THM | CAMPUS GIESSEN | MNI
Mathematik, Naturwissenschaften und Informatik
TECHNISCHE HOCHSCHULE MITTELHESSEN

Bachelorthesis

Konzept und Realisierung eines datentschutzkonformen Software-Analysesytems zur datengetriebenen Planung

zur Erlangung des akademischen Grades

Bachelor of Science

eingereicht im Fachbereich Mathematik, Naturwissenschaften und Informatik an der Technischen Hochschule Mittelhessen

von

Polina Eremina

30. Januar 2024

Referent: Prof. Dr. Dennis Priefer
Korreferent: Samuel Schepp

Erklärung der Selbstständigkeit

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate deutlich kenntlich gemacht zu haben.

Gießen, den 30. Januar 2024 Polina Eremina

Inhaltsverzeichnis

1 Einführung

Die vorliegende Arbeit widmet sich der Konzipierung und Entwicklung eines Analysesystems, das Unterstützung im Entscheidungsprozess bei der Optimierung der Anwendung und des Nutzererlebnisses bieten soll. Dieses Kapitel hat zum Ziel, die Motivation und Herausforderungen zu erläutern, die Forschungsfragen vorzustellen und die methodische Vorgehensweise zu beschreiben.

1.1 Problembeschreibung/Motivation

"Business Intelligence und Analytics (BI & A) sowie der Bereich Big Data Analytics haben in den letzten zwei Jahrzehnten sowohl in der akademischen als auch in der Geschäftswelt zugenommen" [Übers. d. Verf.] (vgl. [Che12]). Gemäß IDC (International Data Corporation) wachsen solche Bereiche wie Big Data, Analyse und KI um 20% jährlich in der Asia-Pazifik-Region (APAC).

Laut Paul Burton, Hauptgeschäftsführer bei IBM APAC, sind erfolgreiche Unternehmen diejenigen, die schnell aus ihren Erfahrungen lernen und expandieren. Dies basiert zu 100% auf der Interpretation und dem Lernen aus Daten (vgl. [Zul22]). Dafür sind idealerweise eine gute Datenarchitektur, eine solide Datenstruktur und die Fähigkeit, Daten zu analysieren, Schlussfolgerungen zu ziehen und dann Entscheidungen zu treffen, erforderlich, so Burton.

Aus diesem Grund sind analytische Systeme zu essenziellen Werkzeugen geworden, die Unterstützung und Hilfe bereitstellen. Eine jüngste Umfrage zu den Trends in der Nutzung von den Analysesystemen im Jahr 2023, durchgeführt von 450 Softwareunternehmen, betont die zunehmende Bedeutung der Sammlung von Produkt- und Nutzerdaten. (siehe Abbildung 1.1) Im Vergleich zu 2022, als 91% der befragten Unternehmen die analytischen Systeme einsetzten, nutzen mittlerweile 97% der Unternehmen diese bereits oder planen, sie bis 2025 zu implementieren. Allerdings geben lediglich 26% an, dass die Datensammlung effizient erfolgt. Ein Anstieg der kommerziellen Nutzung von Analysesystemen von 10% auf 26% im Vergleich zum Vorjahr verdeutlicht diesen Trend. Interessanterweise sammeln 24% der Unternehmen derzeit telemetrische Daten (Nutzungsdaten der Software), die allerdings nicht für weiterführende Auswertung

genutzt werden. Während nur 3% der befragten Unternehmen keine datengetriebene Analyse einsetzen, zögern weitere 5% aufgrund von Bedenken bezüglich Nutzerakzeptanz und Datenschutzverordnungen. (vgl. [Chr23]) Diese Unterschiede weisen auf Potenzial für eine optimierte Datenerfassung und -verarbeitung hin.

Der Einsatz von den Analysesystemen ermöglicht es Unternehmen, die faktenbasierten Entscheidungen zu treffen und daraus resultierend die organisatorische Fähigkeiten sowie die Agilität zu verbessern (vgl. [Pop18]). Ein anschauliches Beispiel ist Netflix, eine Streaming Plattform mit fast 247 Millionen Abonnenten. Die Menge von Daten, die dabei erzeugt wird, ermöglicht Netflix die Erkenntnisse von den Gewohnheiten und Verhalten des Nutzers mithilfe von dem Analysesystem zu gewinnen. Sie haben erkannt, dass die Investition in die Produktion die Serie “House of Cards” dazu beitragen kann, eine noch größere Anzahl von Abonnenten anzuziehen. Basierend auf den gesammelten Nutzerdaten haben sie festgestellt, dass ihr Publikum ein erhöhtes Interesse an Filmen mit dem Schauspieler Kevin Spacey sowie an der britischen Version der Serie zeigt. Darüber hinaus erfreuen sich politische Serien und Filme allgemein positiver Bewertungen auf ihrer Plattform. (vgl. [Pat13])

Daher traf Netflix die Entscheidung, 100 Millionen Dollar in die Unterstützung der Produktion dieser Serie zu investieren, was sich letztendlich als rentabel erwies. Netflix hat einen großen Gewinn erzielt, indem sie bei der Veröffentlichung der Serie 2 Millionen neue Abonnenten gewonnen haben (eine Steigerung um 7% gegenüber dem letzten Quartal) Bei der Ankündigung und Bewerbung der Serie auf der Plattform nutzte Netflix auch die gewonnenen Erkenntnisse und Daten, um die Serie effektiver zu vermarkten. Für Nutzer, die Filme mit Kevin Spacey gesehen haben, wurde ein Trailer mit Ausschnitten aus seinen Auftritten gezeigt. Für das Publikum, das häufiger Inhalte mit weiblichen Hauptfiguren ansieht, wurde ein Trailer aus weiblicher Perspektive präsentiert. (vgl. [Pat13])

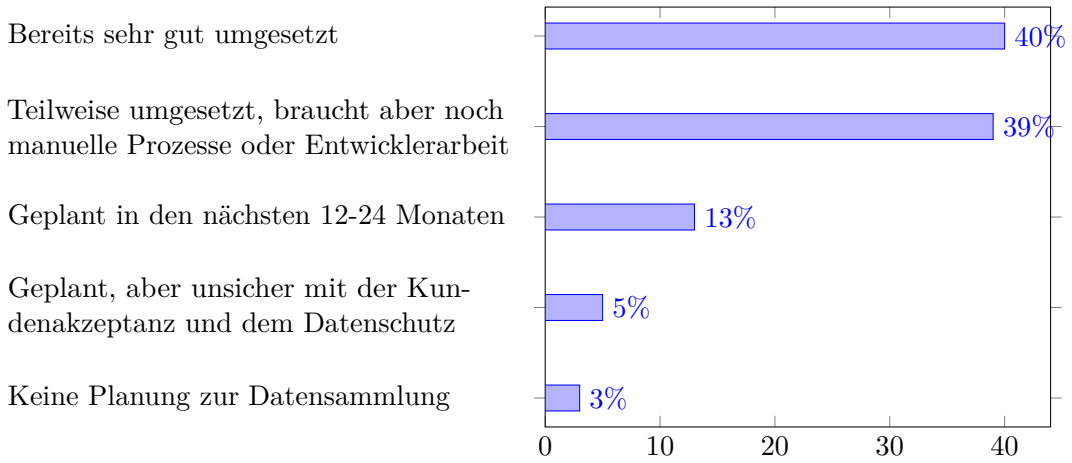

Abbildung 1.1: Analyse Trends von der Befragung 2023 (vgl. [Chr23])

Die Relevanz von Analysesystemen für Unternehmen ist evident. Allerdings zeigen Umfrageergebnisse, dass die Datensammlung in vielen Unternehmen ineffizient erfolgt. Hier besteht Potenzial für Verbesserungen in Bezug auf die Datenerfassung und -verarbeitung. Aufgrund der kommerziellen Tätigkeiten möchten Unternehmen oft nicht wissenschaftlich veröffentlichen, wie die Daten gesammelt und gespeichert werden (vgl. [Can23]). Die Nutzung von Drittanbieterlösungen bietet zwar eine Option, gewährt jedoch nicht die volle Kontrolle über Daten und Privatsphäre.

Einige Unternehmen zögern aufgrund von Bedenken hinsichtlich Nutzerakzeptanz und Datenschutzverordnungen, Drittanbieterlösungen einzusetzen oder selbst ein internes System zu entwickeln. Dies erfordert Fachkenntnisse und Ressourcen, die möglicherweise nicht in allen Unternehmen verfügbar sind. Daher zielt diese Arbeit darauf ab, ein Modell für den Aufbau eines Analysesystems zu konzipieren, das Unternehmen eine Alternative zu Drittanbieterlösungen bietet und ihnen ermöglicht, effektive Tools intern zu implementieren. Dies kann besonders in Branchen wie Medizin, Finanzen, Telekommunikation und Bildung relevant sein, die mit sensiblen Daten arbeiten.

1.2 Ziele dieser Arbeit

Das Ziel dieser Arbeit ist es, eine Unterstützung für die Entscheidungsfindung bei der Optimierung der Anwendung und ein verbessertes Verständnis des Nutzerverhaltens in Form eines Analysesystems zu erschaffen. Dadurch soll eine zielgerichtete Entwicklung ermöglicht werden, die die Priorisierung essenzieller und populärer Funktionen fördert und gleichzeitig unwesentliche Elemente eliminiert.

Aus dieser Zielsetzung ergeben sich folgende Forschungsfragen:

FF1 : Wie können Daten von einer Anwendung am effizientesten erfasst werden?

FF2 : Welche Schritte müssen unternommen werden, um die Datenschutzbestimmungen der DSGVO einzuhalten und die Daten datenschutzkonform zu verarbeiten?

FF3 : Wie könnte eine übersichtliche Darstellung der erfassten Daten aussehen, um eine Menge von Daten wiederzugeben und daraus Muster und Erkenntnisse zu gewinnen?

FF4 : Inwieweit kann das Analysesystem die Entscheidungsfindung im Produktmanagement zur Optimierung der Anwendung und des Nutzererlebnisses unterstützen?

1.3 Vorgehensweise/Methode

Die Durchführung der Arbeit erfolgt deduktiv und mittels quantitativer Methoden. Es wird die Hypothese aufgestellt, dass die Nutzung des Analysesystems zu einem verbesserten Verständnis der Nutzer und ihrer Bedürfnisse sowie zu einer effizienteren Planung führt. Um die optimale Datenerfassung FF1 zu untersuchen, wird eine Literaturrecherche durchgeführt. Des Weiteren wird eine statistische Analyse zur Auswahl geeigneter Metriken angewendet. Zur Überprüfung der Hypothese wird ein Konzept zur Beantwortung der Forschungsfragen wie FF2 und FF3 erarbeitet. In der Realisierungsphase am Beispiel einer Anwendung in der Firma Janitza electronics GmbH wird dieses Konzept exemplarisch umgesetzt. Die weitere Auswertung des realisierten Systems FF4 erfolgt durch Interviews und Beobachtung der Systemnutzung.

1.4 Abgrenzung

Diese Arbeit richtet sich nach der Ausarbeitung eines eigenen Analysesystems ohne Nutzung von Drittanbieterlösungen. Die Nutzung von Nutzertests als Methode zur Datenerhebung wird ausgeschlossen. Der Hauptfokus der Arbeit liegt auf der Entwicklung eines Systems, das datenschutzkonform und erweiterbar ist. Die Visualisierung der Daten wird auf der Benutzeroberfläche angeboten, um Fachkräften wie Produktmanagern eine leichte und schnelle Übersicht zu ermöglichen. Die weitergehende Interpretation erfolgt jedoch je nach Fragestellung manuell und wird nicht automatisiert durchgeführt.

1.5 Struktur der Arbeit

Im Kapitel 1 wird die Relevanz und Problematik des gewählten Themas beleuchtet. Zudem werden die Forschungsfragen formuliert und abgegrenzt, sowie die Methoden zur Untersuchung ausgewählt. Im Kapitel 2 werden die relevanten Hintergründe sowie die theoretische Basis ergänzt. Kapitel 3 gibt die Beschreibung des Konzepts des Analysesystems wieder. Zunächst wird die Systemarchitektur vorgestellt, darauf folgend werden die Metriken und die Datenerfassungsmethoden ausgewählt und beschrieben. Die Vorgehensweise zur datenschutzkonformen Datenverarbeitung wird anschließend dargestellt. Danach werden die Optionen zur Speicherung, der Entscheidungsprozess beim Analysesystem sowie das Visualisierungsmodell ausgearbeitet. Am Ende des Kapitels erfolgt ein Gesamtüberblick des Systems. Das ausgearbeitete Konzept wird im nächsten Kapitel 4 im Rahmen einer Firma umgesetzt. Zuerst wird die Firma und die Problematik aus der Sicht des Unternehmens vorgestellt. Im nächsten Abschnitt wird der aktuelle Stand des Analysesystems betrachtet. Dann erfolgt die eigentliche

Realisierung des Konzepts. Am Ende der Arbeit folgt die Zusammenfassung mit der Auswertung und einem Ausblick auf die nächsten Schritte in diesem Thema im Kapitel 5.

2 Hintergrund

In diesem Kapitel werden die relevanten theoretischen Hintergründe erläutert, um das Konzept besser zu verfolgen und zu verstehen.

2.1 Datenerhebungsmethoden

Um Analyse zu starten, besteht erst die Herausforderung die technischen Möglichkeiten zur Datenerhebung auszuwählen.

2.1.1 Log Dateien

Die ersten Analysesysteme in der Vergangenheit haben die sogenannten Logdateien zur Analyse genutzt. Vorausgesetzt ist die Protokollierung jedes einzelnen Abrufs vom Webserver in einer Logdatei. Die Analyse wird folglich serverseitig durchgeführt. Vorteilhaft ist die Feststellung der Vollständigkeit des Abrufs oder Downloads bei diesem Verfahren. Zu den Nachteilen gehören die Unmöglichkeit, in der Echtzeit Analyse zu betreiben und nicht direkte Webserver Abrufe wie von einem Proxy zu erfassen. (vgl. [Kai10])

So kann eine Log-Datei mit der relevanten Information aussehen (vgl. [Ade10]):

```
1 183.121.143.32 - - [18/Max/2003:08:04:22 +02001 "GET /images/logo. jpg
      HTTP/1.1"
2 200 512 "http://www.wikipedia.org/" Mozilla/5.0 (X11; U; Linux 1686; de-
      DE;rv:1.7.5)
3 183.121.143.32 - - [18/Mar/2003:08:05:03 +0200] "GET /images/bild.png
      HTTP/1.1"
4 200 805 "http://www.google.org/"
```

Listing 2.1: Log File-Datei

Wesentliche Elemente, die bei dieser Methode erfasst werden, sind:

- IP-Adresse

- Zeitstempel
- abgerufene Datei
- Statuscode
- verwendeter Browser
- Referrer

2.1.2 Page Tagging

Eine andere Option besteht darin, den JavaScript-Code in den Quelltext der Seiten einzubinden, sodass der Code jedes Mal aufgerufen wird, wenn die Seite besucht wird (vgl. [Ade10]). Dies ermöglicht die Kommunikation zwischen der eingebauten Schnittstelle und dem Anbieter dieses Codes. Üblicherweise handelt es sich bei diesem Anbieter um einen externen Toolanbieter und auch um einen Application Service Provider (ASP). (siehe Abbildung 2.1) In dieser Arbeit wird dieser Teil übernommen und sich selbst um die Speicherung der erhobenen Daten beim Konzept 3 gekümmert. Da dieses Verfahren clientseitig durchgeführt wird, stellt es die Möglichkeit zur Echtzeitanalyse und zur Erfassung der Abrufe über Proxies im Gegensatz zu den Logdateien dar (vgl. [Kai10]).

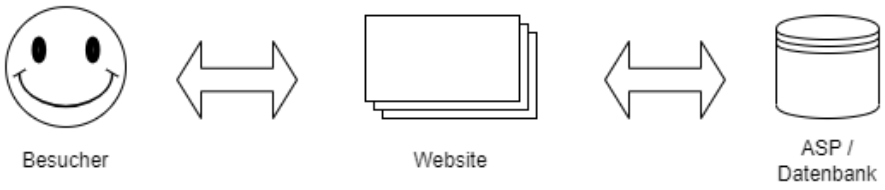

Abbildung 2.1: Page Tagging Modell (vgl. [Ade10])

2.1.3 Ereignis-Tracking

Unter der Voraussetzung der Nutzung des zuvor erwähnten Page Tagging ergibt sich die Möglichkeit zur Messung der Ereignisse, die nur im Browser auftreten und von herkömmlichen HTML-Seitenaufrufen abweichen. Mithilfe des Ereignis-Trackings lassen sich die Interaktionen der Besucher in Bezug auf bestimmte Ereignisse mit komplexeren Elementen und Funktionen auf der Seite messen (vgl. [Hal10]). Dabei stellt sich die Frage, welche Ereignisse und Parameter für die Untersuchung relevant sind und welche Schlussfolgerungen daraus gezogen werden können (vgl. [Kai10]). Im Gegensatz zu bisherigen Verfahren erfolgt diese Methodik objektorientiert. Das bedeutet, dass die einzelnen Komponenten der Seite separat betrachtet und die darauf vorgenommenen

Änderungen beobachtet werden. Die Umsetzung erfolgt durch das Hinzufügen von definierten Parametern an die Stelle des Objekts im Quellcode. (vgl. [Ade10])

2.2 Arten von Daten

Unter dem Begriff "Daten" werden die logischen Sammlungen von den Informationseinheiten zur Verarbeitung verstanden (vgl. [Sta99]). Diese Daten können je nach der Struktur und Zweckbindung in den folgenden Arten unterteilt werden:

2.2.1 Daten nach Merkmal

Die Datenart könnte abhängig von dem Merkmal kategorisiert werden. Dabei lassen sich verschiedene Typen unterscheiden. *Diskrete* Daten sind die natürlichen Zahlen und somit abzählbar wie beispielweise die Kinderzahl. Wenn es nur zwei Ausprägungen gibt, dann ist die Rede von einer diskreten und *dichotomen* Variable, wie etwa dem Geschlecht. Im Gegensatz dazu sind *stetige Daten* rationale Zahlen, die jeden Zwischenwert aus dem Zahlbereich annehmen können, beispielsweise die Zeit in Sekunden. *Quantitative* Daten sind messbare Größen, die als numerische Werte dargestellt werden können. Diese Variablen können entweder diskret oder stetig sein, wie beispielsweise das Alter. *Qualitative* Variablen sind kategorisierbare Daten, deren Werte eine begrenzte Anzahl von Kategorien annehmen und nur begrifflich zu unterscheiden sind, wie zum Beispiel die Wohnsituation. (vgl. [Sch15])

2.2.2 Daten nach Struktur

Eine weitere alternative Klassifikation der Daten besteht in der Unterscheidung der Daten von deren Organisation.

unstrukturierte Daten sind Daten ohne konkrete Aufbau Muster. Beispielsweise sind Text- oder Bilddaten.

strukturierte Daten sind hingegen diejenigen, die eine erkennbare Struktur aufweisen. Dazu gehören die Tabellen, wobei sowohl die zeilenweise als auch spaltenweise Zugehörigkeit der Werten zu den Attributen auffindbar ist (vgl. [Pap19]).

2.3 Metriken und Skalen

Bei der Erhebung von Daten und der Speicherung der Ergebnisse ist es wichtig, die passenden Merkmale festzulegen und diese als ein Ergebnis zu gruppieren, um eine weitere Analyse aufrechtzuerhalten. In der nächsten Unterkapiteln werden die herkömmlichen Skalen erläutert, die eine Zusammenfassung der Ergebnisse in den Gruppen ermöglichen und die Metriken, um die einzelnen Werte zu messen.

