
Aufgabenstellung

Container-basierte Ausführung von Multimedia-Inhaltsanalyseverfahren in GMAF

Abschlussarbeit Multimedia Information Retrieval, SS/WS 2026

Patrick Steinert, MSc

Lehrbeauftragter für Multimedia Informationssysteme und Information Retrieval

am Lehrgebiet Multimedia und Internetanwendungen

Prof. Dr.-Ing. Matthias Hemmje

Fernuniversität Hagen

Universitätsstrasse 1, D-58097 Hagen

Raum: 1 F09, Gebäude 3 (Informatikzentrum, Universitätsstr. 1)

Tel.: +49 2331 987-304

Fax: +49 2331 987-4487

Internet: <https://www.fernuni-hagen.de>

Aufgabenstellung für: Abschlussarbeit Multimedia Information Retrieval

Bearbeitungszeitraum: 2024

Betreuer: Patrick Steinert, M.Sc.

E-Mail: patrick.steinert@fernuni-hagen.de

Skype: marques666

Whatsapp: <https://bit.ly/3PDutUw>

Lehrgebiet Multimedia und Internetanwendungen:

- Daten- und Dokumentenmanagement
- Informations- und Wissensmanagement
- Multimedia
- Informationssysteme und Datenbanken
- Informationsvisualisierung
- Information Retrieval
- Visuelle
- Mensch-Maschine-Interaktion
- Content- und Wissensmanagement im Internet
- Semantic Web
- Digitale Langzeitarchivierung
- Virtuelle Forschungsumgebungen
- Big Data Analyse
- Analyse natürlicher Sprache
- Berufliche Weiterbildung und E-Learning
- Industrie 4.0 und "Factories of the Future"

Die nachfolgenden Beschreibungen für die Aufgabenstellung im Bereich Informatik benennen die nötigen Aufgaben und Teilaufgaben und sollen ebenfalls als Übersicht und Leitfaden für die zu erbringenden Leistungen dienen. Zur Bearbeitung der nachfolgenden Aufgaben werden folgende Fähigkeiten und Kenntnisse vorausgesetzt:

- **Wissenschaftliches Arbeiten:** strukturiertes, analytisches Vorgehen auf Basis der gültigen Konventionen für Aufbau, Zitierweise und inhaltliche Bearbeitung von forschungsrelevanten Themen sind die Voraussetzung für das Schreiben einer Bachelor- oder Master-These. Hierzu gehören die eigenständige Literaturrecherche, die selbstständige Aneignung von Wissen und die Umsetzung von theoretischen Konzepten in die Praxis.
- **Programmiersprachen / Beschreibungssprachen:** Java, XML. Von Vorteil sind Kenntnisse im Umgang mit Eclipse und ein grundlegendes Verständnis von Konzepten aus dem Multimedia Information Retrieval, Multimedia Informationssystemen und Semantik.
- **Sonstige Systeme oder Konzepte:** Objektorientierte Systementwicklung, Design Patterns, Dokumentation, Umgang mit Repositories, Dependency Management (z.B. Maven), grundlegende Konzepte der semantischen Modellierung, RDF, Semantic Web. Erfahrung in der Anwendung von Machine Learning (ML) Modellen sind von Vorteil.
- **Methodik:** Anwendung von Forschungsmethodiken zum strukturierten Lösen von Problemstellungen und zum Erstellen von wissenschaftlich-technischen Dokumenten (z.B. Nunamaker [1]).

1. Themenbeschreibung

In der Aufgabe geht es um ein System zur Verarbeitung der Inhaltsanalyse (engl. Content Analysis) von Metaverseinhalten. Beispiele für die Inhaltsanalyse sind Verfahren zur KI-basierten Objekterkennung, Bildbeschreibungen oder Inhaltszusammenfassung. Die Verarbeitung findet beispielsweise in Multimedia Information Retrieval (MMIR) Systemen statt. Integrative Systeme erlauben es, bestehende Verfahren zur Inhaltsanalyse zu integrieren. In der praktischen Anwendung stellt sich jedoch heraus, dass die Verfahren oft unterschiedliche Laufzeitumgebungen erfordern, die untereinander inkompatibel sein können. Daher kann beispielsweise Verfahren X und Verfahren Y nicht auf dem selben Computer gleichzeitig installiert werden. Containervirtualisierung (engl. OS-Level Virtualization) erlaubt die Kapselung von Laufzeitumgebungen, die zudem noch gut skalieren können. Die Aufgabe beschäftigt sich mit der Integration von Containervirtualisierung für die Inhaltsanalyse im MMIR.

