

Kapitel 7

Fazit und Ausblick

Dieses Kapitel dient dazu, einen abschließenden Überblick über die Inhalte und die wichtigsten Ergebnisse der Arbeit wiederzugeben.

7.1 Fazit

Das Hauptziel dieser Arbeit wurde erreicht, indem die Nebenziele 1.1 und 1.2 abgedeckt wurden.

Um das Nebenziel 1.1 zu erreichen, wurden in dieser Arbeit fünfzig Publikationen zum Thema „Production and Use of Videos Requirements Engineering“ gelesen und ein bestehendes Template ausgebaut. Das Template wurde an einer im ORKG kuratierten Liste von Publikationen zum Thema “Production and Use of Videos Requirements Engineering” angewendet. Das Ziel dabei ist wissenschaftliches Wissen über die Anwendung von Videos im Requirements Engineering aus wissenschaftlichen Publikationen zu akquirieren und um die nachhaltig und langfristig mit dem ORKG zu kuratieren, veröffentlichen und verarbeiten. Mit dem Ausbau des vorhandenen Templates wurden zwei neue Properties *Production costs* und *Video Fidelity* dem Template hinzugefügt. Somit wurde ein Gesamtbild von einem Video Prozess, welcher in einem Artikel stattgefunden hat, erschaffen. Die Daten, die mit Hilfe vom Template eingetragen werden, stellen die Nachhaltigkeit dar, weil sie in einer festen Struktur langfristig bereitgestellt werden.

Das Nebenziel 1.2 wurde erreicht indem ein Jupyter Notebook erstellt wurde, in dem die erfassten Daten im Form von Graphen dargestellt wurden um Beziehungen zwischen den Properties des Templates (Daten der wissenschaftlichen Publikationen) für weitere Analyse rauszufinden. Dadurch, dass das Notebook immer wieder neu ausgeführt und erweitert werden kann, wird auch Wiederverwendbarkeit erreicht. Die mit dem Notebook entwickelten Daten wurden analysiert und daraus Schlussfolgerungen gezogen.

7.2 Ausblick

In dieser Arbeit wurden nur wissenschaftliche Publikationen zum Thema “Production and Use of Videos Requirements Engineering” als Eingangsdaten untersucht. Für die Überprüfung der Signifikanz der Ergebnisse wäre es empfehlenswert den Ansatz an weiteren Publikationen mit anderen Themen anzuwenden, um die Datenbasis zu erweitern und so umfassendere Ergebnisse zu erhalten.

Es wurde geprüft ob ORKG eine mögliche Lösung für nachhaltige und langfristige Aufbereitung der Daten bietet. Zudem wurde geprüft ob ORKG dem Problem der wiederholten und oft redundanter Extraktion und Analyse entgegenwirken kann. Dadurch, dass ORKG solche Features wie Templates hat und eine SPARQL Schnittstelle anbietet, wurde dies erfolgreich geprüft und bestätigt.

Mit der durchgeführten Analyse der Daten wurden zwei offene Forschungslücken entdeckt. In den fünfzig zur Verfügung gestellten Publikationen wurden keine Videos in der Requirements Engineering Phase “Interpretation” angewendet. Des Weiteren wurden nur sehr wenige Interaktionen mit Videos in den Publikationen erwähnt.

Ein nächstmöglicher Schritt wäre die Methode an anderen Forschungsgebieten anzuwenden, um zu prüfen, ob die Methode nicht nur für ein bestimmtes Forschungsgebiet geeignet ist, sondern für möglichst viele Forschungsgebiete gleich gute Ergebnisse liefert.

Ein weiterer Ansatz wäre eine Methode zu entwickeln um das Extrahieren und Übertragen der Daten in das Template zu vereinfachen bzw. vollständig zu Automatisieren, z.B. mit Hilfe von künstlicher Intelligenz und/oder maschinellem Lernen. Da dies zurzeit die zeitaufwändigsten Schritte sind, könnte damit sehr viel Zeit gespart werden. Außerdem kann so viel mehr Daten gesammelt werden, damit keine der gestellten Kompetenzfragen unbeantwortet bleiben.