2.3.1 Skalen

Die Messwerte von Daten verfügen über verschiedene Eigenschaften, die eine Kategorisierung in bestimmten Gruppen ermöglichen. Zu diesem Zweck werden in der Statistik verschiedene Skalen eingesetzt. Die *Nominalskala* ist eine grundlegende Form der Skala, die Gruppierung von Daten durch eine Kategorisierung ermöglicht, wobei die Kategorien eindeutig sein sollten und die gruppierten Daten nicht in einer bestimmten Reihenfolge angeordnet sind. *Ordinalskala* ermöglicht ebenfalls die Gruppierung von Daten in Kategorien, wobei den Daten eine bestimmte Rangfolge zugewiesen ist, jedoch sind die Abstände zwischen den Kategorien nicht messbar oder gleichmäßig verteilt. Im Unterschied dazu verfügen Daten in einer *Intervallskala* über messbare Abstände zwischen den Werten. Die *Ratioskala* besitzt alle Eigenschaften der Intervallskala und verfügt zudem über einen absoluten Nullpunkt, der auf das Fehlen der gemessenen Eigenschaft hinweist. (vgl. [Ste46])

Zusammengefasst stellt die Tabelle eine Übersicht der aufgezählten Skalen mit den Beispielen (siehe Tabelle 2.1):

Skalenniveau	Beispieltyp	Beispiele	Mögliche Rechenoperationen
Nominalskala	Strings, Nummern, Codes	Geschlecht, Telefonnummern	a=b; a!=b
Ordinalskala	Einschätzungen, Summe mehrerer Einzelobjekte	Schmerzintensität, Kundenzufriedenheit	a>b; a<b
Intervallskala	physikalische Messung, Zählung	Temperatur in°C, IQ	d=a-b
Ratioskala	Zählung, Physikalische Messung	Körpergröße, Bewegungsausmaß	q=a/b

Tabelle 2.1: Skalen Vergleich (vgl. [Sch15])

2.3.2 Metriken

Der Begriff “Metrik” stammt aus dem Griechischen und bedeutet “Messung”. Diese quantifizierbaren (beschreibbaren) Größen helfen, objektive Aussagen bei der Analyse zu treffen und können unterteilt werden:

- Die *Anzahl-Metrik* verwendet absolute Werte wie Einzelzahlen, Durchschnittswerte, Summen oder Differenzen (vgl. [Has13]).
- Die *Verhältnis-Metrik* stellt eine Relation zwischen den Anzahl-Metriken her und wird oft mit dem Begriff “pro” gekennzeichnet. Das Ergebnis kann als Prozentsatz oder Quotient ausgedrückt werden. Verhältnis-Metriken kombinieren verschiedene Messgrößen zu einer Kennzahl und sind widerstandsfähig gegen Messungenauigkeiten. Ein Beispiel für eine Verhältnis-Metrik ist “Seitenaufrufe pro Besuche”. (vgl. [Has13])
- *KPIs* (Key Performance Indicators) sind Kennzahlen, die bei der Messung der Zielerreichung und des Erfüllungsgrades behilflich sind (vgl. [Zum12]).
- *Webmetriken* sind Messgrößen von Webseiten, die aussagekräftige Informationen über die Nutzung und den Erfolg der Webseite liefern können (vgl. [Zum12]).

2.4 Analyse und deren Arten

Die gesammelten Daten könnten je nach dem Gebiet und Vorgehen auf verschiedene Arten bewertet werden. Die Analyse könnte nach Datenquellen, Hilfsmitteln und Zielen kategorisiert werden. Für diese Arbeit sind die folgenden Analyseverfahren aus diesen Perspektiven relevant:

2.4.1 Explorative Analyse

Die explorative Datenanalyse (kurz EDA) umfasst eine Reihe von Techniken aus der deskriptiven Statistik und grafischen Verfahren, um die Werte besser zu verstehen. Mit ihrer Hilfe werden Muster untersucht und Anomalien oder Besonderheiten erkannt. Die Grafik dient als Hauptwerkzeug zur Erkenntnisgewinnung. (vgl. [Fah11]) Je nach den Zielen der Analyse werden verschiedene EDA-Techniken eingesetzt (vgl. [Kom16]). (siehe Tabelle 2.2)

Ziel	EDA-Technik
Betrachtung der Verteilung von einer Variable	Histogramm
Ausreißer finden	Histogramm, Streudiagramm, Box-Whisker-Plot
Beschreibung von der Abhängigkeit zwischen einem Expositionsvaribale und einem Ergebnisvariable	2D Streudiagramm +/ Ausgleichungsrechnung
Visualisierung der Abhängigkeit zwischen zwei Expositionsvariablen und einer Ergebnisvariable	Heatmap
Visualisierung von hochdimensionalen Daten	t-SNE oder PCA + 2D/3D Streudiagramm

Tabelle 2.2: EDA-Techniken nach Ziel (vgl. [Kom16])

2.4.2 Webanalyse

Die Webanalyse ist die Art der Analyse, die entlang des Medientyps, in diesem Fall Webkanälen, stattfindet (vgl. [Has13]). Die dafür verwendeten Datensammlungsmethoden wurden im Abschnitt 2.1 erläutert.

Die Webanalyse beschränkt sich jedoch nicht nur auf die Datensammlung und die Erstellung von Berichten, sondern folgt einem zyklischen Prozess zur kontinuierlichen Verbesserung, wie in Abbildung 2.2 verdeutlicht (vgl. [Wai09]).

Abbildung 2.2: Webanalyse Prozess (vgl. [Wai09])

2.4.3 Multikriterielle Entscheidungsanalyse

Die multikriterielle Entscheidungsanalyse (englisch: Multicriteria Decision Analysis, kurz MCDA) ist, wie aus der Bezeichnung ersichtlich, eine Analyse, die bei der Bewertung von Daten mehrere Kriterien berücksichtigt und Unterstützung bei der Entscheidungsfindung bietet.

Die Methodik von MCDA ist aus der Notwendigkeit entstanden, die menschliche Schwierigkeit bei der Berücksichtigung verschiedener Informationsquellen in der Analyse zu bewältigen. Das Hauptziel besteht darin, Alternativen auf der Grundlage mehrerer

Kriterien auszuwerten und auszuwählen, um den Entscheidungsprozess systematisch und strukturiert zu gestalten. (vgl. [Kik05])

Im nächsten Abschnitt werden die Methoden und Hilfsmittel, wie DSS (Entscheidungsunterstützungssysteme) 2.5.1 und AHP (Analytische-Hierarchie-Prozess) 2.5.3, die bei der MCDA eingesetzt werden, genauer erläutert.

2.5 Entscheidungstheorie

Der Entscheidungsprozess im Sinne der Entscheidungstheorie wird als ein Prozess verstanden, bei dem aus mehreren Alternativen ausgewählt wird. Die Entscheidungstheorie selbst kann je nach Forschungsziel unterschieden werden. Die *deskriptive* Entscheidungstheorie beschäftigt sich mit der Erklärung der getroffenen Entscheidungen und sucht nach empirischen Hypothesen, die Vorhersagen über Entscheidungen unter gleichen Umständen ermöglichen könnten. Die *präskriptive* (oder normative) Entscheidungstheorie hingegen befasst sich mit der Ermittlung rationaler Begründungen und Empfehlungen für Entscheidungsprobleme, die von konkreten Fällen abstrahiert werden. (vgl. [Lau18])

2.5.1 Entscheidungsunterstützungssysteme

Im Rahmen der präskriptiven Entscheidungstheorie werden sogenannte Entscheidungsunterstützungssysteme (englisch: Decision Support System, abgekürzt DSS) zur Hilfe entwickelt (vgl. [Mei15]). Diese Systeme verfügen über charakteristische Eigenschaften, die wie folgt formuliert werden können (vgl. [Sto82]):

- Speziell entwickelt, um den Entscheidungsprozess zu erleichtern
- Automatisieren den Entscheidungsprozess nicht, sondern bieten Unterstützung dabei
- Bieten die Möglichkeit zur schnellen Anpassung, abhängig von den Änderungsanforderungen der Entscheidungsträger

Die Entscheidungsunterstützungssysteme können in fünf Kategorien unterteilt werden und werden wie folgt beschrieben (vgl. [Pow02]):

- **Datengetriebene DSS** unterstützt die Analyse der strukturierten Daten. Ein solches System kann Teil des Berichts-, Analysesystems oder des Data Warehousing sein. Zudem bietet das System die wesentliche Unterstützung und umfassende Funktionalität für große Mengen historischer Daten.

- **Modellgetriebenes DSS** bietet Unterstützung für den Zugang und die Bearbeitung von Modellen, einschließlich Optimierungs-, Simulations- und Statistikmodellen
- **Wissensgetriebenes DSS** basiert auf Spezialwissen in einem Bereich und bietet Experten Vorschläge und Empfehlungen bei der Problemlösung
- **Dokumentgetriebenes DSS** befasst sich mit der Sammlung und Verwaltung unstrukturierter Daten wie Textdokumente, Videos und Bilder
- **Kommunikationsgetriebenes DSS** zielt darauf ab, die Gruppenarbeit zu organisieren und kann in Form von Online-Kommunikationskanälen, Aufgabenverteilung und Tools zum Teilen von Dokumenten umgesetzt werden

2.5.2 Objekt- und Metaphase im Entscheidungsprozess

Die Entscheidungstheorie kann in zwei Phasen: Metaphase und Objektphase unterteilt werden.

Die Metaphase befasst sich mit der Vorbereitung, unabhängig von den Methoden und dem Modell, und besteht aus folgenden Schritten, die am Anfang oder am Ende des Entscheidungsprozesses auftreten können (vgl. [Mei15]):

- Definition des Problems
- Festlegung der Ziele und Kriterien
- Suche nach Alternativen
- Umsetzungsphase
- Kontrolle und Feedback

In dieser Phase kann ein DSS keine Hilfe leisten, da die Arbeit für die aufgezählten Schritte nicht automatisiert werden kann, und somit existiert kein Entscheidungsunterstützungssystem für die Metaphase.

Die Objektphase konzentriert sich gezielt auf die Formulierung eines Entscheidungsmodells und dessen Lösung. Hierbei unterstützt ein DSS den Entscheidungsprozess und hilft dabei (vgl. [Mei15]):

- Kriterien zu gewichten
- Alternativen zu suchen

- Optimale Lösung abzuleiten
- Stabilität und Akzeptanz der optimalen Lösung zu bewerten

2.5.3 Der Analytische-Hierarchie-Prozess

Zur Entscheidungsunterstützung wird der analytische Hierarchieprozess (gekürzt AHP), der von Thomas L. Saaty in den 1980 Jahren entwickelt wurde, eingesetzt. AHP stellt eine Entscheidungsmodell dar, das im Entscheidungsprozess dazu beiträgt (vgl. [Saa88]):

- Eine visuelle Darstellung eines komplexen Problems zu entwickeln
- Prioritäten zu messen und zwischen den Alternativen auszuwählen
- Konfliktlösungen zu analysieren
- Vorhersagen zu treffen
- Eine Kosten-Nutzen-Analyse durchzuführen

Der AHP-Prozess erfolgt in zwei Phasen des Entscheidungsprozesses gemäß der präskriptiven Entscheidungstheorie. Die Ausführung des AHPs findet in der Objektphase statt. Wie zuvor beschrieben, stellt die Metaphase die notwendige Vorbereitung zur Generierung des Entscheidungsmodells dar. (vgl. [Mec17])

Der Hierarchieaufbau kann wie in Abbildung 2.3 dargestellt sein. In dieser hierarchischen Struktur werden Alternativen prozessual nach Kriterien bewertet und ein Ergebnis herausgeführt. Der gesamte Ablauf des AHP-Prozesses ist in Abbildung 2.4 widerspiegelt.

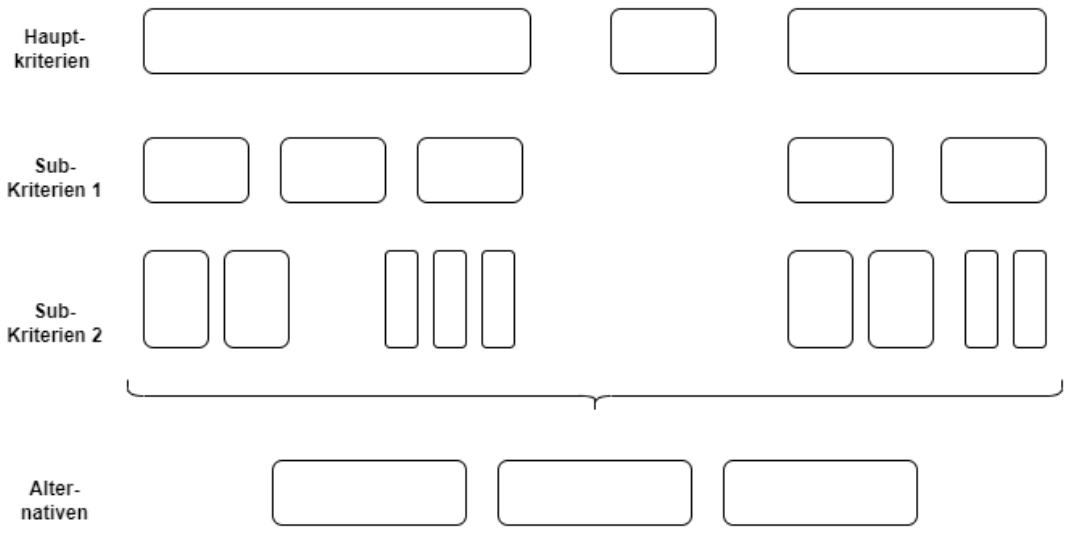

Abbildung 2.3: Hierarchie des AHP (vgl. [Mei15])

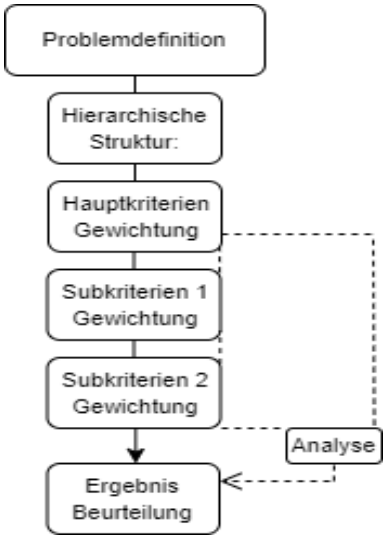

Abbildung 2.4: AHP-Ablauf (vgl. [Mec17])

2.6 Datenschutz

Datenschutz gilt seit dem Volkszählungsurteil von 1983 als ein Grundrecht und nimmt zunehmend an Bedeutung in der Zeit der automatisierten Datenverarbeitung zu. Dies stellt den Schutz des Persönlichkeitsrechts vor Benachteiligung bei der Verarbeitung personenbezogener Daten sicher. (vgl. [Wit08])

2.6.1 Personenbezogene Daten

Der Wirkungsbereich der Datenschutzgesetze betrifft die Sammlung und Speicherung von Daten, die als "personenbezogen" gelten. Dazu gehören Daten, die eindeutig einer Person zugeordnet werden können. Im Gegensatz dazu sind anonymisierte und unpersönliche Daten grundsätzlich nicht betroffen (vgl. [Has13]). Eine Person wird als bestimmbar angesehen, wenn ihre Identität "direkt oder indirekt, insbesondere durch Zuordnung zu einer Kennnummer oder zu einem oder mehreren spezifischen Elementen identifiziert werden kann" (vgl. [Hä08]).

2.6.2 Datenverarbeitung

Ein weiterer wichtiger Begriff im Zusammenhang mit dem Datenschutz ist die Datenverarbeitung. Gemäß der Datenschutz-Grundverordnung (DSGVO) umfasst die Verarbeitung von Daten einen Lebenszyklus, der vom Auslesen bis zum Löschen reicht (vgl. [Kne21a]). (siehe Abbildung 2.5)

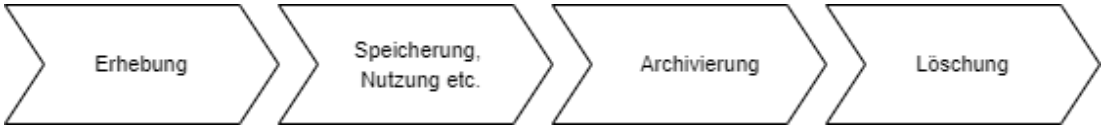

Abbildung 2.5: Typischer Lebenszyklus von (personenbezogenen) Daten (vgl. [Kne21a])

2.6.3 Verbot mit Erlaubnisvorbehalt

Grundsätzlich gilt das sogenannte präventive Verbot mit Erlaubnisvorbehalt, was bedeutet, dass die Nutzung von personenbezogenen Daten untersagt ist, es sei denn, sie ist ausdrücklich erlaubt (vgl. [Gau22]).

Die Erlaubnis liegt vor, wenn sie in den Rechtsgrundlagen erlaubt oder angeordnet ist. Die Rechtsgrundlage ergibt sich aus der Datenschutz-Grundverordnung (DSGVO) und erfordert die Notwendigkeit der Daten und Datenverarbeitung sowie die Einhaltung der Zweckbindung und die Richtigkeit der Daten (vgl. [Gau22]). Zur Datenverarbeitung ist keine behördliche Erlaubnis erforderlich, sondern es müssen gesetzliche Voraussetzungen beachtet werden. Das bedeutet, dass die Verarbeitung personenbezogener Daten nur in dem Maße erlaubt ist, wie es die gesetzlichen Bestimmungen vorsehen, oder wenn eine aktive Einwilligung der betroffenen Personen vorliegt (vgl. [Pet22]).

2.6.4 Identifikation

Die Identifizierung einer Person kann durch verschiedene Merkmale erfolgen (vgl. [GD14]):

- *Direkte Identifikatoren* sind Daten, die eine unmittelbare Identifizierung einer Person ermöglichen, wie beispielsweise der Name, die Handynummer oder die Ausweisnummer.
- *Quasi-Identifikatoren* sind Daten, die in Kombination mit anderen Informationen die Identifizierung einer Person ermöglichen, wie beispielsweise das Geburtsdatum oder die Postleitzahl.
- *Sensitive Attribute* sind Daten, deren Offenlegung in Verbindung mit einer Person unerwünscht ist, wie beispielsweise Informationen über Krankheiten oder Vorstrafen

2.6.5 Anonymisierung

Das Gegenteil von personenbezogenen Daten ist anonyme Informationen, die entweder keinen Bezug zu identifizierbaren Personen haben oder auf nicht identifizierte Personen verweisen. Ein spezieller Fall sind anonymisierte Daten, die zuvor einen Personenbezug hatten, aber durch Anonymisierungsverfahren so verändert wurden, dass sie nicht mehr auf individuelle Personen zurückverfolgt werden können. (vgl. [Kne21a])

Abb.2.6 zeigt den Unterschied zwischen den beiden Ansichten zu den Daten. Aus juristischer Sicht ist die Trennung zwischen personenbezogenen und anonymen Informationen eindeutig. In der praktischen Umsetzung sind die Grenzen zwischen den beiden Begriffen jedoch nicht strikt definiert. (vgl. [Kne21a])

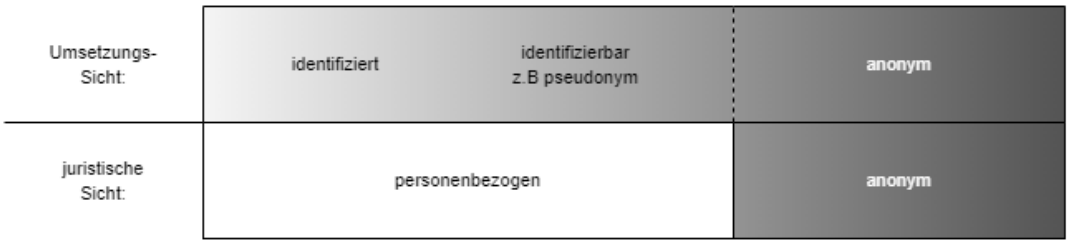

Abbildung 2.6: Stufen der Identifizierbarkeit (vgl. [Kne21a])

Mögliche Verfahren, um den Zugriff Unbefugter auf personenbezogene Daten während der Verarbeitung zu erschweren oder auszuschließen, umfassen:

Informationsreduktion

Bei der Reduktion werden Teile der Informationen, die personenbezogen sind, entfernt. Dieses Verfahren soll die direkte Identifizierbarkeit verhindern. Dennoch besteht die Möglichkeit und das Risiko, dass allein durch eine geringe Menge zusätzlicher Informationen die Identifikation möglich wird. (vgl. [Kne21a])

Pseudoanonymisierung

Gemäß der zuvor gezeigten Abbildung (siehe Abbildung 2.6) gehört pseudonyme Information nicht direkt zu den anonymen Informationen, sondern immer noch zu personenbezogenen.