Zum weiteren Verständnis wird zunächst der Kontext der Aufgabe erklärt.

Multimedia Information Retrieval

Multimedia Information Retrieval [2] beschäftigt sich mit der Fragestellung, wie Multimedia-Inhaltsobjekte (engl. Multimedia Content Objects, z. B. Texte, Bilder, Audio, Video) analysiert und deren Merkmale erfasst und für die Suche genutzt werden können. Typischerweise werden hierfür z. B. Algorithmen zur Objekterkennung in Bildern, Speech-To-Text Verfahren oder auch Mechanismen zur Szenen-Erkennung in Videos eingesetzt, die allesamt Multimedia Merkmale ermitteln können. Merkmale können technischer Natur sein (z.B. Farbwert eines Bildes, Auflösung, Speicherformat, verwendeter Codec), inhaltsbasiert (z.B. Objekt erkannt, das Wort „Jaguar“ erkannt) oder auch semantisch ausgeprägt sein (z.B. erkanntes Objekt ist ein Mensch, mit „Jaguar“ ist das Tier gemeint, der Mensch steht dem Jaguar gegenüber). Für die Extraktion und Verwaltung solcher Merkmale gibt es eine Reihe von Tools und APIs, die sich jeweils auf einen bestimmten Zweck spezialisiert haben. Beispielsweise kann mit Google Vision die Erkennung von Objekten in Bildern unterstützt werden, Shazam ermittelt aus einem Audio-Stream den Titel und den Interpreten eines Musikstücks, usw.

Metaverse

Im Bereich Multimedia sind seit einigen Jahren Entwicklungen in Metaverse-Technologien zu beobachten. Das Metaverse [3],[4] ist ein immerwährender digitaler Multiuser-Raum, ein Zusammenschluss von mehreren immersiven virtuellen Welten [5]. Die virtuellen Welten können Virtuelle Realität [6] (engl. Virtual Reality, VR) und Erweiterte Realität [7] (engl. Extended Reality, XR) beinhalten, bestehen aber immer auch aus multimedialen Inhalten. Am Lehrgebiet Multimedia und Internetanwendungen werden Synergien zwischen MMIR und dem Metaverse erforscht.

Metaverse Recordings

Im Metaverse können Benutzer verschiedene Dinge tun. Was auch immer sie tun, sie können dabei indirekt ihre Sitzung (engl. Session) aufzeichnen [8]. Eine Aufzeichnung ist ein Multimedia-Content-Object (MMCO), z. B. eine Bildschirmaufnahme in Form eines gespeicherten Videos. Metaverse-Aufzeichnungen bezeichnen wir als *Metaverse Recording* (MVR). MVRs können in MMIR-Systeme importiert und auffindbar gemacht werden, um eine breite Palette von Anwendungen zu unterstützen.

Metaverse Recording Retrieval

Im Rahmen der laufenden Dissertation „Supporting Information Retrieval of Metaverse Recordings“ von Patrick Steinert ist das *Processing Framework for Metaverse Recordings* (PFMR) entstanden. Wie in Abbildung 2 dargestellt, beschreibt PFMR die Schritte zur Analyse und Indexierung von *Metaverse Recordings* (MVRs). MVRs sind Aufzeichnungen von Metaverse-Sitzungen, mit denen ein Benutzer seinen Besuch im Metaverse dokumen-