Anhang A

Publikationen zum Thema “Production and Use of Videos Requirements Engineering”

Liste der Publikationen zum Thema “Production and Use of Videos Requirements Engineering”^[1], die als Basis für das Extrahieren der Daten für weitere Analyse und Anwendung des Templates dienten.

- [1] Pam Briggs, Mark Blythe, John Vines, Stephen Lindsay, Paul Dunphy, James Nicholson, David Green, Jim Kitson, Andrew Monk, and Patrick Olivier. Invisible design: Exploring insights and ideas through ambiguous film scenarios. In Proceedings of the Designing Interactive Systems Conference, DIS '12, page 534–543, New York, NY, USA, 2012. Association for Computing Machinery.
- [2] Bernd Bruegge, Oliver Creighton, Maximilian Reiß, and Harald Stangl. Applying a video-based requirements engineering technique to an airport scenario. In 2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering - Beyond Mere Descriptions and with More Fun and Games, pages 9–11, 2008.
- [3] Françoise Brun-Cottan and Patricia Wall. Using video to represent the user. *Commun. ACM*, 38(5):61–71, may 1995.
- [4] P.L. Brouse, N.A. Fields, and J.D. Palmer. A multimedia computer supported cooperative work environment for requirements engineering. In [Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics, pages 954–959 vol.2, 1992.
- [5] Miroslav Bojic, Areti Goulati, Dalila Szostak, and Panos Markopoulos. On the effect of visual refinement upon user feedback in the context

¹<https://orkg.org/list/R182413>

- of video prototyping. In Proceedings of the 29th ACM international conference on Design of communication, pages 115–118, 2011.
- [6] Gregor Broll, Heinrich Hussmann, Enrico Rukzio, and Raphael Wimmer. Using video clips to support requirements elicitation in focus groups-an experience report. In SE 2007 Workshop on Multimedia Requirements Engineering, 2007.
- [7] Melanie Busch, Oliver Karras, Kurt Schneider, and Maike Ahrens. Vision meets visualization: Are animated videos an alternative? In Nazim Madhavji, Liliana Pasquale, Alessio Ferrari, and Stefania Gnesi, editors, Requirements Engineering: Foundation for Software Quality, pages 277–292, Cham, 2020. Springer International Publishing.
- [8] Amel Bennaceur, Ciaran McCormick, Jesús García Galán, Charith Perera, Andrew Smith, Andrea Zisman, and Bashar Nuseibeh. Feed me, feed me: An exemplar for engineering adaptive software. In Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self- Managing Systems, SEAMS '16, page 89–95, New York, NY, USA, 2016. Association for Computing Machinery.
- [9] Olesia Brill, Kurt Schneider, and Eric Knauss. Videos vs. use cases: Can videos capture more requirements under time pressure? In International Working Conference on Requirements Engineering: Foundation for Software Quality, pages 30–44. Springer, 2010.
- [10] Melanie Busch, Jianwei Shi, Lukas Nagel, Johann Sell, and Kurt Schneider. Vision video making with novices: A research preview. In Vincenzo Gervasi and Andreas Vogelsang, editors, Requirements Engineering: Foundation for Software Quality, pages 251–258, Cham, 2022. Springer International Publishing.
- [11] Bernd Bruegge, Harald Stangl, and Maximilian Reiss. An experiment in teaching innovation in software engineering: Video presentation. In Companion to the 23rd ACM SIGPLAN Conference on Object-Oriented Programming Systems Languages and Applications, OOPSLA Companion '08, page 807–810, New York, NY, USA, 2008. Association for Computing Machinery.
- [12] Oliver Creighton, Martin Ott, and Bernd Bruegge. Software cinema-video-based requirements engineering. In 14th IEEE International Requirements Engineering Conference (RE'06), pages 109–118, 2006.
- [13] Andy Darby, Emmanuel Tseklevs, and Pete Sawyer. Speculative requirements: design fiction and re. In 2018 IEEE 26th International Requirements Engineering Conference (RE), pages 388–393. IEEE, 2018.