Bei dieser Methode wird der Personenbezug durch ein Pseudonym ersetzt und gespeichert, um die Zuordnung für Berechtigte aufrechtzuerhalten (vgl. [Gau22]). Dies hinterlässt

jedoch eine potenzielle Sicherheitslücke, durch die Unbefugte die Person anhand des Pseudonyms identifizieren können, wenn sie zusätzliche Informationen erhalten. Ein solcher Fall wurde beispielsweise anhand der Deanonymisierung von Netflix-Nutzern mit nur geringfügigen Abonnentendaten nachgewiesen (vgl. [Nar08]).

Generalisierung und Unterdrückung der Daten

Durch die Generalisierung der Daten können viele Informationen beibehalten werden, während gleichzeitig die Identifikatoren verallgemeinert werden. Dies erfolgt beispielsweise durch die Generalisierung einzelner Attribute personenbezogener Daten, wie dem Geburtsdatum. Das Löschen von Datenzeilen in der Datensammlung wird als Unterdrückung bezeichnet und eignet sich zur Anonymisierung von weniger Subjekten. Dabei werden Quasi-Identifikatoren gelöscht, die nur einmal in den Daten vorkommen. (vgl. [Pet22])

In beiden Fällen wird die sogenannte k-Anonymität erreicht, wobei die Anonymisierung mit steigendem Wert von k zunimmt. Die k-Anonymität liegt vor, wenn in der Datentabelle mindestens k Datensätze (Zeilen) für die Quasi-Identifikatoren die gleichen Werte aufweisen. (vgl. [Swe02])

Name	Geburtsjahr	PLZ	Geschlecht	Diagnose
John Doe	1982	33098	Männlich	Migräne
Thomas Muster	1982	33098	Männlich	Erkältung
Max Maier	1983	33098	Männlich	Rheuma
Otto Normal	1983	33098	Männlich	Depression
Jane Doe	1985	33100	Weiblich	Heuschnupfen
Lieschen Müller	1985	33100	Weiblich	Hypochondrie
Erika Musterfrau	1983	33098	Weiblich	Übergewicht
Jane Average	1983	33098	Weiblich	Migräne

Tabelle 2.3: Daten vor der Generalisierung (vgl. [Pet22])

Geburtsjahr	PLZ	Geschlecht	Diagnose
1982–1983	33098	Männlich	Migräne
1982–1983	33098	Männlich	Erkältung
1982–1983	33098	Männlich	Rheuma
1982–1983	33098	Männlich	Depression
1983–1985	33*	Weiblich	Heuschnupfen
1983–1985	33*	Weiblich	Hypochondrie
1983–1985	33*	Weiblich	Übergewicht
1983–1985	33*	Weiblich	Migräne

Tabelle 2.4: Ergebnis der Generalisierung (vgl. [Pet22])

In der Tabelle 2.3 werden die Daten dargestellt, die k-Anonymität mit k = 4 aufweisen. In den vier Zeilen mit Quasi-Identifikatoren (Geburtsdatum, PLZ, Geschlecht) treten die gleichen Werte auf. Ausgehend von diesen Daten könnte die Generalisierung eingesetzt werden, um die k-Anonymität zu erreichen. Im ersten Schritt sollten die Identifikatoren wie Namen gelöscht werden. In dem nächsten Schritt werden die Werte bei den Quasi-Identifikatoren generalisiert. In diesem Beispiel wird das Attribut mit dem Geschlecht beibehalten und die anderen beiden werden generalisiert. Zusammenfassend sieht das Ergebnis nach der Generalisierung wie in der Tabelle 2.4 aus.

3 Konzept

In diesem Kapitel wird das Konzept des Analysesystems erarbeitet und vorgestellt

3.1 Systemarchitektur

Basierend auf der Definition von Entscheidungsunterstützungssystemen (DSS) im Abschnitt 2.5.1 strebt diese Arbeit danach, ein **datengetriebenes** und gleichzeitig **modellgetriebenes** DSS zu konzipieren. Dabei werden die Daten als Quellen betrachtet, und zur Unterstützung von Entscheidungstreffen für die Optimierung der Anwendung wird Visualisierung als Teil des modellierten DSS angeboten. Ein solches System ermöglicht es, datengetrieben eine Vielzahl historischer Daten zu sammeln und diese grafisch zu modellieren. (vgl. [Pow02])

Im Buch vom Allhou wurden die Schritte zusammengefasst, die bei der Entwicklung eines Analysesystems relevant sind, um das System effizient zu gestalten. Dies umfasst das *Sammeln* von Anforderungen und Daten, das *Aggregieren* der Daten gemäß den Anforderungen, die *Segmentierung* der Daten in logische Teile für einen tieferen Kontext, die *Integration* von Daten aus verschiedenen Quellen und Systemen sowie die *Visualisierung* der Daten für eine klare Vermittlung der Analyseergebnisse und eine erste explorative Datenanalyse (EDA). Darüber hinaus gehört die *Interpretation* der Kausalität hinter den Daten dazu, um deren Bedeutung und Möglichkeiten zu erfahren. (vgl. [Alh16])

Diese Schritte zur Implementierung des Systems können ein generelles Modell des Bottom-Up Ansatzes repräsentieren, der vorschlägt, Teile auf niedriger Ebene zu erstellen, um sie später zu integrieren und so ein großes Teil zu erstellen. Ein solcher Ansatz minimiert Änderungen in der Informationsdarstellung und sollte daher den Anforderungen entsprechen, die von den Entscheidungsträgern für das System gestellt werden. (vgl. [Dah22])

Im Bereich Data Warehouse wird das System grundsätzlich in drei Schichten unterteilt, um ein Framework zum Entscheidungsprozess aufzubauen (vgl. [Nan19], [Sit17]) In der *unteren Schicht* erfolgt die Datenerhebung von verschiedenen Datenquellen. Die *mittlere*

Schicht übernimmt die Speicherung der Daten. Schließlich bildet die *obere Schicht* eine Benutzeroberfläche, über die Nutzer auf die gespeicherten Daten zugreifen und die Analyse durchführen können.

Bei der Entwicklung eines Entscheidungsunterstützungssystems für den Vertrieb wurde anhand des Bottom-up-Ansatzes und Schichten die Architektur des entwickelten Systems erst formuliert und später von Dahr realisiert (vgl. [Dah22]). In dieser Arbeit wird dieser Ansatz zur Betrachtung des Systems übernommen. Die Schritte zur Entwicklung des Analysesystems von Alhlou werden in Schichten aufgeteilt und von unten nach oben betrachtet. Die Datensammlung, als erste Phase, bildet die Datenschicht. Dieser Schritt entspricht dem Sammeln von Anforderungen und Daten. Die nächste Ebene ist die Logikebene, in der Daten transformiert werden, einschließlich Aggregation, Segmentierung und Integration, was den Schritten im Buch von Allhou entspricht. Schließlich repräsentiert die oberste Ebene die Visualisierungsschicht, in der die Visualisierung und Interpretation der Daten stattfinden. Die vorläufige Systemarchitektur ist in Abbildung 3.1 dargestellt, wobei jede Schicht ihre spezifischen Funktionen in der Datenverarbeitung und Entscheidungsunterstützung erfüllt.

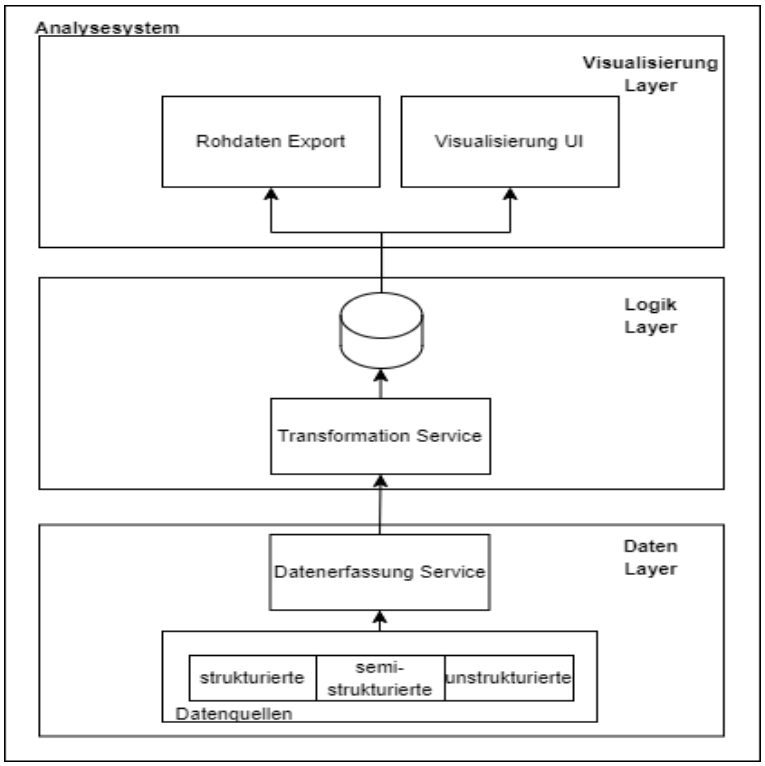

Abbildung 3.1: Vorläufige Systemarchitektur

3.2 Bestimmung der erfassten Metriken

In der Literaturübersicht im Bereich der Webanalyse hat Palomino die Gruppen von Metriken, die in anderen Arbeiten zur Messung des Nutzererlebnisses eingesetzt wurden, in drei Kategorien zusammengefasst (vgl. [Pal21]):

- Generische Metriken: Dies sind grundlegende Metriken, die in den meisten Analysesystemen gefunden werden.
- Klickbasierte Metriken: Diese können in Tools gefunden werden, die den Schwerpunkt auf Click-Analyse legen.
- Ereignisbasierte Metriken: Hierbei handelt es sich um Metriken, die speziell für Forschungszwecke formuliert wurden. Diese sind nicht in Drittanbieterlösungen zu finden.

Die Auswahl der genannten Metriken kann je nach Ziel variieren. Wenn allgemeines Wissen über das Nutzerverhalten gewünscht ist, reichen die generischen Metriken vollkommen aus (vgl. [Pal21]). Laut Kirsch sind jedoch nur die generischen Metriken, die auf Seitenstufe Daten erheben, ineffektiv, wenn es um die Optimierung und Verbesserung der Anwendung geht (vgl. [Kir20]). Zu diesen Zwecken werden Metriken benötigt, die feinere Details auf der Unterseitebene liefern können. Für diese vertiefenden Einblicke sind beide Kategorien von Metriken, wie klick- und ereignisbasierte, passend (vgl. [Pal21]).

In der folgenden Tabelle 3.1 sind mögliche Metriken je nach Kategorie aufgeführt:

Kategorie	Beispiele
generische Metriken	Ladungszeit, Anzahl der Besucher, Anzahl der Sitzungen, Dauer der Sitzung
klickbasierte Metriken	Anzahl der Klicks, Anzahl der Klicks pro Link
ereignisbasierte Metriken	Gebrauchte Zeit beim Registrieren, meist verwendetes Element (bekommt mehr Klicks), Rate der Interaktion des Nutzers mit anderen Elementen

Tabelle 3.1: Beispiele der Metriken nach Kategorie [Pal21]

3.3 Auswahl der Datenerfassungsmethoden

Obwohl es die Studien über die Datensammlung mit Open-Source Webanalysetools gibt, ist die Anzahl der akademischen Publikationen, die die Funktionsweise der Datenerfassungsmethoden beleuchten, so gut wie nicht vorhanden (vgl. [Can23]). Aus der Untersuchung nach einem generellen Vorgehen, um die optimale Datenerfassung und weitere Analyse zu ermöglichen, stellte Kitchens auch einen Mangel an Forschungen und wissenschaftlichen Arbeiten fest, die solches Modell untersucht oder entwickelt haben (vgl. [Kit18]).

Um trotzdem die Forschungsfrage FF1 zu beantworten, wird in dieser Arbeit eine hybride Methodik angeboten und ausgearbeitet, indem man wie in Big Data Analyse verschiedenen Arten der Daten sammelt. Hybride Datenerfassung ermöglicht eine genauere Statistik und Analyse durchzuführen (vgl. [Jyo17]). Beispielweise funktioniert die Kombination der Methoden wie Page Tagging (bzw. Ereignis Tracking) und Log Dateien in Webbereich besser als die Techniken getrennt (vgl. [Kum20]).

Die drei weiter vorgestellten Methoden zur Datenerfassung stellen sich die Möglichkeit, die verschiedene Arten der Daten zu sammeln. Zudem bietet dieses hybride Vorgehen den Freiraum, um die Präferenz und Schwerpunkt an bestimmte der genannten Methoden zu setzen, falls es nötig oder gewünscht ist.

3.3.1 Ereignis-Tracking

Ereignisbasierte Analyse, manchmal auch “absichtbasierte” Analyse bezeichnet, gewährt tiefere Einblicke in die Nutzung von Webseiten oder Apps durch die Nutzer. Die Verlagerung von der Sitzungs- und Seitenansichts-Tracking Methode zu einer ereignisbasierten Analyse markiert eine bedeutsame Verschiebung im Verständnis des Nutzerverhaltens. Anstelle der Frage “Was sehen die Nutzer” konzentriert sich diese Analyse auf “Was machen die Nutzer”. Dieses Echtzeitwissen trägt zur Anpassung des Benutzererlebnisses an die tatsächlichen Bedürfnisse der Kunden bei. (vgl. [Res21])

Eine genauere Erklärung der Funktionsweise der Methode wurde im Abschnitt 2.1 vorgestellt.

Ein Beispiel aus der Industrie ist Google Analytics (GA4), das erfolgreich die Ereignisbasierte Methodik einsetzt. In der aktuellsten Rezension für GA4 Plattform wird als Schlussfolgerung zusammengefasst, dass dieser Ansatz es ermöglicht, jedes spezifische Nutzerverhalten auf einer detaillierten Ebene zu verfolgen (vgl. [McG23]).

Mit dieser Methode können sowohl klick- als auch ereignisbasierte Metriken, wie im Abschnitt 3.2 vorgestellt, erfasst werden.

Zusammengefasst sind die Vorteile dieser Methode:

- Echtzeit-Adaptivität: Diese Analyse bietet die Möglichkeit, das Benutzererlebnis unmittelbar an die tatsächlichen Bedürfnisse und Handlungen der Kunden anzupassen.
- Fokus auf das Nutzerverhalten: Die Methode konzentriert sich auf die Interaktion zwischen dem Kunden und dem System.
- Gezielte Erfassung: Das Tracking erfolgt auf Mikroebene und ermöglicht eine präzise Auswahl der relevanten Ereignisse, die untersucht werden sollen.

Erfassung beim Ereignis-Tracking

Mithilfe von DOM (Dokument Object Model) Listenern lassen sich die Aktionen des Nutzers bei einem Element auf der Webseite beobachten (vgl. [Alh16]). Obwohl es üblich ist, ein Frontend Framework zu verwenden, das einige davon eine virtuelle DOM zur Verfügung stellt (wie React, Vue), ist die Nutzung von sogenannten Event-Handlern beim Framework anstelle von DOM-Listenern ebenfalls möglich (vgl. [Upp22]) Je nach dem spezifischen Framework werden sich schließlich nur die Benennungen leicht unterscheiden. Die möglichen DOM-Listener / Event-Handler für die Ereignisse könnten folgendermaßen sein:

- onclick (beim Click)
- onmouseover (Schweben über ein Element)
- onmouseout (Schweben weg von einem Element)
- onkeydown (nach dem Drück bestimmter Taste)
- onkeyup (nach Loslassen einer Taste)
- onchange (Bei der Änderung an einem Element)
- onfocus (Element wird fokussiert)
- onblur (Element verliert Fokus)
- onscroll (beim Scrollen)

Generelle Struktur

Zur Ereigniserfassung wird eine generelle Struktur festgelegt, um die benannte Vielzahl an den Event-Handler zu beschreiben (vgl. [Ade10]):

- Kategorie: generelle Beschreibung des Ereignisses (Pflicht)
- Aktion: beschreibt die Tätigkeit innerhalb des Objektes (Pflicht)
- Label: genauere Beschreibung der Tätigkeit (optional)
- Ereignis Wert: Je nach Metrik könnte es z.B die Anzahl von Clicks sein (optional)

Data Layer: Zwischenspeicherung

Die gesammelten Daten könnten vorübergehend im Browser im Data Layer abgelegt werden, um sie zu einem späteren Zeitpunkt zu übermitteln. Der Data Layer stellt sich als JavaScript Objekt oder Array dar (vgl. [Alb23]). Beim Laden einer neuen Seite wird JavaScript im Browser gereinigt, und so wird der Data Layer bei jeder Seite neu angelegt (vgl. [Web15]). Daher ist das Sammeln von Aktionen in kleinen Paketen im Data Layer möglich, solange eine Seite geöffnet ist.

Asynchrone Übermittlung

Sobald die Nutzeraktionen im Data Layer zwischengespeichert werden, besteht die Herausforderung, diese Informationen zu übermitteln. Da der Nutzer die Seite jederzeit schließen kann, können die Daten bei einer POST-Anfrage verloren gehen. Daher ist es wichtig, die Übertragung der Daten asynchron durchzuführen. Dies könnte mittels der **sendBeacon**-Methode umgesetzt werden. Diese Methode ist für die Versendung kleiner Pakete von Analysedaten gedacht und vermeidet die Probleme, die sonst bei anderen Techniken wie XMLHttpRequest auftreten (vgl. [Doc23]).

Da diese Methode nur eine kleine Menge der Daten übermittelt, reicht die einmalige Methode beim Schließen der Seite mit der steigenden Anzahl der gesammelten Ereignis-Objekte im Data Layer nicht aus. Der Speicher-Limit kann vom Browser unterschiedlich sein und beträgt ungefähr 64 KB (vgl. [Wan21a]) Deswegen ist eine intervalle Sendung in der Zwischenzeit der Nutzersitzung notwendig. Dabei sollte am Ende jedes Intervalls auch der Data Layer gereinigt werden. Um die Daten dann wieder pro Nutzer oder Sitzung zu analysieren, könnte eine UUID pro Sitzung / Nutzer behilflich sein. Diese UUID kann in jedem Paket als Metadaten eingeführt werden, um die Daten wieder zu vereinigen und in Form eines Berichts zusammenzufassen.

Die Sendung beim Schließen der Seite könnte mithilfe von verschiedenen Varianten umgesetzt werden. Die Nutzung von *visibilitychange* und *pagehide* hat sich als sicherste Methoden erwiesen, in Bezug auf Stabilität und Plattformunabhängigkeit. *beforeunload* und *unload* bergen die Gefahr von Datenverlust und sind daher nicht die optimale Option dafür. Zudem wird *beforeunload* mit mobilen Geräten nicht unterstützt. (vgl. [Wit23])

Schematische Darstellung

Insgesamt könnte die Datenerfassung beim Ereignis-Tracking wie in der Abbildung 3.2 aussehen.