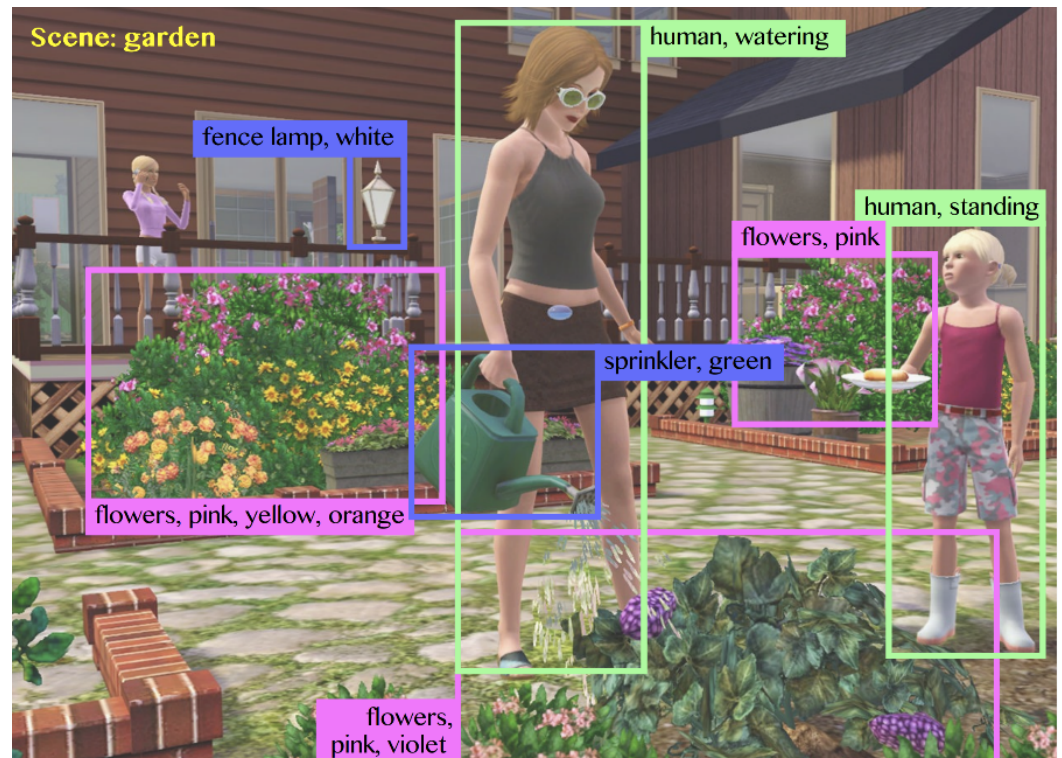


Abbildung 1. Beispiel einer Objekt-, Interaktions- und Szenenerkennung in einer Metaverseaufnahme.

tiert. Die Aufgabenstellung ordnet sich in diese Dissertation ein, indem Sie bestimmte Problemstellungen in der *MVR-Analyse*, insbesondere der *MVR Feature Extraction*, *MVR Data Mining* und *MVR Indexing* beschreibt.

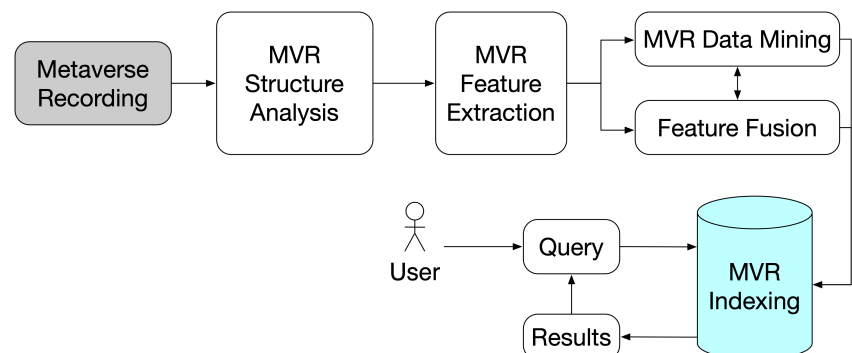


Abbildung 2. Processing Framework for Metaverse Recordings

Zur Beantwortung der Forschungsfrage (engl. Research Question, RQ) **RQ3** aus der Dissertation, "Can an MMIR system be developed that supports MVR and its various media and data types?", ist die Zielsetzung der Forschung die Modellierung, Implementierung und Evaluierung eines "MVR-extended MMIR Systems" festgelegt worden. Diese Abschlussarbeit ist an der RQ3 angelehnt und untersucht die Integration von Verfahren zur Inhaltsanalyse.