- [14] Samuel A Fricker, Kurt Schneider, Farnaz Fotrousi, and Christoph Thuemmler. Workshop videos for requirements communication. *Requirements Engineering*, 21(4):521–552, 2016.
- [15] Mark Gall and Brian Berenbach. Towards a framework for real time requirements elicitation. In 2006 First International Workshop on Multimedia Requirements Engineering (MERE’06 - RE’06 Workshop), pages 4–4, 2006.
- [16] P. Haumer, K. Pohl, and K. Weidenhaupt. Requirements elicitation and validation with real world scenes. *IEEE Transactions on Software Engineering*, 24(12):1036–1054, 1998.
- [17] M. Jirotko and P. Luff. Supporting requirements with video-based analysis. *IEEE Software*, 23(3):42–44, 2006.
- [18] Kurt Schneider Karras, Jil Klünder. Enrichment of Requirements Specifications with Videos - Enhancing the Comprehensibility of Textual Requirements. September 2016.
- [19] Oliver Karras. Software professionals’ attitudes towards video as a medium in requirements engineering. In Marco Kuhrmann, Kurt Schneider, Dietmar Pfahl, Sousuke Amasaki, Marcus Ciolkowski, Regina Hebig, Paolo Tell, Jil Klünder, and Steffen Küpper, editors, *Product-Focused Software Process Improvement*, pages 150–158, Cham, 2018. Springer International Publishing.
- [20] Oliver Karras, Ahmad Hamadeh, and Kurt Schneider. Enriching requirements specifications with videos-the use of videos to support requirements communication. *Softwaretechnik-Trends*, 38(1):51–52, 2018.
- [21] Oliver Karras, Eklekta Kristo, and Jil Klünder. The potential of using vision videos for crowdre: Video comments as a source of feedback. In 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), pages 298–305, 2021.
- [22] Stephan Kiesling, Oliver Karras, and Kurt Schneider. Reqvida– requirements video analyzer. *GI Softwaretechnik-Trends*, 36(3), 2015.
- [23] Oliver Karras, Stephan Kiesling, and Kurt Schneider. Supporting requirements elicitation by tool-supported video analysis. In 2016 IEEE 24th International Requirements Engineering Conference (RE), pages 146–155, 2016.
- [24] Oliver Karras, Svenja Polst, and Kathleen Späth. Using vision videos in a virtual focus group: Experiences and recommendations. arXiv preprint arXiv:2011.00965, 2020.

- [25] Oliver Karras and Kurt Schneider. Software professionals are not directors: What constitutes a good video? In 2018 1st International Workshop on Learning from other Disciplines for Requirements Engineering (D4RE), pages 18–21, 2018.
- [26] Oliver Karras and Kurt Schneider. An interdisciplinary guideline for the production of videos and vision videos by software professionals. arXiv preprint arXiv:2001.06675, 2020.
- [27] Oliver Karras, 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*, 162:110479, 2020.
- [28] H. Kaiya, M. Saeki, and K. Ochimizu. Design of a hyper media tool to support requirements elicitation meetings. In *Proceedings Seventh International Workshop on Computer- Aided Software Engineering*, pages 250–259, 1995.
- [29] Oliver Karras, Carolin Unger-Windeler, Lennart Glauer, and Kurt Schneider. Video as a by-product of digital prototyping: Capturing the dynamic aspect of interaction. In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pages 118–124, 2017.
- [30] Stephen Lindsay, Daniel Jackson, Guy Schofield, and Patrick Olivier. Engaging older people using participatory design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1199–1208, 2012.
- [31] Wendy E. Mackay and Anne Laure Fayard. Video brainstorming and prototyping: Techniques for participatory design. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems, CHI EA '99*, page 118–119, New York, NY, USA, 1999. Association for Computing Machinery.
- [32] Clara Mancini, Yvonne Rogers, Arosha K Bandara, Tony Coe, Lukasz Jedrzejczyk, Adam N Joinson, Blaine A Price, Keerthi Thomas, and Bashar Nuseibeh. Contravision: exploring users' reactions to futuristic technology. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 153– 162, 2010.
- [33] Wendy E. Mackay, Anne V. Ratzner, and Paul Janecek. Video artifacts for design: Bridging the gap between abstraction and detail. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '00*, page 72–82, New York, NY, USA, 2000. Association for Computing Machinery.