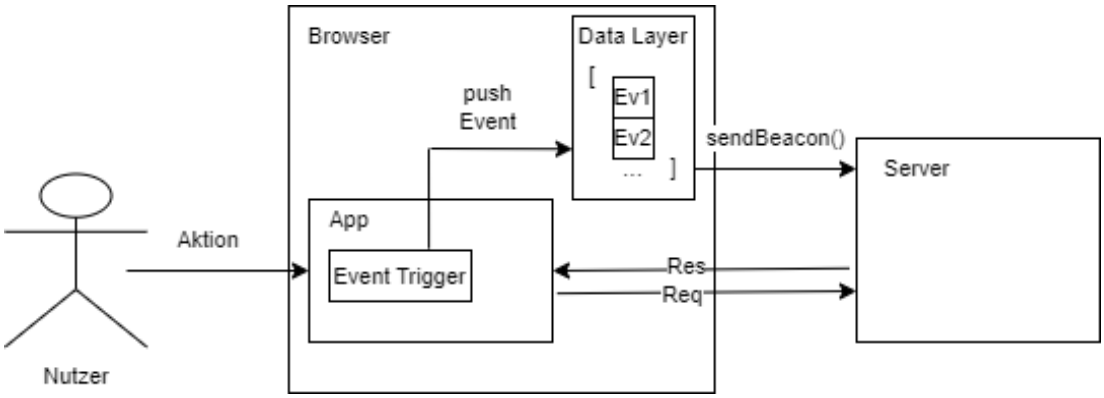

Abbildung 3.2: Ereignis-Tracking Verlauf

3.3.2 Extraktion der Daten

Die Herausforderung, nützliche Daten zu extrahieren, hat das Interesse von Forschern geweckt, die die Datenextraktion in den Bereichen Datenbanken, Mustererkennung, maschinelles Lernen und Datenvisualisierung weiter untersuchen (vgl. [AS19]). Insbesondere im Kontext von Data Warehouse durchlaufen Daten aus verschiedenen Quellen den etablierten Prozess der Extraktion, Transformation und Ladung (ETL), bevor sie in einer einheitlichen Struktur in einer Datenbank gespeichert werden. In der Extraktionsphase werden die Daten erhoben, die später die im geschäftlichen Entscheidungsprozess genutzt werden (vgl. [Nam22]). Somit werden wir die Extraktion weiterhin als eine Methode der Datenerfassung betrachten und sie in das Konzept des Analysesystems integrieren. Mithilfe dieser Methode können die ereignisbasierten Metriken aus Abschnitt 3.2 erhoben werden, die für jede Anwendung und jedes Ziel bei der Entscheidungsfindung spezifisch sein können.

Datenquellen

Als Quellen können strukturierte und semi-strukturierte Daten aus bereits bestehenden Datenbanken oder APIs gewählt werden. Der Vorteil der Datenextraktion aus semi-strukturierten Daten besteht darin, dass solche Extraktion in verschiedenen Bereichen, beispielsweise in Bildung, Werbung oder Wohnungsverwaltung, angewendet werden kann (vgl. [AS19])

In der Fallstudie der Oxfam Firma wurde am Beispiel von Twitter die Datenerfassung der JSON-Darstellung von Tweets mittels APIs vorgestellt. Aufgrund der guten Organisationsstruktur des Formats ermöglicht es zudem, die Daten für die maschinelle Verarbeitung weiterzuverwenden. (vgl. [Sch20])

Transformation

Wie im ETL-Prozess ist die Transformation der Daten ein wichtiger Schritt, um die Daten konsistent und vollständig zu halten. Hierbei werden die Daten bereinigt und gemappt. Duplikate, fehlende Werte, Fehler in den Daten sowie NULL/Not-NULL Werte werden hier korrigiert. Die Gruppierung der Daten kann auch durch Aggregationsfunktionen (SUM, MIN, MAX, AVG) erfolgen. Je nach Wertetyp können wir die Daten anhand von deskriptiven Statistiken zusammenfassen, indem die Daten nach Skalen, wie im Abschnitt 2.3, gemappt werden. (vgl. [Aze19])

Automatisierung

Die Extraktion der Daten kann in Form automatisierter Jobs realisiert werden. Die Automatisierung dieses Prozesses kann Zeit sparen und mögliche Fehler reduzieren. Die geplanten Jobs, die zu bestimmten Zeitpunkten ausgeführt werden, stellen sowohl die Aktualität der Daten als auch die gleichmäßige Verteilung der Auslastung des Systems sicher. Zusammengefasst sieht die Extraktion wie in Abbildung 3.3 aus. (vgl. [See23])

3.3.3 Log Datei Alternative

Für die Sammlung von Nutzerdaten von der Webseite gelten Log Dateien als grundlegendes Konzept. Die Analyse dieser Dateien kann einen Überblick über die Leistung der Seite bieten und Unternehmen dabei unterstützen, effektivere Betriebsstrategien zu entwickeln (vgl. [Wan21b]).

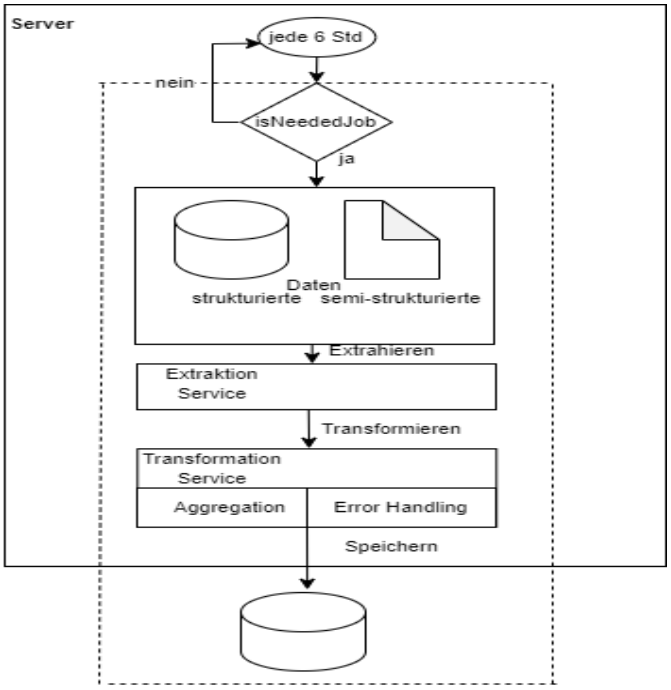

Abbildung 3.3: Das System der Extraktion

In mehreren Forschungen im Bereich Web Mining werden jedoch die Probleme und der hohe Aufwand beschrieben, der mit der Vorverarbeitung der gesammelten Log Dateien verbunden ist. Die gesammelten Dateien enthalten viele unnötige Daten und Rauschen, sodass diese Vorverarbeitungsstufe unvermeidbar ist. (vgl. [Jay17], [Roy20], [Can23])

Als alternative Methode zu den Log Dateien wurde in einem kürzlich erschienenen Artikel die Log-API vorgestellt. Die Nutzung der Log-API ermöglicht es, die Datensammlung intern zu halten und selbst zu kontrollieren, indem die Sammlung serverseitig und als Teil der Software im Hintergrund ausgeführt wird. Dieser Ansatz könnte sowohl für die Webanalyse als auch für das Web Usage Mining (WUM) verwendet werden. Zudem sind die erhaltenen Daten mittels API bereits strukturiert und können so direkt gespeichert werden, was die Notwendigkeit der Vorverarbeitungsstufe vermeidet. Die möglichen Informationen und Quellen umfassen hauptsächlich allgemeine / generische Metriken (siehe Tabelle 3.2), die auch für das Monitoring der Leistung der Webseite genutzt werden können. (vgl. [Can23])

In Abbildung 3.4 wird die Funktionsweise der API dargestellt. Wenn ein Nutzer die Webseite öffnet, wird eine Anfrage an den Server geschickt. In diesem Moment wird die Log-API Anfrage gestartet und die Sitzungsinformationen gesammelt. Anschließend werden die Seiteninformationen versandt. Sobald die Anwendung gestartet ist und ihre Daten gesendet wurden, werden zusätzlich die Ladezeit und andere mögliche Anwendungsdaten von der Seite angefragt:

Daten	Quelle
Benutzeragent (Browser, OS)	HTTP Request
IP-Adresse	Netzwerkprotokoll
vorhergehende Seite	HTTP Request
Geolokalisierung	Extern
Seitenansicht	HTTP Request
Nutzer Profil	Anwendung
Anwendungsspezifische Daten	Anwendung
Sitzung	Anwendung

Tabelle 3.2: Mögliche Erfassungsdaten und deren Quellen (vgl. [Can23])

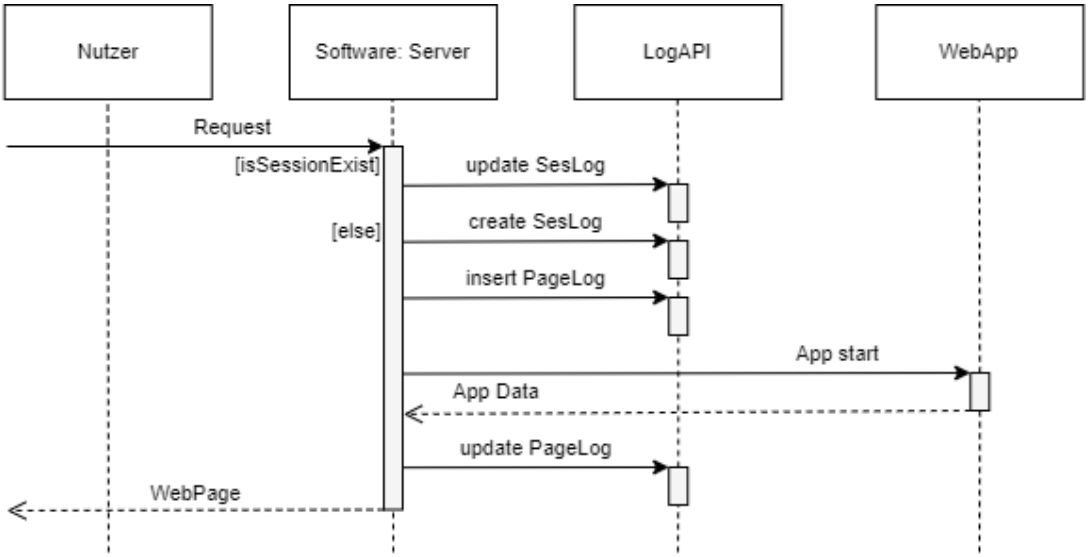

Abbildung 3.4: Sequenzdiagramm der Funktionsweise der LogAPI (vgl. [Can23])

3.4 Datenschutzkonforme Datenverarbeitung

Ein Teil der Nutzer und Unternehmen hat Bedenken bezüglich persönlicher Daten, die bei der Verwendung des Analysesystems gesammelt werden könnten. Der Prozentsatz der Nutzer, die ein Tracking ihrer Daten als unerwünscht betrachten, variiert von 12,5% bis zu 76% (vgl. [Alb23]). Zudem wurde in einer Studie von 2019 festgestellt, dass bereits 92% der Datenerhebung auf der Seite durchgeführt wird, bevor die Nutzer dem wirklich zugestimmt haben (vgl. [SR19]). Die Einhaltung der Datenschutzverordnung und die Anonymisierung der Daten könnten die Akzeptanz der Analysesysteme in der Öffentlichkeit steigern. In diesem Abschnitt wird eine mögliche Vorgehensweise im Umgang mit personenbezogenen Daten vorgestellt, um somit die Forschungsfrage FF2 zu beantworten.

3.4.1 Datenminimierung und Zweckbindung

Laut DSGVO sind Datenminimierung und Zweckbindung grundlegende Prinzipien, die bei der Datenerhebung berücksichtigt werden sollten. Die Menge der gesammelten Daten, die einen Personenbezug haben, sollte grundsätzlich angemessen und beschränkt sein.

Im Hinblick auf die Datenminimierung ist zu beachten, dass auch die vorrätige Speicherung rechtlich untersagt ist. Da Daten in diesem Fall keine konkrete Zweckbindung haben, entspricht eine solche Datenverarbeitung nicht den rechtlichen Grundlagen (vgl. [Amo22]). Somit steht die Datenminimierung im Zusammenhang mit der Zweckbindung, sodass die Datenverarbeitung nur auf die Daten beschränkt ist, die zur Erreichung des verfolgten Ziels dienen (vgl. [Kü20]).

Eine solche Datenverarbeitung entspricht der Rechtsverträglichkeit. Außerdem gewährleistet dieser Ansatz nicht nur die Umsetzung der Grundrechte der Nutzer, sondern bietet auch Vorteile im Vergleich zu anderen internationalen Wettbewerbsprodukten, die regelmäßig in den Datenskandalen auftauchen. (vgl. [Thi20])

3.4.2 Keine Cookies von Drittanbietern

Durch Cookies ermöglicht es die Webseite, den Nutzer wiederzuerkennen und sein Verhalten nachzuvollziehen. Dabei werden die Cookies in zwei Kategorien unterteilt (vgl. [Kam20]):

- First Party Cookies: Auch als HTTP-Cookies bekannt, werden zur Optimierung des Benutzererlebnisses direkt von der Webseite gesetzt. Sie können auch sensitive Daten enthalten, die jedoch meistens vom Nutzer selbst eingegeben werden (außer der IP-Adresse).
- Third Party Cookies: Diese werden von einem Drittanbieter auf der Webseite eingebunden, um den Nutzer domainübergreifend zu verfolgen.

Im Jahr 2017 führte Apple Intelligent Tracking Prevention (ITP) in Safari ein, was dazu führte, dass die Sammlung von First Party Cookies eingeschränkt wurde. Später, im Jahr 2020, wurde auch die Sammlung von Third Party Cookies in Safari blockiert. (vgl. [Alb23]) Andere Browser wie Firefox, Chrome und Microsoft Edge beschränken ebenfalls die Verbreitung von Third Party Cookie Daten (vgl. [Kam20]). Die Tendenz der letzten Jahre zeigt, dass die Verwendung von Drittanbietern Cookies abnimmt und schnell veraltet (vgl. [Ket22]). Stattdessen könnte eine Umstellung auf First Party Cookies und Ereignis-Tracking in Betracht gezogen werden.

Beim Tracking besteht weiterhin die Möglichkeit, sensible Daten zu sammeln, einschließlich solcher, die zur Identifizierung einer Person führen können. Gemäß der Datenschutzgrundverordnung (DSGVO) ist entweder die rechtliche Grundlage oder eine aktive Einwilligung für die Datenverarbeitung erforderlich, und dementsprechend sollten personenbezogene Daten anonymisiert behandelt werden.

3.4.3 Anonymisierungsverfahren

Trotz des Fehlens standardisierter Verfahren zur Anonymisierung personenbezogener Daten hat Kneuper die zu berücksichtigende Schritte zusammengefasst (vgl. [Kne21b]):

- Identifikation der betroffenen Daten und deren geplanter Nutzung
- Initiale Risikobetrachtung
- Basis Anonymisierung
- Erneuerte Risikobetrachtung
- Weitergehende Anonymisierung

Die einzelnen Schritte werden in den weiteren Abschnitten erläutert. Der zusammengefasste Verlauf ist ebenfalls in der Abbildung 3.5 dargestellt.

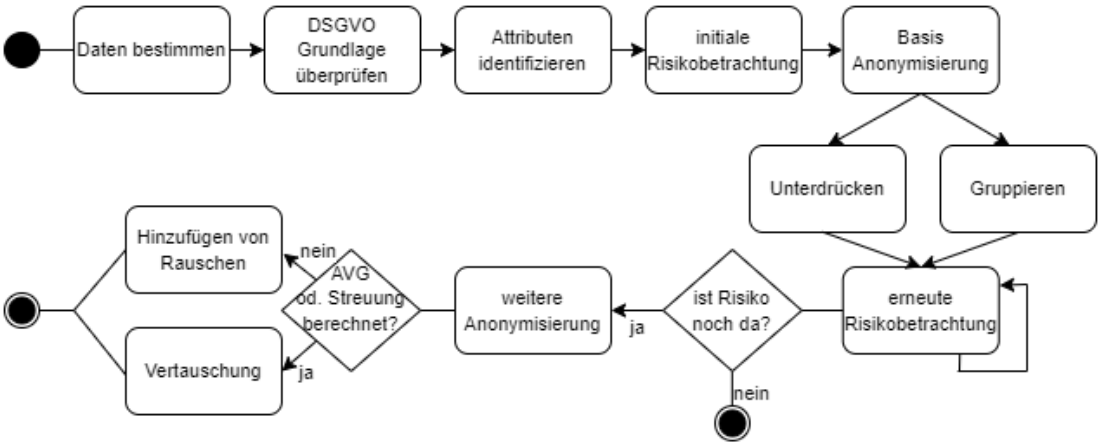

Abbildung 3.5: Aktivitätsdiagramm des Anonymisierungsverfahrens

Identifikation der betroffenen Daten und deren geplanter Nutzung

In diesem Schritt sollten die folgenden Maßnahmen vorgenommen werden (vgl. [Kne21b]):

- Aufklärung über die Art, Zielsetzung, Nutzen und Relevanz der Daten
- Überprüfung der rechtlichen Grundlage für die Verwendung und Veröffentlichung der anonymisierten Daten
- Identifizierung der Arten von Attributen (Identifizierer, Quasi-Identifizierer)

Initiale Risikobetrachtung

Im nächsten Schritt wird das Risiko im Zusammenhang mit möglicher *Schadenshöhe* und *Schadenwahrscheinlichkeit* betrachtet. Bei der *Schadenshöhe* wird von einem ungünstigen Szenario mit großem Schaden ausgegangen, um den Schutzbedarf der Daten einzuschätzen. In der initialen Risikobetrachtung wird die *Schadenwahrscheinlichkeit* zunächst nur durch die Verfügbarkeit der Daten berücksichtigt, da noch keine Maßnahmen zum Datenschutz festgelegt wurden. Verfügbarkeit bezieht sich hier auf den Zugriff auf die Daten. Das Risiko des Missbrauchs steigt proportional zur Öffentlichkeit der Daten. (vgl. [Kne21b])

Nach der Bewertung von Schadenshöhe und Schadenswahrscheinlichkeit kann das Risiko eingestuft werden. In der DSGVO werden Stufen wie “geringes Risiko”, “Risiko” und “hohes Risiko” ohne weitere konkrete Beschreibung unterschieden. Die Einschätzung der Risiken kann mithilfe einer Matrix erfolgen, wobei das Risiko durch Schadenshöhe und Schadenswahrscheinlichkeit für einzelne Fälle bewertet wird. (vgl. [Kre22])

Basis Anonymisierung

Zur Anonymisierung der Daten werden zunächst die grundlegenden Schritte unternommen. Die theoretischen Grundlagen zur Identifizierung und Anonymisierung wurden im Unterkapitel 2.6 erläutert. Die Identifikatoren werden hierbei gelöscht, während die Quasi-Identifikatoren mithilfe von Generalisierung und Unterdrückung angepasst werden.