Problembeschreibung: Multimedia-Inhaltsanalyse und Indexierung

Das am Lehrgebiet entwickelte MMIR-System Generic Multimedia Analysis Framework (GMAF) [9] wurde so konzipiert, dass es vielfältige Verfahren zur Analyse von Multimedia, die sogenannte Feature Extraction, unterstützen kann. Dazu hat GMAF eine

Plugin-Schnittstelle, mit der Analyseverfahren als Plugin implementiert und eingebunden werden können, beispielsweise YoloObjectDetection für die Objekterkennung in Bildern, AvatarDetector für die Erkennung von Avataren in Videos oder HOIPlugin für die Interaktionserkennung in Videos. Die Plugins werden lokal ausgeführt. Daher müssen auf dem System die jeweiligen Installations- und Systemvoraussetzungen erfüllt sein, z. B. bestimmte Python-Versionen und Bibliotheken. Das führt zu Problemen, wenn Plugins untereinander inkompatible Voraussetzungen haben: Beispielsweise benötigt *Plugin A* Python 3.8 mit einer bestimmten PyTorch-Version, während *Plugin B* Python 3.11 mit einer neueren, dazu inkompatiblen PyTorch-Version erfordert. In solchen Fällen können nicht beide Plugins auf demselben System ausgeführt werden.

Ein weiteres Problem ist die Dauer der Analyse. Die MMIR-Indexierung erfordert die Analyse und Verarbeitung großer Mengen an multimodalen Daten, wie Texten, Bildern, Audio und Video, um relevante Informationen zu extrahieren und zugänglich zu machen. Die Indexierung großer Sammlungen ist zeitintensiv – beispielsweise kann die Analyse von 1.000 Videos mehrere Tage dauern. Im einfachen Fall werden die Aufgaben sequenziell bearbeitet. Einige Verfahren nutzen intensiv den Hauptprozessor, während andere stark mit einem oder mehreren Grafikprozessoren arbeiten. Eine Parallelisierung oder Verteilung der Arbeit ist in GMAF unzureichend berücksichtigt. Zum Beispiel hat GMAF zwar die Möglichkeit, die Multimediasammlung auf mehrere Systeme zu verteilen, jedoch wird die Indexierung eines Elements dann nur auf dem Knoten ausgeführt, auf dem das Element gespeichert ist. Eine dynamische Verteilung der Indexierung ist nicht vorgesehen. Da verteilte Systeme mehrere Rechner oder Knoten nutzen, um komplexe Aufgaben zu bewältigen, entsteht die Herausforderung, die Inhaltsanalyse so zu gestalten, dass sie über verschiedene Knoten hinweg effizient durchgeführt werden kann.

Aufgabenstellung

Im Rahmen dieser Arbeit ist ein Lösungsansatz zu modellieren, zu implementieren und zu evaluieren, der es ermöglicht, GMAF-Plugins mit unterschiedlichen Abhängigkeiten konfliktfrei auszuführen. Dazu soll ein Lösungsansatz mit Containervirtualisierung verfolgt werden, bei der jedes Plugin mitsamt seiner Laufzeitumgebung isoliert betrieben werden kann. Containervirtualisierung, wie z. B. Docker oder OpenContainer, virtualisiert den Prozess, also die Instanz eines Programms, die die Analyse durchführt, inklusive der Programmabhängigkeiten. Abbildung 3 zeigt einen möglichen Lösungsansatz für die Ausführung. Dazu muss jedoch GMAF die einzelnen Container starten und beenden sowie mit den Prozessen kommunizieren können, um die Analyseaufträge zu senden und die Ergebnisse abzuholen. Dies soll mit einer HTTP-API oder einem Messaging-System (z. B. REST-API, RabbitMQ) gelöst werden.

Die Umsetzung soll für mindestens drei GMAF-Plugins erfolgen, die mit dem Betreuer abzustimmen sind. Möglich sind beispielsweise YoloObjectDetection (Objekterkennung), AvatarDetector (Avatar-spezialisierte Objekterkennung) und HoiDetector (Human-Object-Interaction Detection).

Für eine skalierte und/oder parallele Ausführung ist ein Lösungsansatz zu modellieren, zu implementieren und zu evaluieren. Durch die Verwendung von Containervirtualisierung können die Verfahren auch auf einem verteilten System dynamisch ausgeführt werden, z. B. in einem Kubernetes-Cluster, der lokal (z. B. mit Minikube oder k3s) oder in einer Testumgebung betrieben wird. Abbildung 4 zeigt einen möglichen Lösungsansatz. Dadurch könnten die Verfahren möglicherweise eine höhere Effizienz in der Verarbeitungszeit und der Ressourcenauslastung erreichen.