- [34] Lukas Nagel and Oliver Karras. Keep your stakeholders engaged: Interactive vision videos in requirements engineering. In 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), pages 51–57, 2021.
- [35] Lukas Nagel, Jianwei Shi, and Melanie Busch. Viewing vision videos online: Opportunities for distributed stakeholders. In 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), pages 306–312, 2021.
- [36] Raphael Pham, Sebastian Meyer, Ingo Kitzmann, and Kurt Schneider. Interactive multimedia storyboard for facilitating stakeholder interaction: Supporting continuous improvement in it-ecosystems. In 2012 Eighth International Conference on the Quality of Information and Communications Technology, pages 120–123, 2012.
- [37] Chris Rupp, Jan Dovica, Michael Bezold, Markus Wiener, and Benedict Fryer. Pimp my spec: Tuning von spezifikationen durch videos. *Softwaretechnik-Trends*, 39(1):15–16, 2019.
- [38] Tom A. Rodden, Joel E. Fischer, Nadia Pantidi, Khaled Bachour, and Stuart Moran. At home with agents: Exploring attitudes towards future smart energy infrastructures. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, page 1173–1182, New York, NY, USA, 2013. Association for Computing Machinery.
- [39] Rick Rabiser, Norbert Seyff, Paul Grunbacher, and Neil Maiden. Capturing multimedia requirements descriptions with mobile re tools. In 2006 First International Workshop on Multimedia Requirements Engineering (MERE'06 - RE'06 Workshop), pages 2–2, 2006.
- [40] Kurt Schneider and Linda Marilena Bertolli. Video variants for crowdre: How to create linear videos, vision videos, and interactive videos. In 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), pages 186–192, 2019.
- [41] Kurt Schneider, Melanie Busch, Oliver Karras, Maximilian Schrapel, and Michael Rohs. Refining vision videos. In International Working Conference on Requirements Engineering: Foundation for Software Quality, pages 135–150. Springer, 2019.
- [42] Harald Stangl and Oliver Creighton. Continuous demonstration. In 2011 Fourth International Workshop on Multimedia and Enjoyable Requirements Engineering (MERE'11), pages 38–41. IEEE, 2011.
- [43] Kurt Schneider. Focusing spontaneous feedback to support system evolution. In 2011 IEEE 19th International Requirements Engineering Conference, pages 165–174, 2011.

- [44] Kurt Schneider, Oliver Karras, Anne Finger, and Barbara Zibell. Reframing societal discourse as requirements negotiation: Vision statement. In 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), pages 188–193, 2017.
- [45] Yuhui Shan, Lin Liu, and Fei Peng. Megore: Multimedia enhanced goal-oriented requirement elicitation experience in china. In 2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering - Beyond Mere Descriptions and with More Fun and Games, pages 37–41, 2008.
- [46] Peter Vistisen and Søren Bolvig Poulsen. Investigating user experiences through animation-based sketching. In Motion Design Education Summit 2015, pages 29–38. Routledge, 2016.
- [47] Peter Vistisen and Søren Bolvig Poulsen. Return of the vision video: Can corporate vision videos serve as setting for participation? 2017.
- [48] Han Xu, Oliver Creighton, Naoufel Boulila, and Bernd Bruegge. From pixels to bytes: Evolutionary scenario based design with video. In Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering, FSE '12, New York, NY, USA, 2012. Association for Computing Machinery.
- [49] Han Xu, Oliver Creighton, Naoufel Boulila, and Ruth Demmel. User model and system model: the yin and yang in user-centered software development. In Proceedings of the 2013 ACM international symposium on New ideas, new paradigms, PUBLIKATIONEN ZUM THEMA “PRODUCTION AND USE OF VIDEOS REQUIREMENTS ENGINEER and reflections on programming and software, pages 91–100, 2013.
- [50] K. Zachos and N. Maiden. Art-scene: enhancing scenario walkthroughs with multi-media scenarios. In Proceedings. 12th IEEE International Requirements Engineering Conference, 2004., pages 360–361, 2004.