In der Webanalyse wird besonders versucht, Quasi-Identifikatoren wie die IP-Adresse beizubehalten. Da im Fall der IP-Adresse die Reidentifizierung der Person nicht ausgeschlossen ist, sollten mindestens acht der letzten Stellen gestrichen werden. Auf diese Weise wird die Unterdrückung der Daten durchgeführt. Üblicherweise erfolgt eine Verkürzung auf 24 Bit für IPv4 und auf 40 Bit für IPv6. (vgl. [Luc19])

Erneute Risikobetrachtung

Nach der Durchführung der Basisanonymisierung wird das Risiko der Schadenwahrscheinlichkeit erneut betrachtet. Es wird abgeschätzt, inwiefern der Datensatz eine Person eingrenzt und ob durch zusätzliche Informationen die Anonymität aufgehoben werden kann. Dabei werden besonders die Daten betrachtet, die für die Analyse nicht gelöscht, sondern generalisiert wurden. Diese Überprüfung sollte regelmäßig durchgeführt werden, insbesondere wenn es Änderungen an den Analysemethoden oder den gesammelten Daten gibt. (vgl. [Kne21b])

Weitergehende Anonymisierung

Nach Bedarf wird eine zusätzliche Anonymisierung durchgeführt, falls das Risiko weiterhin besteht. Neben den bereits genannten Maßnahmen wie Generalisierung und Unterdrückung werden Daten vertauscht und Rauschen hinzugefügt. Diese Techniken gehören zu den Randomisierungsmethoden. Die Kombination mit den Generalisierungsmethoden ermöglicht einen stärkeren Schutz der Daten. (vgl. [Kre22])

- *Vertauschung* ist eine geeignete Maßnahme, wenn ein Mittelwert oder eine Streuung über die Spalten berechnet wird. In den Spalten (pro Person) werden die Werte vertauscht.
- *Hinzufügen von Rauschen* bedeutet eine Veränderung der Daten, sodass kleine Fehler integriert werden, um die Identifikation der Einzelperson zu verhindern. Das Endergebnis dabei wird nicht oder nur teilweise beeinträchtigt.

3.5 Datenpersistenz

Da in dieser Arbeit ein eigenes Analysesystem konzipiert wird, sollte auch die Speicherung der Daten selbst durchgeführt werden. Die Auswahl der passenden Datenbankart könnte mithilfe der ACID- und CAP-Theorien unterstützt werden.

Um zwischen SQL- und NoSQL-Datenbanken zu wählen, hilft an dieser Stelle das formulierte CAP-Theorem von Brewer. Es gibt drei Eigenschaften wie Konsistenz, Verfügbarkeit und Partitionstoleranz bei den Datensystemen, wobei eine Datenbank höchstens nur zwei davon besitzen und garantieren könnte (vgl. [Bre12]):

C := Konsistenz entspricht dem Vorhandensein einer einzigen aktuellen Kopie der Daten

A := Hohe Verfügbarkeit dieser Daten für die Änderungen

P := Partitionstoleranz: Toleranz gegenüber Netzwerkteilungen

Relationale Datenbanken ermöglichen normalerweise Konsistenz und Verfügbarkeit, während NoSQL-Datenbanken eine hohe Verfügbarkeit und Partitionstoleranz bieten (vgl. [Vet23]).

Zudem sollten Faktoren wie Menge und Art der Daten bei der Auswahl berücksichtigt werden. Im Bereich Big Data werden viele verschiedene Daten, einschließlich unstrukturierter, erhoben. Wegen der ACID-Eigenschaften (Atomarität, Konsistenz, Isolation, Dauerhaftigkeit) fällt es relationalen Datenbanken schwer, Skalierbarkeit für große Datenmengen anzubieten und mit semi- und unstrukturierten Daten umzugehen (vgl. [Ima20]) Die Komplexität ist dabei hoch, da relationale Datenbanken der Tabellenstruktur folgen, daher müssen die Daten angepasst werden (vgl. [Pal20]).

Die Skalierbarkeit der Daten in NoSQL-Datenbanken ist durch die Trennung der Datenverwaltung und Datenspeicherungsfunktionalitäten gestattet. So ermöglichen diese Datenbanken die Speicherung unstrukturierter Daten und das schnelle Vornehmen von Änderungen ohne Umschreibungen, da das Schema flexibel ist. (vgl. [Ima20])

Im Abschnitt 3.3 wurden drei Datenerhebungsmethoden vorgestellt. Der Ereignis-Tracking sammelt und verschickt die Daten mehrmals pro Sitzung eines Nutzers. Die Menge der gesammelten Daten könnte potenziell groß sein, wenn viele Ereignisse getrackt werden. Zudem ist die Datenmenge ansteigend mit der Anzahl der Sitzungen und Nutzer, die aktiv die Anwendung nutzen. Zudem können nicht alle Ereignisse die gleiche Struktur haben, beispielsweise brauchen nicht alle Ereignisse eine Anzahl von Klicks zu haben. Deswegen sollte hier die Option mit der NoSQL-Datenbank in Betracht gezogen werden.

Für die beiden weiteren Methoden wie Extraktion und Log-API ist der Einsatz von relationalen Datenbanken ebenfalls möglich. Da die Daten in der Extraktionsmethode aggregiert werden, sind sie nach der Aggregation strukturiert. Hiermit ist nur die Menge der Daten unbekannt und könnte mit der Zeit auch wachsen, da die Extraktion von verschiedenen Quellen wie im Big Data Bereich erfolgt. Wie im Abschnitt 3.3 über die Log-API erwähnt wurde, ist der Vorteil dieser Methode, dass bereits strukturierte Daten erhoben werden und keine Anpassung erfordern.

Wir gehen davon aus, dass das Analysesystem nicht nur für eine Anwendung in einem Unternehmen angewendet wird, sondern eine mehrfache Relation ermöglicht. Daher sollten für jede Anwendung aus den drei Datenerhebungsmethoden die Ergebnisse gespeichert und dargestellt werden. Unter Beachtung all dieser Kriterien könnten die folgenden Optionen für die Speicherung in Frage kommen:

- Option 1: Der Einsatz einer NoSQL-Datenbank sowohl für das Endergebnis für eine Anwendung als auch für alle Datenerhebungsmethoden in jeweiligen Kollektionen.
- Option 2: Die Kombination von NoSQL- und SQL-Datenbanken. Für jede Datenerfassungsmethode könnte je nach Bedarf ein eigenes Datenbankmodell für die Zwischenspeicherung angewendet werden. Am Ende könnten die Daten als Endergebnis zu einer Struktur zusammengeführt werden.
- Option 3: Einsatz einer relationalen Datenbank. Hierbei sollte der Verwaltungsaufwand mit der steigenden Anzahl von Daten und möglichen Änderungen an der Datenstruktur berücksichtigt werden.

3.6 Visualisierungsmodell

Die Visualisierung von Daten kann den Wissenserwerb, die Kommunikation und die Argumentation in der Organisation unterstützen. Die Erkennung von Mustern aus der visuellen Darstellung ermöglicht zudem das Erlangen entscheidungsrelevanten Wissens. (vgl. [Lti20])

Um Muster erkennen zu können, ist ein Modell erforderlich, das den Wissenserwerb als Prozess betrachtet und unterstützt. In mehreren Arbeiten könnte eine schematische Beschreibung des Kenntniserwerbs von einem Visualisierungsmodell gefunden werden (vgl. [Sac16] [Kei10]). So wurde auch ein visuelles Modell vorgestellt, wobei die wichtigen Eigenschaften formuliert wurden, die bei einem Erkenntnisprozess für Zuschauer hilfreich sein können (vgl. [Bur05]).

- Aufmerksamkeit
- Kontext
- Übersicht
- Interaktionsmöglichkeiten
- Detaillierte Ansicht

Wie in Abbildung 3.6 gezeigt, werden die Kenntnisse von einem Sender zum Empfänger durch die Visualisierungskomponente vermittelt. Die Visualisierung gliedert sich in drei Phasen. Zuerst sollte die Visualisierung die **Aufmerksamkeit** des Betrachters gewinnen, beispielsweise durch interaktive Dashboards. Anschließend sollte der **Kontext** des Problembereichs mithilfe einer **Übersicht** dargestellt werden, gefolgt von **Handlungsoptionen**. Aus diesen drei Aspekten wird schließlich die **detaillierte Ansicht**

angeboten, auf die der Betrachter fokussieren und interagieren kann, um die Bedeutung hinter den dargestellten Daten zu untersuchen. (vgl.[Lti20])

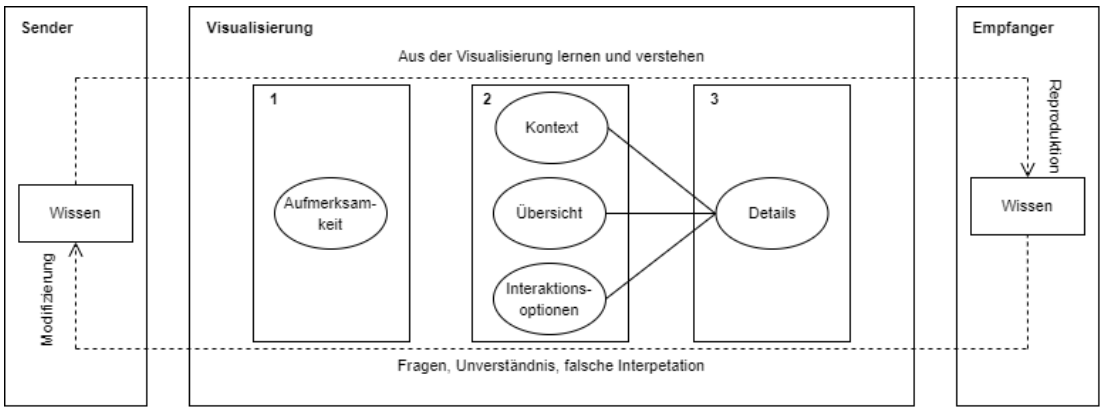

Abbildung 3.6: Visuelles Erkenntnismodell (vgl. [Bur05])

Für das visuelle Modell dieser Arbeit sollten die Wege und Optionen erarbeitet werden, wie die beschriebenen Eigenschaften erreicht werden können, um eine übersichtliche Darstellung zur Unterstützung der Entscheidungsfindung bei der Optimierung der Anwendung sowie des Nutzererlebnisses bereitzustellen.

Miller hat aus der Untersuchung von Arbeiten, die DSS und Visualisierung entwickelt haben, insgesamt 42 Designempfehlungen in drei Kategorien formuliert, nämlich Interface, Information und Interaktion (vgl. [Mil18]).

Eine der Hauptempfehlungen besteht darin, die Präsentation möglichst schlicht und einfach zu halten, sodass nur die wichtigen Elemente auf der Seite erscheinen. Unsere Benutzeroberfläche wird daher eine Übersicht enthalten, in der die aktuellen Kennzahlen zunächst in einer Vorschau angezeigt werden. Bezüglich der Positionierung werden die Kennzahlen gruppiert und lokalisiert, um die Suche auf dem Bildschirm zu erleichtern und die Aufmerksamkeit auf die zentralen Elemente zu lenken (vgl. [Mil18]) Der Kontext zum betrachteten Bereich und zur Anwendung wird durch die Beschriftung der Elemente erreicht.

Benutzer haben die Möglichkeit, von jeder angezeigten Metrik zu einer detaillierten Ansicht zu navigieren. Diese Ansicht wird im Vordergrund ein Diagramm mit dem Verlauf der Daten anzeigen. Der Vergleich der Daten und die Kombination der Darstellungen können beim Verständnis der Daten helfen (vgl. [Lti20]) Hierzu könnte eine tabellarische Darstellung auf derselben Seite angeboten werden. Da unser Analysesystem eine datengetriebene DSS ist und daher historische Daten sammeln und speichern wird, könnte zur Interaktion die Aggregation nach Zeit angeboten werden, um den Verlauf auf täglicher, wöchentlicher, monatlicher und jährlicher Basis anzeigen zu können. Da wir

keine Drittanbieterlösungen im System nutzen, kann auch die Exportierungsmöglichkeit der Rohdaten angeboten werden.

Aufgabe Nummer			
1. Wert abrufen	2. Extremum finden	3. Bereich bestimmen	4. Verteilung charakterisieren
5. Anomalien finden	6. Cluster finden	7. Korrelationen/Trends	8. Vergleiche durchführen

Tabelle 3.3: Visualisierungsaufgaben

Visualisierung	**Datentyp**	**Visualisierungsaufgaben**							
		1	2	3	4	5	6	7	8
Liniendiagramm	Tabelle: eine quantitative Variable, einen geordneten Schlüssel	X	X	X				X	X
Balkendiagramm	Tabelle: eine quantitative Variable, einen kategorisierten Schlüssel	X	X	X					X
Gestapeltes Balkendiagramm	Multidim. Tabelle: einen quantitativen Variable, zwei kategorisierten Schlüssel	X	X	X					X
Kreisdiagramm	Tabelle: eine quantitative Variable, einen kategorisierten Schlüssel	X	X						X
Histogramm	Tabelle: eine quantitative Variable	X	X		X				X
Scatterplot	Tabelle: Zwei quantitative Variable	X	X	X	X	X	X	X	X

weiter auf der nächsten Seite

Flächendiagramm	Tabelle: eine quantitative Variable, einen geordneten Schlüssel	X	X	X				X	X
Gestapeltes Flächendiagramm	eine quantitative Variable, einen kategorisierten Schlüssel, einen geordneten Schlüssel	X	X	X				X	X
Bubble chart	Multidim. Tabelle: Drei quantitative Variablen	X	X	X	X	X	X	X	X
Choropleth Map	Geographische Geometrie mit Tabelle: eine quantitative Variable pro Region	X	X						X
Treemap	Tree und Tabelle: eine quantitative Variable pro Knoten	X	X						X

Tabelle 3.4: Visualisierungsmöglichkeiten anhand von Datentyp und Aufgaben (vgl. [Lee17])

Als nächstes sollten wir uns entscheiden, welche Art des Diagramms zur Anzeige der Daten gewählt wird. Lee (2017) stellte in seinem Artikel "VLAT: Development of a Visualization Literacy Assessment Test"[Lee17] 12 Darstellungsmöglichkeiten vor, abhängig von den Aufgaben und Datentypen. Die Aufgaben sind die Erkenntnisziele, die mit der Visualisierung angestrebt werden. Dabei werden acht Aufgaben unterschieden: Wert abrufen, Extremum finden, Bereich bestimmen, Verteilung charakterisieren, Anomalien finden, Cluster finden, Korrelationen/Trends und Vergleiche durchführen.

Die Datenvisualisierung wird nicht nur durch die Aufgaben unterschieden, sondern auch durch den Datentyp. Zusammengefasst sind die Visualisierungsmöglichkeiten in der Tabelle 3.4 dargestellt, wobei die üblichen Datentypen sowie die Aufgaben gekennzeichnet sind. Die Aufgaben sind durchnummeriert, und die Beschreibung der Aufgaben mit der entsprechenden Nummer ist nochmals in der Tabelle 3.3 dargestellt.

3.7 Entscheidungsprozess

Insgesamt wird das konzipierte Analysesystem als DSS zur Unterstützung im Entscheidungsprozess zur Optimierung und Verbesserung der Anwendung eingesetzt, um die Forschungsfrage FF4 zu beantworten. In dem Hintergrundkapitel wurde die Definition eines Entscheidungsprozesses sowie die Entscheidungstheorie im Abschnitt 2.5.2 erläutert. Basierend auf diesen Kenntnissen wird der Entscheidungsprozess zur Optimierung der Anwendung grundsätzlich in zwei Phasen unterteilt, nämlich Meta- und Objektphase.

In der Metaphase wird die wichtige Vorbereitung vor der Nutzung des Analysesystems vorgenommen. Das Problem sowie die Ziele mit Kriterien werden festgelegt. Nachfolgend erfolgt die Umsetzung des Systems. In der Objektphase wird das Entscheidungssystem zur Gewichtung der Kriterien und Ableitung der optimalen Lösung eingesetzt. Die Nutzung des DSS kann als analytischer Hierarchieprozess (AHP) betrachtet werden. Die Schritte beim AHP sind in der Abbildung 2.4 im Abschnitt 2.5.2 dargestellt. Zusammengefasst lässt sich der gesamte Ablauf wie in der folgenden Abbildung 3.7 darstellen.

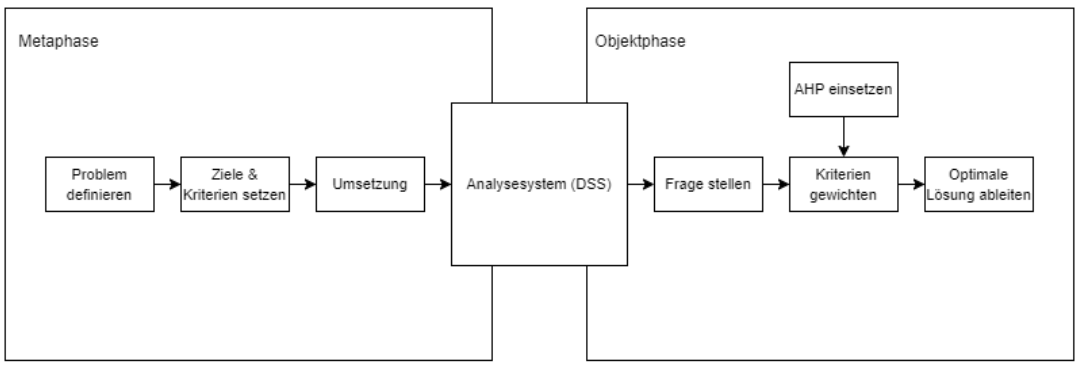

Abbildung 3.7: Entscheidungsprozess schematisch

3.8 Systemüberblick

Aus dem zyklischen Ablauf der Webanalyse aus dem Abschnitt 2.4.2 sowie den vorgestellten Teilen des Systems stellt sich zusammendfassend solches Aktivitätsdiagramm wie in der Abbildung 3.8 vor:

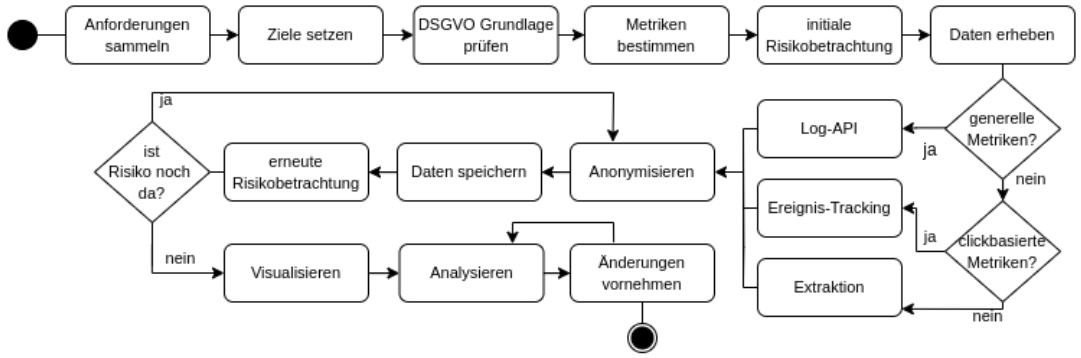

Abbildung 3.8: Aktivitätsdiagramm des Systems

Im Entscheidungsprozess zur Optimierung der Anwendung und datengetriebenen Planung beginnt das System mit der Metaphase, wo die Sammlung der Anforderungen und der anschließenden Zielsetzung für eine Problemstellung erfolgt (siehe 3.7). Die relevanten Daten werden auf DSGVO-Grundlage geprüft, und falls die Grundlage vorhanden ist, werden die Metriken gemäß der Unterteilung im Abschnitt 3.2 bestimmt. Dabei erfolgt eine initiale Risikobetrachtung, wie im Abschnitt 3.4.3 beschrieben. Basierend auf den relevanten Metriken werden die Datenerhebungsmethoden ausgewählt, die am besten für die Erhebung dieser Daten geeignet sind. Wie im Abschnitt 3.3 für die Datenerfassungsmethoden erläutert, ist die Log-API hauptsächlich für generelle Metriken vorgesehen, während das Ereignis-Tracking für clickbasierte Metriken und die Extraktionsmethode für ereignisbasierte Metriken geeignet sind. Vor der Speicherung der Daten werden, falls erforderlich, Anonymisierungsverfahren durchgeführt, wie im Abschnitt 3.4.3 beschrieben. Anschließend erfolgt die Auswahl einer der drei Speicherungsoptionen aus dem Abschnitt 3.5. Nachfolgend wird eine erneute Risikobetrachtung durchgeführt, und bei Zustimmung erfolgt die Auswahl von Visualisierungsmöglichkeiten, die zur Art und Zielsetzung der erfassten Daten passen. Die Auswertung im Rahmen des Entscheidungsprozesses wird durchgeführt, wobei zur Unterstützung AHP als Vorgehensweise eingesetzt werden kann.