Die Umsetzung soll ebenfalls für mindestens drei GMAF-Plugins in Abstimmung mit dem Betreuer erfolgen.

Die weiteren Abschnitte der Aufgabenbeschreibung erläutern die formellen Aspekte der Arbeit.

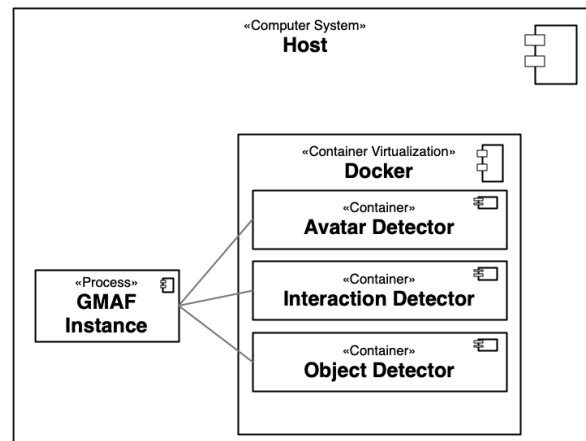


Abbildung 3. Möglicher Ansatz für Problem 1

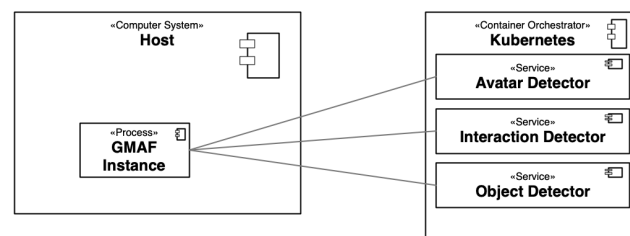


Abbildung 4. Möglicher Ansatz für Problem 2

2. Theoretischer Teil

Der theoretische Teil der Arbeit wird nach der Methodik von Nunamaker [1] aufgebaut und methodisch hergeleitet. Dies bedeutet, dass sich auch die Dokumentenstruktur von Exposé und Ausarbeitung auf die Nunamaker Methodik stützt und entsprechende Forschungsziele der Nunamaker-Typen (Beobachtung, Theoriebildung, Implementierung und Experiment) definiert werden. Daneben folgt die Arbeit den UCSD-Konzepten von Norman & Draper [10]. Als Modellierung wird die UML [11] verwendet. Im einem Exposé formulieren Sie zunächst Ihre Forschungsfragen und leiten die zugehörigen Forschungsziele ab. Danach wird dieses Exposé freigegeben und stellt auch das erste Kapitel Ihrer späteren Ausarbeitung dar. Die praktische Durchführung der Forschungsziele (Recherche, Modellierung, Implementierung und Evaluierung) gehört zum praktischen Teil der Arbeit, die Dokumentation der Ergebnisse in der Arbeit wiederum zum Theoretischen Teil. Im Theoretischen Teil führen Sie nach dem ersten Kapitel in den Stand der Wissenschaft und Technik ein und stellen die für das Thema relevanten Konzepte vor. Auf dieser Basis kann dann die Modellierung beschrieben werden, gefolgt von der Implementierung und Evaluierung. Weitergehende Informationen zum Thema Aufbau von wissenschaftlich-technischen Dokumenten finden Sie unter [12]. Das Kennwort, um die Inhalte ansehen zu können lautet FUH-SW!.

3. Praktischer Teil

Der praktische Teil der Arbeit enthält zunächst die Recherche in den jeweiligen Bereichen der Arbeit, zum Beispiel über eine Literaturrecherche oder Expertenbefragung. Diese wird gefolgt von der Modellierung und die programmiertechnische Umsetzung der Modellierung sowie die experimentelle Überprüfung. Als Programmiersprache kommt typischerweise Java oder Python zum Einsatz. Die Verwendung von Standard-Entwicklungstools wie Eclipse, GIT, Maven, Javadoc wird vorausgesetzt. Im Rahmen des praktischen Teils erfolgt auch die Erstellung von Test-Cases und deren Dokumentation, eine Evaluierung der

Ergebnisse inklusive der Wahl und Begründung einer geeigneten Evaluierungs-Methodik, sowie ein Ausblick auf zukünftige Ergänzungen. Im Rahmen der Implementierung folgen Sie den GoF Design Patterns und orientieren Ihre Arbeit an den bestehenden Strukturen. Hierbei sind insbesondere das *Command Pattern*, das *Abstract Factory Pattern*, *Singleton Strukturen* und das *Model-View-Controller Prinzip* zu berücksichtigen. Den Einsatz und die Verwendung der Patterns stimmen Sie mit Ihrem Betreuer ab.