Literaturverzeichnis

- [1] Beat Akeret. Grafische darstellungen. In *Rechnen im Labor*, pages 183–196. Springer, 2019.
- [2] Sören Auer, Allard Oelen, Muhammad Haris, Markus Stocker, Jennifer D’Souza, Kheir Eddine Farfar, Lars Vogt, Manuel Prinz, Vitalis Wiens, and Mohamad Yaser Jaradeh. Improving access to scientific literature with knowledge graphs. *Bibliothek Forschung und Praxis*, 44(3):516–529, 2020.
- [3] Victor R Basili. Software modeling and measurement: the goal/question/metric paradigm. Technical report, Institute for Advanced Computer Studies, Department of Computer Science, University of Maryland, 1992.
- [4] Egon Börger, Bärbeö Hörger, David Parnas, and Dieter Rombach. Requirements capture, documentation and validation (dagstuhl seminar 99241). Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 1999.
- [5] Jennifer Brings, Marian Daun, Markus Kempe, and Thorsten Weyer. On different search methods for systematic literature reviews and maps: Experiences from a literature search on validation and verification of emergent behavior. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, pages 35–45, 2018.
- [6] Frank J Furrer. Eine kurze geschichte der ontologie. *Informatik-Spektrum*, 37(4):308–317, 2014.
- [7] Jürgen Heller. Semantic structures. In *Knowledge structures*, pages 117–149. Springer, 1994.
- [8] Markus Kaindl. Springer natures scigraph+ insights. 2018.
- [9] Oliver Karras, Eduard C Groen, Javed Ali Khan, and Sören Auer. Researcher or crowd member? why not both! the open research knowledge graph for applying and communicating crowdre research.

- In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 320–327. IEEE, 2021.
- [10] Oliver Karras, 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*, 162:110479, 2020.
- [11] B. Kitchenham and S Charters. Guidelines for performing systematic literature reviews in software engineering, 2007.
- [12] Stefan Klampfl, Michael Granitzer, Kris Jack, and Roman Kern. Unsupervised document structure analysis of digital scientific articles. *International journal on digital libraries*, 14(3):83–99, 2014.
- [13] John Lyons. *Introduction to theoretical linguistics*, volume 510. Cambridge university press, 1968.
- [14] Laura Haak Marcial and Bradley M Hemminger. Scientific data repositories on the web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10):2029–2048, 2010.
- [15] Sacha Martin, Aybüke Aurum, Ross Jeffery, and Barbara Paech. Requirements engineering process models in practice. In *7th Australian workshop on requirements engineering. Deakin University, Melbourne, Australia*, pages 41–47, 2002.
- [16] N NSB. Publication output: Us trends and international comparisons. *Science and Engineering Indicators*, 2019, 2020.
- [17] Allard Oelen, Markus Stocker, and Sören Auer. Smartreviews: towards human-and machine-actionable reviews. In *International Conference on Theory and Practice of Digital Libraries*, pages 181–186. Springer, 2021.
- [18] Jorge Pérez, Marcelo Arenas, and Claudio Gutierrez. Semantics and complexity of sparql. *ACM Transactions on Database Systems (TODS)*, 34(3):1–45, 2009.
- [19] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, pages 1–10, 2008.
- [20] Klaus Pohl and Chris Rupp. *Basiswissen requirements engineering: Aus- und Weiterbildung nach IREB-Standard zum certified professional for requirements engineering foundation level*. dpunkt. verlag, 2021.

- [21] Alida Rohde. Verifikation von empfehlungen zur produktion und nutzung von vision videos basierend auf subjektiven videoqualitätsbewertungen. B.S. thesis, Hannover: Institut für Praktische Informatik, 2020.
- [22] Markus Stocker, Tina Heger, Artur Schweidtmann, Hanna Ówiek-Kupczyńska, Lyubomir Penev, Milan Dojchinovski, Egon Willighagen, Maria-Esther Vidal, Houcemeddine Turki, Daniel Balliet, et al. Skg4eosc-scholarly knowledge graphs for eosc: Establishing a backbone of knowledge graphs for fair scholarly information in eosc. *Research Ideas and Outcomes*, 8:e83789, 2022.
- [23] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, 1(1):396–413, 02 2020.
- [24] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