Das Aktivitätsdiagramm des Analysesystems dient als Leitfaden für die schrittweise Realisierung und wird im folgenden Kapitel 4 angewendet.

4 Realisierung

Dieses Kapitel befasst sich mit der Umsetzung des ausgearbeiteten Konzepts in der Umgebung eines Unternehmens.

4.1 Janitza electronics GmbH

Die Firma Janitza electronics GmbH ist ein mittelhessisches Unternehmen, das im Bereich Energiemanagement Komplettlösungen von Messgeräten bis zur Software anbietet (vgl. [jan23]). Die Software namens GridVis stellt einen Monolith dar, der verschiedene Dienste zur Verfügung stellt.

4.1.1 Berichtseditor-Anwendung

Eine der Webanwendungen von Janitza ist der sogenannte Berichtseditor, der die Möglichkeit bietet, für ein besseres Energiemanagement Berichte zu erstellen und dabei die gemessenen Werte wie Strom oder Spannung graphisch darzustellen. Die Anwendung verfügt über vielseitige Funktionen, um die Berichte möglichst benutzerdefiniert zu gestalten. Diverse Darstellungsmöglichkeiten werden in Form von Objekten angeboten, die einem Bericht hinzugefügt werden können, um die Werte zu visualisieren. Zum Beispiel kann der Nutzer mit einem Objekt namens Liniendiagramm sowohl historische als auch Echtzeitwerte von einem Gerät auslesen und darstellen lassen.

4.1.2 Definierung der Ziele und Anforderungen für Berichtseditor

Im Fall der Berichtseditor-Anwendung ergaben sich Herausforderungen und Bedarf hinsichtlich des besseren Verständnisses des Nutzerverhaltens mit der Anwendung. Die gewünschten Erkenntnisse wurden ausschließlich aus Kundengesprächen und Nutzertests gewonnen. Für das Produktmanagement ist es entscheidend, das Nutzerverhalten zu verstehen, um die Planung und Entwicklung zukünftiger Anwendungsfunktionen zu unterstützen. Ein Beispiel hierfür ist die Arbeit des Entwicklungsteams an der Erstellung

von Vorlagen für den Berichtseditor. Hierbei wäre ein genaues Verständnis der beliebten und bevorzugten Anwendungs- und Darstellungsfunktionen von Vorteil.

Angesichts der Herausforderungen in der Firma liegt der Hauptfokus bei Entscheidungsfindungen auf dem Nutzerverhalten, also wie die Benutzer mit der Anwendung interagieren. Daher verfolgen wir bei der Umsetzung das Ziel, einen besseren Überblick über die genutzten und weniger genutzten Funktionen zu schaffen.

4.2 Analyse der bestehenden Anwendung

Im Rahmen eines Hackathons wurde in der Firma damals ein Versuch unternommen, eine Schnittstelle zu implementieren, um die Informationen aus dem GridVis periodisch als Ergebnis an einen Server zu senden und zu speichern.

4.2.1 Aktueller Stand des internen Analysesystems

So verfügt GridVis über ein Analytics-Modul, wobei die Klasse `Analytics Manager` die Funktionalität für das Hinzufügen und Speichern der Ergebnisse bereitstellt. Die Ergebnisse werden im String-Format im Key-Value Store zwischengespeichert, für den Fall, dass der GridVis-Server unterbrochen wird. Sobald der Server wieder läuft, werden die Ergebnisse synchronisiert und mit einer POST-Anfrage vom Key-Value Store zum `ID Server` geschickt. An dieser Stelle wird das Ergebnis ausgelesen und in einer Sammlung von dokumentenbasierten Datenbanken gespeichert. Der Datenfluss ist noch einmal in der Abbildung 4.1 dargestellt.

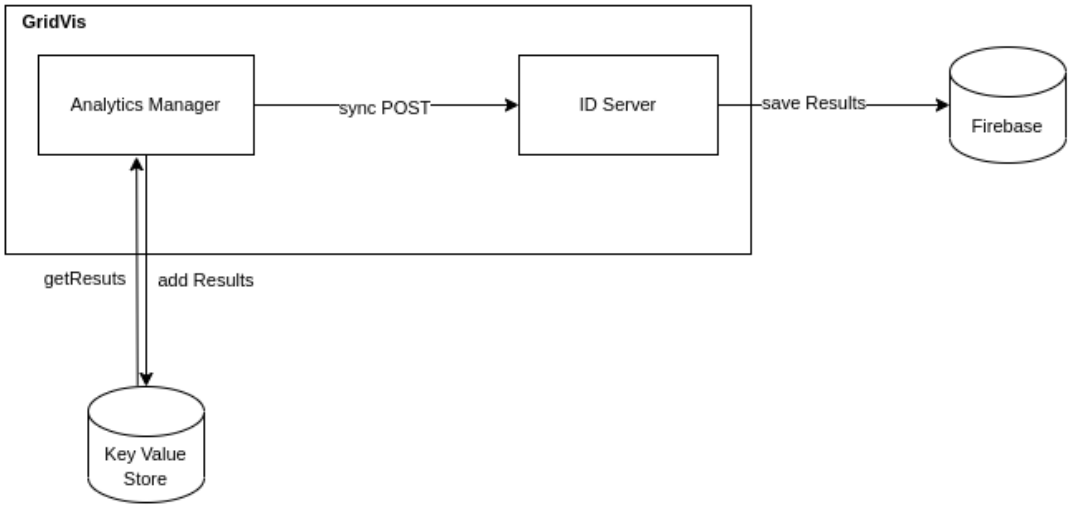

Abbildung 4.1: Verlauf des aktuellen Analysesystems

4.2.2 DSGVO Grundlage

Vor dieser Arbeit hat sich die Firma bereits mit der DSGVO als Grundlage für die Analyse der Software befasst. Bevor Kunden die Software nutzen, wird eine Zustimmung zu den allgemeinen Geschäftsbedingungen (AGB) eingeholt. Diese Bedingungen umfassen auch die Erhebung von Daten, um die Nutzung der Software zu ermöglichen und zu optimieren. Das bedeutet, dass die Datenerhebung für die Analyse der Software nicht untersagt ist. Das Ziel der Analyse ist ebenfalls vorhanden und im Abschnitt 4.1.2 formuliert. Die Datenverarbeitung sollte weiterhin datenschutzkonform erfolgen, um keine Rückschlüsse auf die Nutzer zuzulassen und deren Deanonymisierung zu verhindern.

Die folgenden Schritte zur Einhaltung der DSGVO sind die initiale Risikobetrachtung (siehe 4.4), die Anonymisierung (siehe 4.5.3) sowie die erneute Risikobetrachtung (siehe 4.7). Diese Schritte werden nach der Auswahl der Metriken im nächsten Abschnitt durchgeführt und beschrieben.

4.3 Auswahl der Metriken

Im Konzept 3.2 wurden die drei Kategorien von Metriken vorgestellt. Die Zielsetzung 4.1.2 betont das Verhalten und die Interaktion mit der Anwendung als vorrangig. Für die Analyse der Berichtseditor-Anwendung könnten sowohl klick- als auch ereignisbasierte Metriken eingesetzt werden, während generelle Metriken keine wertvollen Informationen für unsere Zwecke liefern. Um das Verhalten mit den Objekten in der Anwendung genauer zu verfolgen, müssen die Metriken präziser formuliert werden. Als ereignisbasierte Metriken werden gesammelt:

- Anzahl der Seiten
- Durchschnittliche Anzahl der Objekte pro Seite
- Verteilung von Objekten
- Verteilung von Objekten (gewichtet je nach aktiver Nutzung/Öffnung der Berichte)
- Durchschnittliche Messwertauswahl
- Verteilung des genutzten Seitenformats
- Verteilung bei der Nutzung der zeitlich automatisierten Berichterzeugung

Als klickbasierte Metriken könnten wir folgende Metriken sammeln:

- Die Anzahl der Klicks für einen Objekttyp
- Die Anzahl der Mausbewegungen für einen Objekttyp

Die Kombination aus klick- und ereignisbasierten Metriken könnte einen umfassenderen und aussagekräftigeren Überblick über die tatsächliche Nutzung der Objekte/Funktionen in der Anwendung liefern. Die Datenerhebung dafür wird im Abschnitt 4.5 umgesetzt.

4.4 Initiale Risikobetrachtung

Nach der Bestimmung der Metriken sollten wir die Schadenshöhe sowie die Schadenwahrscheinlichkeit, wie im Abschnitt 3.4.3 dargelegt, betrachten. Wie aus den formulierten Metriken hervorgeht, werden grundsätzlich keine personenbezogenen Informationen aus der Anwendung gesammelt. Trotzdem stellen wir ein ungünstiges Szenario vor, bei dem ein unberechtigter Zugriff auf die gespeicherten Daten erfolgt.

Die Speicherung der Daten wird im Abschnitt 4.6 weiter erläutert. Die erfassten Daten werden pro Projekt gesammelt und gespeichert. Dabei wird der `projectLicenceKey` verwendet, der eine bereits verschlüsselte Form für die Zuordnung der Daten darstellt. Selbst wenn die Kollektionen/Tabellen, die die Informationen pro Schlüssel speichern, offenbart würden, besteht immer noch keine Verbindung zu einer bestimmten Person. Die Schadenshöhe sowie die Schadenswahrscheinlichkeit wären in diesem Fall gering. Nur im Fall der vollständigen Offenlegung aller Tabellen zur Lizenz- und Benutzerverwaltung in der Firma könnte eine Reidentifizierung erfolgen, indem nachgeschaut wird, ob der jeweilige Nutzer zu dem Projekt gehört. Die Schadenshöhe wäre dabei hoch, jedoch unwahrscheinlich.

Nach der Bewertung der Schadenshöhe und Schadenswahrscheinlichkeit könnte das Risiko als gering eingestuft werden. Da keine Personenbezüge bei der Erfassung vorhanden sind und die Zuordnung zu einem Projekt nur im Falle der vollständigen Veröffentlichung aller Daten zur Lizenz- und Benutzerverwaltung besteht, ist das Risiko als gering anzusehen.

4.5 Datenerhebung

Gemäß der für die Untersuchung der Anwendung festgelegten Metriken im Abschnitt 4.3 und dem Vorgehen beim Systemüberblick 3.8 können die Methoden ausgewählt werden, um die gewünschten Metriken zu erfassen. Da die Log-API im Wesentlichen allgemeine Informationen liefert, ist diese Methode für die Umsetzung in der Firma nicht relevant. In den folgenden Abschnitten konzentrieren wir uns auf die Umsetzung von zwei anderen

Methoden, nämlich dem Ereignis-Tracking zur Sammlung klickbasierter Metriken und der Extraktion zur Erfassung ereignisbasierter Metriken.

4.5.1 Umsetzung vom Ereignis-Tracking

Durch die Implementierung des Ereignis-Trackings werden die klickbasierten Metriken aus der Anwendung des Berichtseditors gesammelt. Hiermit wird der Ablauf, wie auch bei der Vorstellung der Methode im Konzept in der Abbildung 3.2, eingehalten.

Zuerst sollte die Integration in der bestehenden Infrastruktur überlegt werden. Der Berichtseditor befindet sich in einem Webmodul des Monorepos namens Graphdesigner. In diesem Repository gibt es auch eine andere ähnliche Anwendung namens Dashboardeditor. Beide Anwendungen nutzen durch ihre Abhängigkeiten Module, die beispielsweise die Objektsimplementierung haben. Auf diese Weise wird der Code ohne Wiederholungen genutzt. Genauso könnte das Ereignis-Tracking als ein Modul definiert werden, damit es später auch in weiteren Anwendungen integriert werden kann. Wie in der Abbildung 4.2 dargestellt, kann das Ereignis-Tracking im Graphdesigner-Monorepo unter anderem in Webmodulen platziert werden, sodass die Anwendungen in diesem Repository das Ereignis-Tracking durch die Angabe von Abhängigkeiten in der App nutzen können.

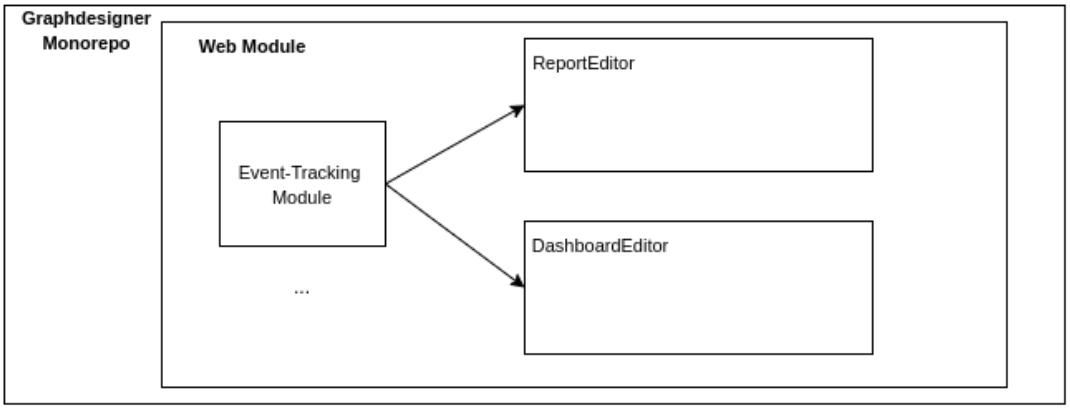

Abbildung 4.2: Ereignis-Tracking Module integriert im Graphdesigner Infrastruktur

Die Implementierung des Ereignis-Trackings kann entweder durch eigene Funktionen oder durch die Verwendung einer Bibliothek mit entsprechender Funktionalität erfolgen. Ein Beispiel hierfür wäre die React-Tracking Bibliothek, die durch Komponenten, Dekoratoren oder Hooks das deklarative Tracking ermöglicht (vgl. [Gay18]). Die Verwendung einer solchen Bibliothek könnte die Implementierungszeit sparen, aber einige Überlegungen sollten berücksichtigt werden. Es gibt jedoch potenzielle Nachteile in der Bibliothek, wie die fehlende Berücksichtigung von Duplikaten beim Speichern im Data-Layer und die Verwendung einer beliebigen Struktur als Partial in der Hook-

Funktion für Ereignisse, ohne die Verwendung eines Interfaces. Um diese potenziellen Nachteile zu überwinden, könnte ein eigenes Interface und die Hook-Funktion für das Ereignis-Tracking implementiert werden, die zusammen mit der Bibliothek eingesetzt werden.

Als nächstes bestimmen wir die Hauptfunktionalität, die das Ereignis-Tracking-Modul liefern sollte, um das Tracking zu ermöglichen. Für bessere Lesbarkeit und Übersichtlichkeit wird die Logik in separate Funktionen nach Aufgaben aufgeteilt:

- Funktionale Komponente: Diese Funktionale Komponente erhält die zu verfolgende Anwendung als Kind, um das Tracking in der gesamten App zu ermöglichen.
- Ereignis-Interface: Die Beschreibung der Struktur des Ereignisses, das verfolgt werden soll.
- Funktion für Ereignisintegration: Diese Funktion wird in das Ereignis integriert.
- Funktion für Datenversendung: Diese Funktion wird für die regelmäßige Versendung der Daten im Intervall und beim Schließen der Anwendung verwendet.
- Hilfe-Funktionen: Diese Funktionen dienen der Überprüfung von Duplikaten, dem Aktualisieren der Daten, der Bereinigung des Data-Layers, etc.

Zum Tracking in der gesamten Anwendung, ohne die Daten in der Hierarchie des Komponentenbaums zu übermitteln, wird eine funktionale Komponente definiert und verwendet. Wie in Abbildung 4.3 zu sehen ist, nimmt diese Komponente namens `TrackedApp` die Props mit Metadaten sowie die Kind-Komponente entgegen. Als Kind wird selbst die Anwendung übermittelt, in unserem Fall der Berichtseditor. Zudem erfolgt hiermit die Duplikatbehandlung vor dem Hinzufügen der Daten in den Data-Layer.

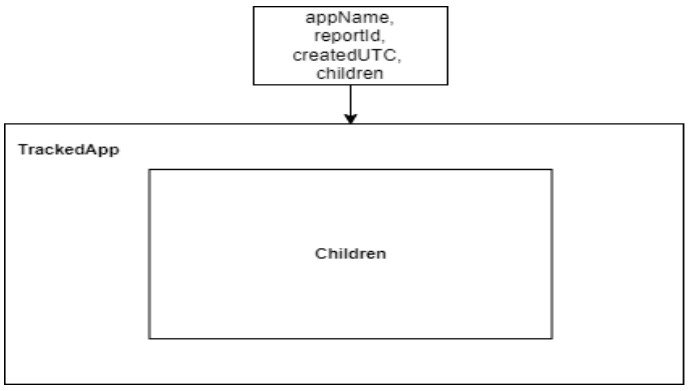

Abbildung 4.3: Funktionale Komponente

Wie im Konzept zum Ereignis-Tracking bereits erwähnt wurde, ist eine Struktur für die Sammlung der Ereignisse erforderlich. Das Interface könnte gemäß dem Konzept wie in der folgenden Abbildung 4.4 aussehen:

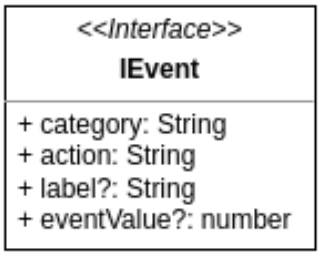

Abbildung 4.4: Interface für Ereignisse

Um ein Ereignis zu tracken, wird die Funktion definiert. Die Rückgabe ist eine Callback-Funktion namens `handleEvent`, die für das Tracking in den gewünschten Ereignissen verwendet wird und ein Ereignis mit der Struktur `IEvent` erwartet. Die Information über die Interaktion mit den Objekten könnte mithilfe von Ereignissen wie `onMouseDown` und `onMouseEnter` verfolgt werden. Anstelle dieser Ereignisse im Code wird die Callback-Funktion aufgerufen. Somit werden die Ereignisse gesammelt und gezählt, sobald der Nutzer ein Objekt mit der Maus trifft oder darauf klickt. In der Abbildung 4.5 ist der gesamte Verlauf des Trackings für diese Fälle dargestellt.