4. Betreuung

Die inhaltliche und fachliche Betreuung der Arbeit wird von Patrick Steinert übernommen. Hierzu steht Ihnen jederzeit die Kommunikation per Skype, Telefon, E-Mail zur Verfügung, regelmäßige Abstimmungstermine mit dem Betreuer helfen Ihnen, die Arbeitsergebnisse strukturiert voranzutreiben und geben Ihnen die Gelegenheit, durch unmittelbares Feedback zügig und effizient die Ziele der Arbeit zu erreichen. Allgemeine Fragen sollten Sie über das Moodle Forum stellen, da so alle von der Antwort profitieren. Diverse "Quality-Gates" stellen sicher, dass die Ausarbeitung auch die Anforderungen der Universität und des Lehrgebietes erfüllt und somit zu einem erfolgreichen Ergebnis führen wird.

5. Aufgaben im Vorfeld

Bevor Sie sich dafür entscheiden, das Thema zu bearbeiten, sollten Sie sich einen groben Überblick zu den benötigten Voraussetzungen oder Hintergründen im Rahmen einer Recherche verschaffen. Im Literaturverzeichnis dieses Dokuments sind eine Reihe von Quellen hierzu aufgeführt. Dies ist Bestandteil der ersten Einsendeaufgabe. Anschließend beginnen Sie mit der Erstellung eines Exposé, welches den genauen Rahmen Ihrer Arbeit absteckt. Die Gliederung des Exposé sieht hierbei wie folgt aus:

1. **Einleitung** (Aus den Entwicklungen in der Welt erfolgt die Beschreibung des Themas und Einordnung in ein übergeordnetes Themengebiet)
 - 1.1. **Motivation** (Entwicklungen in der Forschungsgruppe, Persönliche Motivation der Arbeit)
 - 1.2. **Problembeschreibung** (Formale Beschreibung des Problems durch die Definition von Problemfeldern)
 - 1.3. **Forschungsfragen** (Definition der Forschungsfragen zu den Problemfeldern)
 - 1.4. **Methodik und Ableitung der Forschungsziele** (z.B. nach Nunamaker)
 - 1.5. **Ansatz** (Beschreibung, wie die Forschungsziele gruppiert werden und wo sie im Rahmen der späteren Arbeit behandelt werden)
 - 1.6. **Aufbau** (Festlegung der Struktur Ihrer Arbeit)
 - 1.7. **Zeitplan** (inklusive Meilensteinplanung und Gantt-Diagramm auf Basis der Forschungsziele)
 - 1.8. **Zusammenfassung**

Die Ergebnisse fassen Sie im Rahmen eines Dokumentes (Exposé) in ca. 5 bis 10 Seiten zusammen. Das Exposé ist die Basis für das weitere Dokument und gibt einen für alle Beteiligten definierten Rahmen der Arbeit vor. Dadurch werden unter Anderem folgende Punkte sicher gestellt:

- Überprüfung der formalen und schriftlichen Fähigkeiten des Studenten zur Erstellung und Ausarbeitung einer Fachpraktikum-Ausarbeitung.
- Vermittlung der Herangehensweise für die Strukturierung des wissenschaftlichen Diskurses
- Entwicklung eines Grundverständnisses für die zu erstellende Fachpraktikum-Ausarbeitung

6. Die Ausarbeitung der Abschlussarbeit

Durch die Erstellung des Exposé haben Sie bereits einen wichtigen Teil der Fachpraktikum-Ausarbeitung fertig gestellt: die saubere wissenschaftliche Verankerung, die daraus resultierende Ableitung der Forschungsziele, sowie den Ansatz und die Struktur der Arbeit. Deshalb können Sie die im Rahmen des Exposé erstellten Kapitel direkt in Ihre Fachpraktikum-Ausarbeitung übernehmen (bis einschl. 1.6). Nun erfolgt:

- eine weitergehende, selbständige Literaturrecherche
- eine Beschreibung des aktuellen Stands von Wissenschaft und Technik
- das Konzeptionelle Design der Architektur für die Umsetzung und Implementierung
- die praktische Einarbeitung (Proof-Of-Concept Implementierungen, Kennenlernen der APIs und Frameworks)
- das Erstellen von Test-Szenarios, ggfs. auch die Organisation von Test-Daten
- die Implementierung
- eine Evaluierung des Ergebnisses mit Vergleich und Ausblick

Die jeweiligen Ergebnisse halten Sie in Ihrer Fachpraktikum-Ausarbeitung fest und folgen dabei analog zu den definierten Forschungszielen einer logischen Struktur:

1. **Einleitung** (folgt dem Exposé bis einschl. 1.6)
2. **Stand der Wissenschaft und Technik** (Für jedes Forschungsziel vom Typ Beobachtung wird ein Kapitel vorgesehen und der zugehörige Stand der Wissenschaft beschrieben. Am Ende werden remaining challenges" definiert und zusammengefasst.)
3. **Modellierung** (folgt dem Aufbau von Kapitel 2 und sieht für jedes Forschungsziel vom Typ Theoriebildung ein Kapitel vor)
4. **Implementierung** (folgt dem Aufbau von Kapitel 2 und bearbeitet Forschungsziele vom Typ Implementierung)
5. **Evaluierung** (folgt dem Aufbau von Kapitel 2 und bearbeitet Forschungsziele vom Typ Experiment)
6. **Diskussion** (Wie kann das Gesamtergebnis eingeordnet werden? Welche remaining challenges" verbleiben? Was kann Teil von future work sein? Zusammenfassung)

Die Arbeit ist so ausgelegt, dass der Anteil von Theorie und Praxis jeweils 50% beträgt. Durch die hier bereits vorgeschlagene Herangehensweise werden Sie zu einer effizienten, wissenschaftlich korrekten und nachvollziehbaren Arbeitsweise angeleitet. Im Folgenden finden Sie eine Reihe weitergehender Links und Referenzen, die Ihnen einen Einstieg und Überblick über das Thema erleichtern. Sollten Sie Interesse an diesem Thema haben, kontaktieren Sie bitte einen der genannten Ansprechpartner.

References

1. Nunamaker Jr, J.F.; Chen, M.; Purdin, T.D. Systems development in information systems research. *Journal of management information systems* **1990**, *7*, 89–106. Publisher: Taylor & Francis.
2. Raieli, R. *Multimedia information retrieval: theory and techniques*; Chandos information professional series, Chandos Publishing: Oxford, 2013.
3. Anderson, J.; Rainie, L. The metaverse in 2040 **2022**.
4. MetaVRse. The Metaverse Manifesto, 2022.
5. Mystakidis, S. Metaverse. *Encyclopedia* **2022**, *2*, 486–497. doi:10.3390/encyclopedia2010031.
6. Wikipedia. Virtual world, 2023. Page Version ID: 1141563133.
7. Wikipedia. Augmented reality, 2023. Page Version ID: 1142793928.
8. Steinert, P.; Wagenpfeil, S.; Frommholz, I.; Hemmje, M.L. Towards the Integration of Metaverse and Multimedia Information Retrieval. 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRaine); IEEE: Milano, Italy, 2023; pp. 581–586. doi:10.1109/MetroXRaine58569.2023.10405728.
9. Wagenpfeil, S.; Kevitt, P.M.; Hemmje, M. Smart Multimedia Information Retrieval. *Analytics* **2023**, *2*, 198–224. doi:10.3390/analytics2010011.

10. Norman, D.A.; Draper, S.W., Eds. *User centered system design: new perspectives on human-computer interaction*; L. Erlbaum Associates: Hillsdale, N.J, 1986.
11. Object Management Group. Unified Modeling Language, 2011.
12. Wagenpfeil, S. Erstellen von wissenschaftlich-technischen Dokumenten. <http://www.stefan-wagenpfeil.de/nunamaker/>, 2022. Kennwort: FUH-SW!