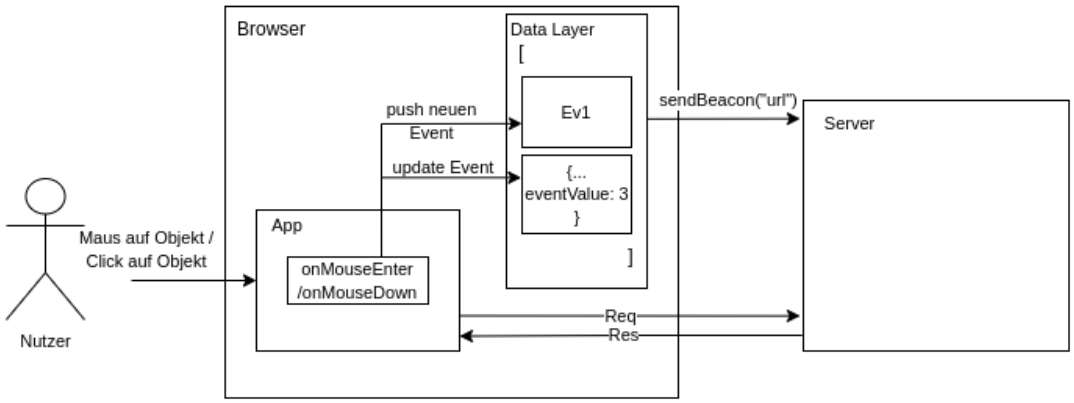

Abbildung 4.5: Ereignis-Tracking Verlauf für die Objekte

Zum Versand der Daten aus Data-Layer wird eine Funktion definiert. Dabei werden zwei Fälle für den Versand betrachtet. Im ersten Fall werden die Daten im Intervall von einer Minute verschickt, sobald die Elemente im Data-Layer die Obergrenze von 60KB erreichen. Der andere Fall tritt beim Schließen der Seite auf. In beiden Szenarien wird, wie im Konzept vorgestellt, `sendBeacon` zum Versand angewendet, wobei die

Daten im String-Format übermittelt werden. Bei jedem Versand wird der Data-Layer bereinigt.

Die Speicherung der gesammelten Daten beim Ereignis-Tracking wird im Abschnitt 4.6.1 behanndelt.

4.5.2 Umsetzung vom Extraktion

Als weitere Datenerhebungsmethode wird die Extraktion vorgestellt, die ereignisbasierte Metriken über den Berichtseditor sammelt. Die Extraktion wird als eigenständiges Modul im REST-Teil des Monorepo Graphdesigners positioniert, in dem sowohl der Berichtseditor als auch der Dashboardeditor liegen. Die Integration in der bestehenden Architektur ist in Abbildung 4.6 dargestellt.

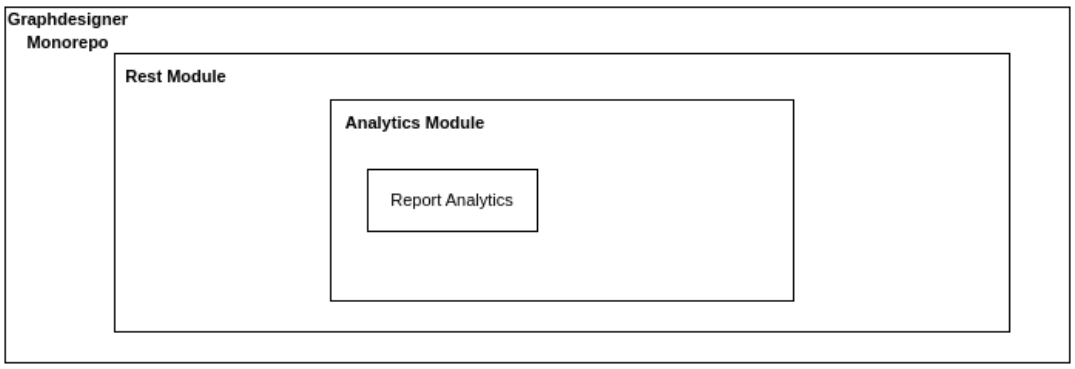

Abbildung 4.6: Integration der Extraktion Methode in Graphdesigner Monorepo

Um die Zukunftsfähigkeit des Systems sicherzustellen und das Analysesystem nicht nur auf den Berichtseditor zu beschränken, sondern auch für andere Anwendungen im Unternehmen zu nutzen, werden möglichst generische Funktionen und Interfaces formuliert. Diese sollen die Erweiterbarkeit des Systems ermöglichen. Die Struktur des Analytics-Moduls ist wie folgt organisiert:

- Ein automatisierter Job, der beim Serverlauf im festgelegten Intervall Informationen sammelt.
- Eine Klasse zur Erstellung der Ergebnisse im JSON-Format.
- Eine eigene Parser-Klasse, um für eine Instanz Informationen und Metriken zu erheben.
- Eine Analytics-Klasse für ein Projekt und für jede Anwendung, von der Daten erhoben werden.

- Hilfsfunktionen für das Mapping und die Erstellung von Skalen.

In der Abbildung 4.7 ist das UML-Diagramm für die Analytics-Klassen dargestellt. Ein Interface wird bereitgestellt, das zwei Methoden enthält. Die erste Methode, `collectAnalytics`, sollte den Parameter Projekt akzeptieren und die Informationen jeweils für eine Anwendung oder einen interessierenden Bereich sammeln. Die zweite Methode, `createJson`, sollte einen Parameter vom Typ `AnalyticsJson` akzeptieren, um die gesammelten Daten im JSON-Format für das Endergebnis zu konvertieren. `AnalyticsJson` ist eine Klasse mit generischen Methoden, die Ergebnisse in eine einheitliche Form bringen.

Die Klasse `ProjectAnalytics` überschreibt die Methoden des `IAnalytics`-Interfaces und dient zur Koordination der Instanzen, die Analytics von verschiedenen Anwendungen behandeln. In der Methode `collectAnalytics` ruft diese Klasse die Analytics der anderen Instanzen für jedes Projekt auf. Gleichzeitig werden die überschriebenen Methoden der Instanzen in der Methode `createJson` aufgerufen, um das JSON-Ergebnis zu erstellen. Um das Endergebnis für den gesamten Analytics zu erzeugen, wird die Methode `buildAnalytics` zur Verfügung gestellt.

In der Klasse `ReportsAnalytics` werden in der Methode `collectAnalytics` zunächst die Berichte für das jeweilige Projekt abgefragt. Anschließend wird iterativ für jedes Dokument die Information extrahiert und in einer Variable der Klasse gespeichert. Bei den Objekten handelt es sich dabei um die private Multimap-Variable. In der weiteren Methode `createJson` wird je nach angegebenem Typ ein formatiertes Ergebnis im JSON-Format in der Instanz von `AnalyticsJson` geschrieben.

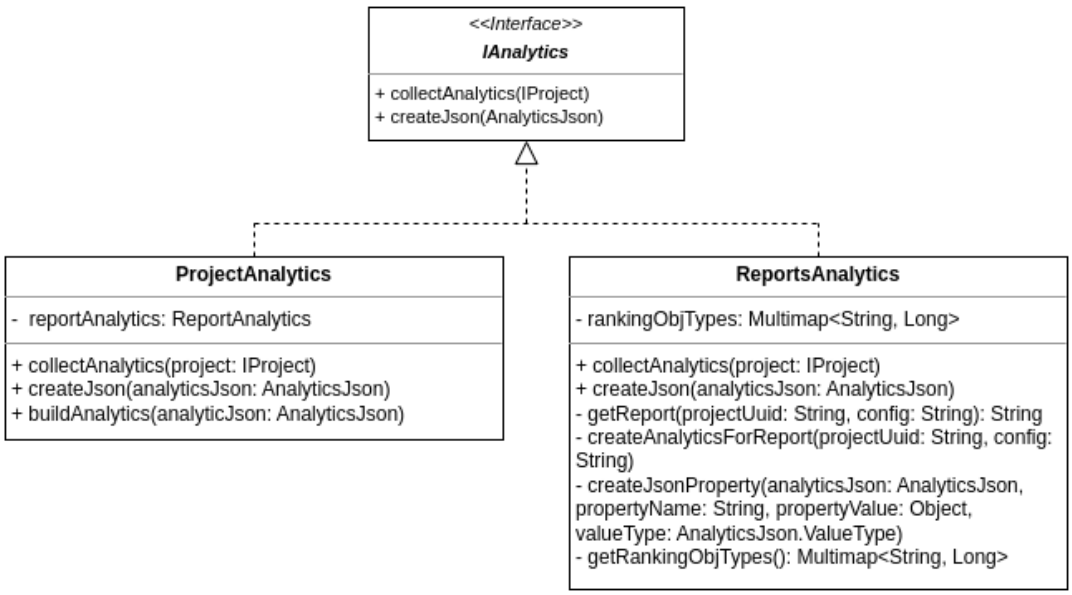

Abbildung 4.7: UML-Diagramm für Analytics Instanzen

Für jede Anwendung benötigen wir eine eigene Parser-Klasse, die die gewünschten Metriken aus JSON, einer API oder einer Datenbank für eine Instanz (Objekt oder Dokument) extrahiert. Im Fall des Berichtseditors möchten wir Informationen über die in den Dokumenten enthaltenen Objekte erhalten. Jeder Kunde bzw. jedes Unternehmen, das die Berichtseditor-Anwendung nutzt, erstellt Dokumente innerhalb eines Projekts. Die Struktur eines Dokuments wird auf dem Server in Form von JSON gespeichert. In diesem JSON befinden sich Informationen über die vorhandenen Objekte auf einer Seite im Dokument. Dadurch können wir zunächst herausfinden, wie viele Objekte eines bestimmten Typs in einem Dokument auftreten. Aus dieser statischen Sicht des Dokuments können wir auch eine gewichtete Anzahl der Objekte nach Typ erfassen, da JSON auch Metadaten enthält, einschließlich der Anzahl der Öffnungen dieses Dokuments. Dies ermöglicht uns, zu unterscheiden, ob es generelle Auftritte eines Objekts gibt oder ob sie von der Öffnung des Dokuments abhängen. In der Parser-Klasse für Berichte wird die Erfassung von Objekten nach Typ und Anzahl umgesetzt. Die Werte werden aus dem JSON anhand des Schlüssels von jeder Seite des Dokuments extrahiert und in einem Ergebnis festgehalten.

In der Abbildung 4.8 wird der Ablauf des automatisierten Jobs zur Extraktion und Übermittlung von Analytics dargestellt. Der Job enthält eine Überprüfung, ob an diesem Tag bereits ein Ergebnis gesammelt wurde. Zudem wird hier das Intervall für die Ausführung auf sechs Stunden eingestellt. Falls der Job die Überprüfung besteht, wird eine Instanz von `ProjectAnalytics` erstellt. Dabei werden iterativ für jede Analytics-Anwendung die Daten gesammelt. Anschließend wird die bereits implementierte `AnalyticsManager`-Instanz (siehe Abschnitt 4.2.1) genutzt, um die Daten im Key-Value-Store zu speichern. Aus dieser Datenbank werden die Daten später synchronisiert und an den ID-Server verschickt, um sie in einer dokumentenbasierten Datenbank zu speichern.

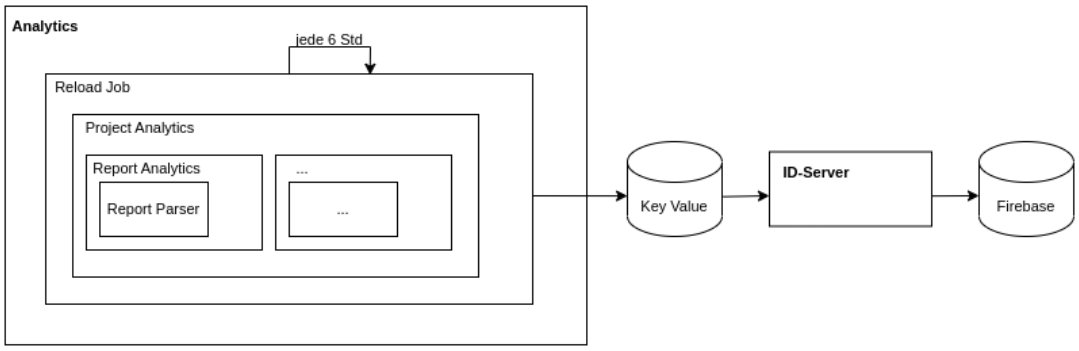

Abbildung 4.8: Verlauf der Extraktion Methode für Berichtseditor

4.5.3 Anonymisierung der Daten

Um den Personenbezug sowie die Möglichkeit eines Rückschlusses auf eine Person mit zusätzlichen Informationen zu vermeiden, wird das Anonymisierungsverfahren durchgeführt. Im Abschnitt 3.4.3 wurden die grundlegenden Techniken zur Anonymisierung beschrieben.

Bei der Sammlung der Daten sollte als allererstes der Personenbezug vermieden werden, soweit es möglich ist. Das bedeutet, es werden nur die für die interessierenden Metriken benötigten Daten gesammelt. Die Identifikatoren werden bei beiden Datenerhebungsmethoden gelöscht und nicht gesammelt. Die IP-Adresse als Quasi-Identifikator, um einen Nutzer zu unterscheiden, ist für das Verständnis der Interaktion mit der Anwendung nicht erforderlich und wird daher ausgeschlossen.

Die gesammelten Daten werden pro Projekt gruppiert und gespeichert. Das Gesamtergebnis wird anschließend aus allen Daten nach der Gruppierung gebildet. Somit wird die Reidentifikation durch Datenbankabfragen und bei der Visualisierung vermieden.

Zusammengefasst wurden bei der Datenerhebung zum Großteil erst die Unterdrückung und die weitere Gruppierung der Daten angewendet, um die Anonymisierung zu gewährleisten. Zudem haben die bereits formulierten Metriken aus dem Abschnitt 4.3 den Aufwand zur Anonymisierung gespart, da diese Metriken von vornherein keinen direkten Personenbezug aufweisen.

4.6 Speicherung der Daten

Im Konzept wurden drei Optionen zur Speicherung vorgeschlagen. Bei der Realisierung im Unternehmen wurde die Datenspeicherung vom alten Modul für die Datenextraktion bereits übernommen. Dabei wurden zwei verschiedene Arten von Datenbanken angewendet, nämlich Key-Value Store und dokumentenbasierte Datenbank. Dieses Vorgehen stimmt mit der Option 2 im Konzept überein. In den nächsten Abschnitten werden wir die Speicherung für die einzelnen Datenerhebungsmethoden sowie für das Gesamtergebnis genauer betrachten.

4.6.1 Speicherung beim Ereignis-Tracking

Die Daten aus dem Ereignis-Tracking sollten mittels einer POST-Anfrage auf dem Server landen. Der Datenfluss ist unnormiert und kann je nach Nutzer unterschiedlich sein. Zudem können die Daten von einem Projekt und einem Bericht mehrmals am Tag, sogar stündlich, vorkommen. Die Nutzung und Umsetzung einer relationalen Datenbank

für solche Datenmengen ist nicht optimal. Die Begründung dafür kann im Abschnitt 3.5 vom Konzept gelesen werden.

Daher können wir die gesammelte Information in einer NoSQL-Datenbank speichern. Wie in der Extraktion können wir einen Key-Value Store anwenden, um die Zwischenspeicherung der Daten zu ermöglichen. Das hat den Vorteil, dass wir ein Gesamtergebnis pro Projekt mit den Daten aus beiden Methoden in `ProjectAnalytics` abfragen können und die bereits bestehende Implementierung anwenden können. Als Schlüssel wird aus den Metadaten beim Ereignis die Document-ID und das Erstellungsdatum angewendet. Die Abbildung 4.9 stellt die graphische Darstellung der Umsetzung dar.

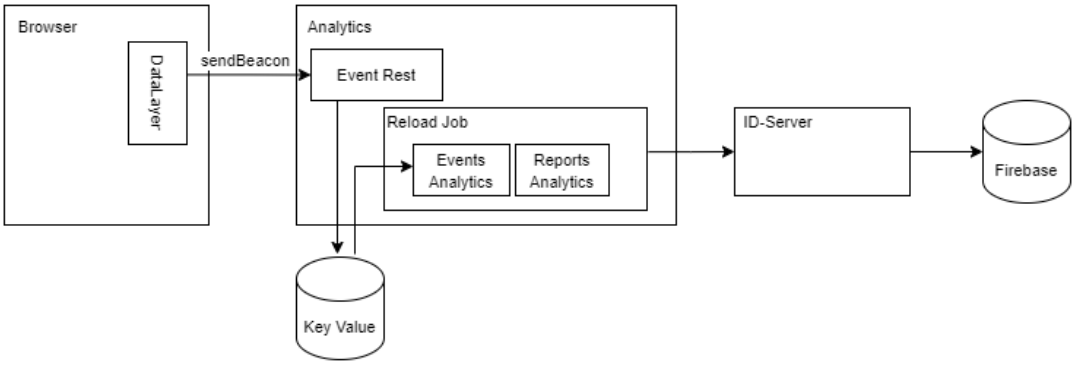

Abbildung 4.9: Speicherungsablauf beim Ereignis-Tracking

4.6.2 Speicherung bei der Extraktion

In der Extraktionsmethode wurde der Ablauf vorgestellt, bei dem die Daten vom Server zum ID-Server-Dienst gelangen und von dort in der dokumentenbasierten Datenbank gespeichert werden (siehe Abschnitt 3.5). Abhängig vom `ProjectLicenceKey` werden die Daten, die vom Server stammen, in einem Dokument in der Kollektion `Analytics` gespeichert. Täglich wird unter einem Dokument ein leeres `Analytics Job` vorbereitet, um das Analytics-Ergebnis aus dem GridVis Server zu sammeln. Falls an einem Tag kein Serverlauf für das Projekt stattfindet, wird kein leeres Job am nächsten Tag für dieses Projekt erzeugt, und ein unbenutzter Job wird erneut verwendet. Bei erfolgreichem Verlauf wird die Job-ID im Key-Value-Store gespeichert, wobei die IDs der abgeschlossenen Jobs gesammelt werden. Genau diese IDs werden für die Überprüfung bei automatisierten Jobs verwendet, um festzustellen, ob die Ausführung erforderlich ist.

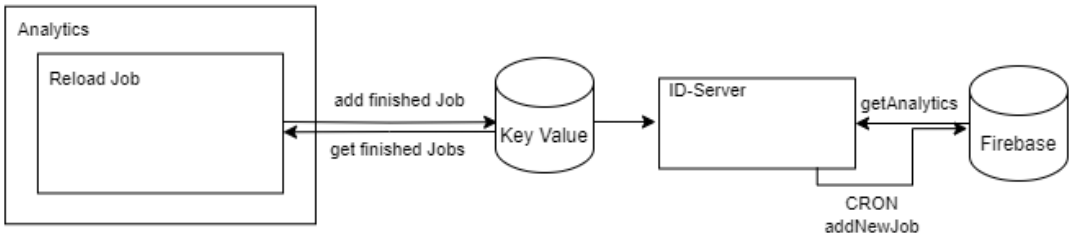

Abbildung 4.10: Speicherungsablauf bei der Extraktionsmethode

4.6.3 Speicherung vom Gesamtergebnis

Sobald ein Ergebnis vom Server in der dokumentenbasierten Datenbank landet, wird es in der `Analytics`-Kollektion unter einem Projektschlüssel gespeichert. Da das Analysesystem auch zukünftig Daten aus anderen Anwendungen sammeln soll, wird eine weitere Kollektion mit dem Namen `analyticsStatistics` erstellt. Diese Kollektion sollte jeweils Unterkollektionen für verschiedene Anwendungen haben. Beispielsweise könnte für den Berichtseditor der Name `Report` lauten. In dieser Kollektion können wir die Speicherung von Ergebnissen für die Benutzeroberfläche beibehalten. Da wir sowohl historische als auch aktuelle Daten anbieten möchten, könnte unter `Report` die Unterkollektion `History` erstellt werden. Darunter könnten Ergebnisse für den aktuellen Tag mit einer CRON-Funktion für alle Berichte der Projekte aus der Kollektion `Analytics` erstellt werden. Auf diese Weise werden im Laufe der Zeit historische Daten gesammelt. Der Datenfluss bei der Speicherung wird erneut in der Abbildung 4.11 dargestellt.

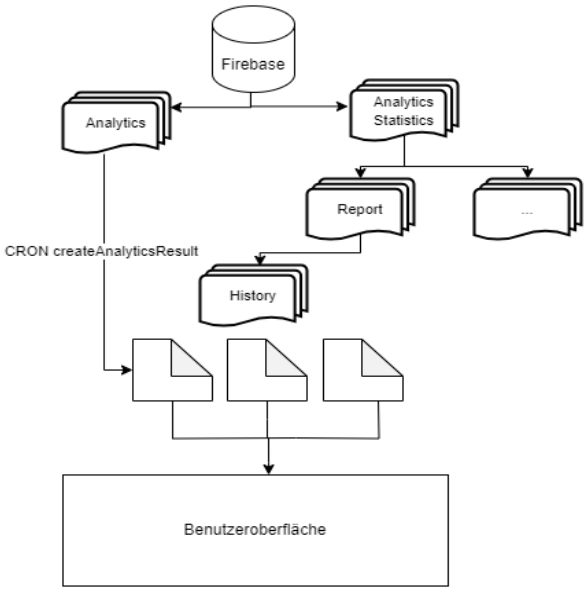

Abbildung 4.11: Speicherungsablauf bei Analytics Gesamtergebnis

4.6.4 Speicherung und Verwaltung der erlaubten Nutzer für Analytics

Das entwickelte Analysesystem wird vorerst in der Firma begrenzt eingesetzt und weiterhin in einer Testphase verwendet. Dafür wird vor der Datensammlung die Analytics-Funktion mit `pasw-feature-analytics=ON` getoggelt und geprüft. Nur wenn dieser Feature Toggle eingeschaltet ist, wird Analytics beim Nutzer im Projekt erfasst.

Bei der Speicherung der Extraktionsmethode wurde bereits eine CRON-Funktion für das Erzeugen eines neuen Jobs für die Datensammlung erwähnt. Diese Funktion erzeugt nicht nur Jobs für bereits vorhandene Projekte, sondern fügt automatisch neue Projekte hinzu, die überwacht werden müssen. Dies wird erreicht, indem die erlaubten Nutzer für Analytics aus der Kollektion abgefragt werden. In der Kollektion für die Analytics-Nutzer (`analyticsUsers`) werden die `ownerId`s der Lizenz eines Projekts als Dokument eingefügt und abgefragt. Aktuell wird die begrenzte Anzahl der internen Nutzer in der Kollektion angelegt. Für die weitere Testphase könnten `ownerId`s von den Nutzern benutzt werden, die zu der Firma gehören. Es kann mithilfe der Email-Endung `@janitza.de` beim Hinzufügen der erlaubten Nutzer gefiltert werden.

4.7 Erneuerte Risikobetrachtung

Nach der Sammlung und Speicherung der Daten wird die Risikobetrachtung erneut durchgeführt. Hierbei schätzen wir die Möglichkeit einer Personeneingrenzung im Datensatz und ob Rückschlüsse mit zusätzlichen Informationen möglich wären.

Die Kollektionen, die das Gesamtergebnis beispielsweise für den Berichtseditor speichern, gruppieren die Daten und geben keine Personeneingrenzung an. Selbst auf der Benutzeroberfläche werden die Daten als Gesamtergebnis gruppiert angezeigt, sodass weder eine Person noch ein Projekt direkt zugeordnet werden kann.

Wie bereits in der initialen Risikobetrachtung (siehe Abschnitt 4.4) erwähnt wurde, besteht das Risiko nur im Fall, wenn die Lizenz-Kollektion sowie die Nutzer-Kollektion, die für die gesamte Projekt- und Benutzerverwaltung eingesetzt ist, offengelegt werden. Die einzige Kollektion mit potenzieller Gefahr ist die `Analytics`-Kollektion, die die Daten aus dem GridVis pro Projektschlüssel erhebt und speichert. Mit zusätzlichen Informationen, durch Offenlegung der Lizenz-Kollektion, ist die Zuordnung zu einem Nutzer nicht auszuschließen. Selbst in der Lizenz-Kollektion ist es möglich, die `ownerId` sowie die Nutzer-IDs dieses Projekts herauszufinden. Wenn dazu noch die weiteren Kollektionen mit den Nutzerdaten wie `UserDetails` offen werden, besteht ein Risiko. Da die Zugehörigkeit zu einem Projekt nicht direkt auf eine Person schließen lässt, ist die Zuordnung nicht unmittelbar und bei der Offenlegung gehemmt. Zudem enthält

die Verwaltung der Nutzer, die das Analysesystem in der Testphase ausprobieren, verschlüsselte `OwnerId`. Auch bei Offenlegung der Lizenz-Kollektion gibt es keine direkten Rückschlüsse auf die Person, sondern nur mit zusätzlichen Informationen.

Da das entwickelte Analysesystem in die bestehende Infrastruktur integriert wird, überschreitet die Datensicherheit die Grenzen dieser Arbeit hinaus. Die gesammelten und gespeicherten Daten für den Analytics wurden anonymisiert und haben keinen Personenbezug. In der Firma wird ständig die Sicherheit des gesamten Systems überprüft, daher gehen wir davon aus, dass die erforderlichen Maßnahmen während der Entwicklung getroffen wurden. Das Risiko wird wie bei der initialen Risikobetrachtung als gering eingestuft. Weitere Anonymisierungsverfahren sind daher nicht erforderlich.

4.8 Visualisierung: Benutzeroberfläche

Zur Visualisierung der gesammelten Daten wird das Modell aus dem Konzepts Abschnitt 3.6 eingesetzt. Wie auch dort beschlossen wurde, wird eine Übersichtsseite zunächst für den Nutzer angezeigt, wobei die Kennzahlen von Metriken gruppiert und nach Relevanz positioniert werden.

Die gesammelten Metriken aus dem Berichtseditor sollten priorisiert werden, um die zentrale Position und die Aufmerksamkeit der Nutzer auf die relevanteste Information zu lenken. Da die Zielsetzung für den Berichtseditor auf das Verständnis der Interaktion mit der Anwendung und Funktionen liegt, werden die Metriken über die Objektverteilung mehr Raum einnehmen und prominenter präsentiert. Andere Metriken erscheinen als kleine Karten. Bei einem Klick auf “Details” wird sowohl für die kleineren als auch für die größeren Karten eine detaillierte Ansicht angezeigt. Die umgesetzte Startseite sieht wie in der Abbildung 4.12 aus. Die Aggregation wird ebenfalls interaktiv angeboten, beispielsweise werden die Werte in der Abbildung für den Tag dargestellt.

Auf der detaillierten Ansicht wird eine kombinierte Darstellung der Daten aus einem Diagramm und einer Tabelle angeboten. Neben der Visualisierung der Daten werden auch die Aggregation und der Datenexport zur Interaktion angeboten. Aus der Tabelle über die möglichen Darstellungsmöglichkeiten im Konzept (siehe Tabelle 3.4) können wir somit die Entscheidung über die passende Darstellungsform treffen. Zwar decken Diagramme wie Bubble Chart und Scatterplot alle visuellen Erkenntnisziele ab, trotzdem sind sie nicht für alle Datentypen geeignet. Deshalb ist es wichtig, die Datentypen unserer Metriken noch zu betrachten.

Die gesammelten Daten über Objekte stellen eine Nominalskala dar und sind somit in Gruppen unterteilt. Die Werte haben den numerischen Typ. Zudem wird die zeitliche Aggregation der Daten angeboten und als ein Schlüssel pro Wert angewendet. Insgesamt

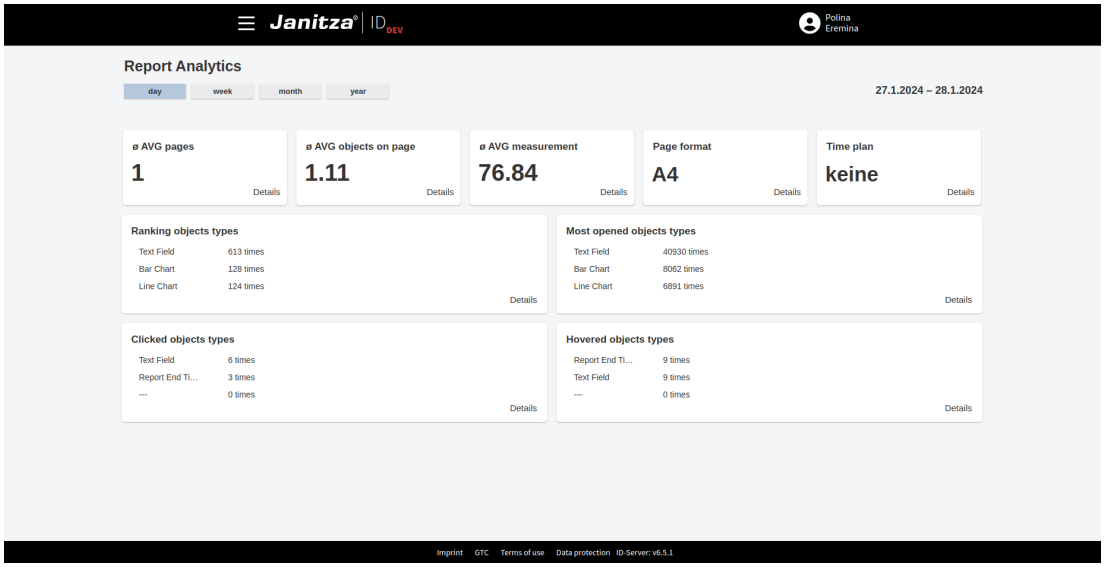

Abbildung 4.12: Umsetzung der Startseite

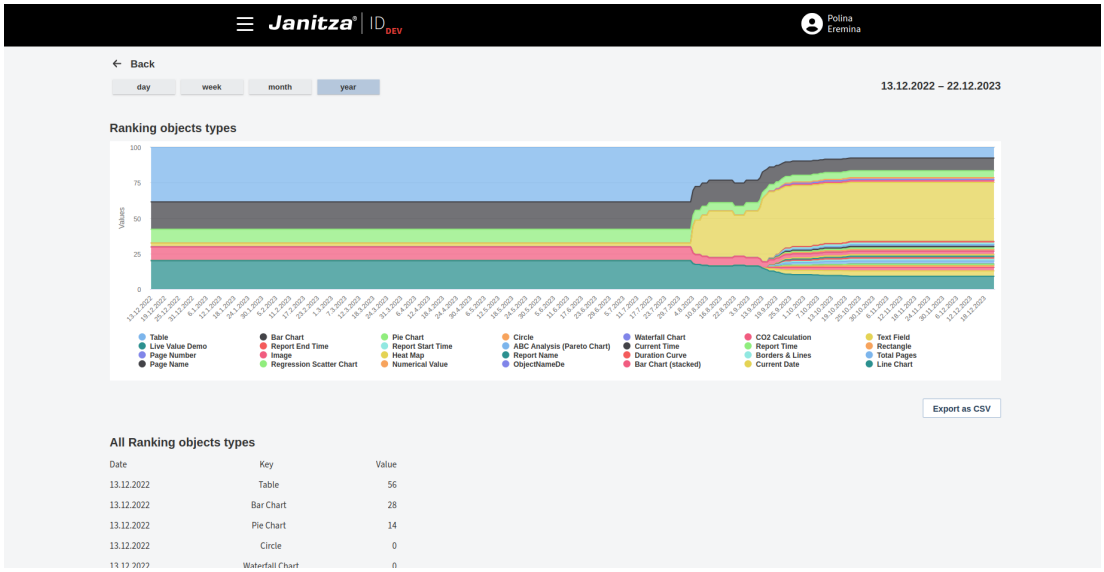

Abbildung 4.13: Umsetzung der Detailseite für die Metriken mit den zwei Schlüsseln

sind somit ein Wert und zwei Schlüssel im Fall der Objekte vorhanden (kategorisierte Schlüssel als Objekttyp und geordnete Schlüssel als Datum). Aus dieser Betrachtung ist ein gestapeltes Flächendiagramm für die Visualisierung möglich. Die anderen Metriken wie durchschnittliche Anzahl stellen einen numerischen Wert mit dem geordneten Schlüssel (Datum) dar. Deswegen können zwischen Linien- und Flächendiagramm gewählt werden.

Auf der umgesetzten Detailsansicht wird je nach Metriks Datentyp entweder die Ansicht mit dem gestapelten Flächendiagramm oder mit dem Liniendiagramm angezeigt (siehe Abbildungen 4.13 und 4.14). In beiden Fällen wird die Aggregation zur Anpassung des Zeitraums angeboten. Nach dem Diagramm folgt die Tabelle mit den Daten für

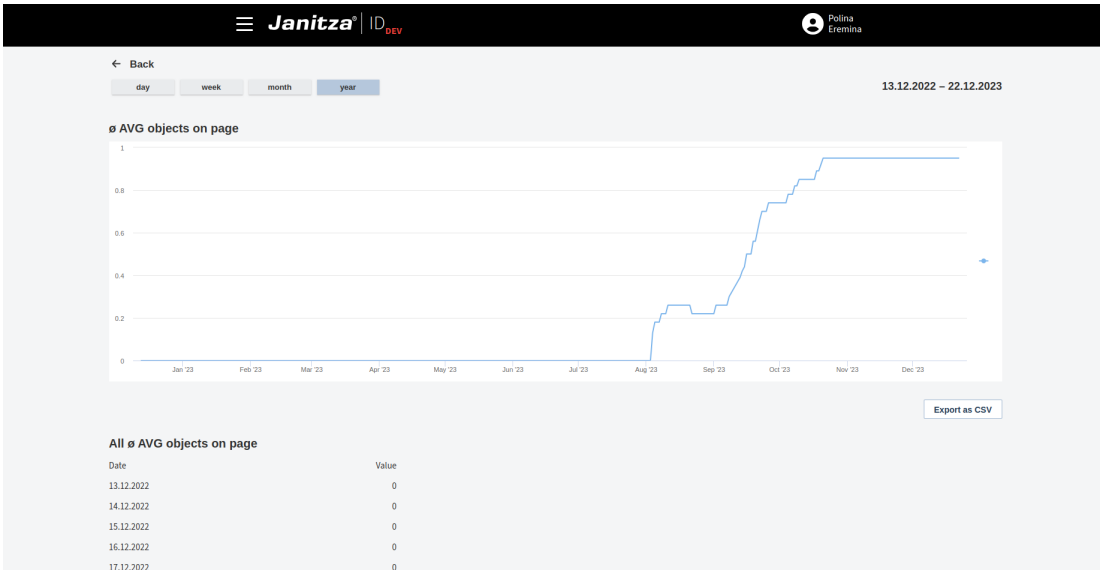

Abbildung 4.14: Umsetzung der Detailseite für numerische Metriken mit einem geordneten Schlüssel

den ausgewählten Zeitbereich. Beim Flächendiagramm werden die Schlüssel als Labels unter dem Diagramm angezeigt, die eine Interaktion und die Anpassung der Darstellung ermöglichen, indem beim Klick auf das Label die Daten zur Darstellung entweder ein- oder ausgeschlossen werden.

5 Zusammenfassung

In diesem Kapitel werden die Ergebnisse sowie die Forschungsfragen auf die Beantwortung und Erfüllung bewertet und diskutiert. Nachfolgend werden die mögliche weitere Ansätze in diesem Themenbereich sowie die nächsten Schritten in der Forschung berschrieben.

5.1 Fazit

Im Rahmen dieser Arbeit wurde ein Modell zur Entwicklung eines Analysesystems konzipiert. Dieses Modell wurde bei der Firma Janitza unter Berücksichtigung der spezifischen Anforderungen und Ziele des Unternehmens umgesetzt.

Im ersten Kapitel werden für diese Arbeit vier Forschungsfragen formuliert und in der Arbeit untersucht. Die erste Frage FF1 bezieht sich auf mögliche Datenerfassungsmethoden, die effizient eingesetzt werden könnten. Im Konzept 3.3 wurden die drei möglichen Datenerhebungsmethoden mit deren Schwerpunkt und Vor-/Nachteilen vorgestellt. Somit kann je nach gezielten und geforschten Metriken die passende Methode gewählt werden. Die Forschungsfrage wurde beantwortet. Die nächste Forschungsfrage FF2 zielt sich ein Vorgehen zu konzipieren, das für die Einheiltung der DSGVO dienen sollte. Im Abschnitt 3.4.3 zeigt das Akitivitätsdiagramm, das Vorgehensweise für eine datenschutzkonforme Datenverarbeitung. Dieses Verfahren wurde in der Realisierungsphase eingehalten. Nach der erneuten Risikobetrachtung wurde das Risiko als geringes eingestuft. Somit lässt sich diese Frage auch beantworten. Bei der Forschungsfrage FF3 geht es um die Untersuchung der übersichtliche Visualisierungsmöglichkeiten. Es wurde ein Visualisierungsmodell entwickelt, das als Ziel die Erkenntnissprozess und Analyse der Daten unterstützen sei. Zudem wurden das Vorgehen bei der Auswahl der Diagrammen vorgestellt, indem es je nach Erkenntnissziel und Datentyp eine Unterscheidung und Klassifikation der Diagrammen zur Verfügung gestellt wird. (siehe Tabelle 3.4)

Die letzte Forschungsfrage FF4 bezieht sich auf den Beitrag des entwickelten Analysesystems zum Entscheidungsprozess bei der Optimierung der Anwendung und Nutzererlebnisses. Im Hintergrund sowie im Konzept werden die Grundlagen und das Vorgehen bei der Entscheidungsfindung unter Verwendung des Analysesystems (DSS) vorgestellt. Der Maßstab, inwieweit das Analysesystem bereits ein Entscheidungstreffen unterstützen

könnte, lässt sich nicht beantworten, da dies von spezifischen Fragen und Problemstellungen abhängt. In der Firma Janitza wurde das Analysesystem zur Abdeckung der Anforderung nach der Untersuchung der beliebten und weniger beliebten Funktionen umgesetzt (siehe Abschnitt 4.1.2). Da das System noch nicht im vollen Betrieb eingesetzt ist, konnte die tatsächliche Interaktion der Kunden noch nicht untersucht werden. In Abschnitt 5.4 werden die nächsten Schritte beschrieben, um die Anforderungen in der Firma weiter zu erfüllen.

Die Stichprobe ist also unzureichend, um die benötigten Beweise über den Beitrag zum Entscheidungsprozess zu sammeln. Da ein Verlauf und eine Verteilung über die Zeit für die interessierenden Metriken dargestellt werden, ist es zudem erforderlich, mehr Daten über die Zeit zu sammeln, um aussagekräftige Schlussfolgerungen zu ziehen. Die aufgestellte Hypothese ist somit weder bestätigt noch widerlegt. Das Ergebnis der Arbeit lässt sich daher nicht eindeutig interpretieren.

5.2 Auswertung

Im Nachhinein könnten die Methoden, Hypothese sowie die auftretenden Probleme und Risiken bei der Arbeit diskutiert und ausgewertet werden.

Erstens werden die Gefährdungen der internen Validität betrachtet werden. Dazu gehört die geringe statische Aussagekraft, die durch geringe Stichprobe und kleine Teststärken bedingt ist. Eine weitere Gefährdung besteht in der formulierten Hypothese, dass das Analysesystem zur Entscheidungsfindung bei der Optimierung der Anwendung und des Nutzererlebnisses beitragen und helfen sollte. Diese ist von Anfang an stark mit der letzten Forschungsfrage verbunden. Diese Abhängigkeit und Zielsetzung hatten zur Folge, dass ohne die vollständige Beantwortung der Forschungsfrage FF4 die Hypothese auch nicht beantwortet werden konnte. Entweder sollte die Hypothese umformuliert werden, sodass auch andere Forschungsfragen mehr Gewichtung bekommen hätten, oder die Arbeit könnte sich nur auf die Evaluation des bereits bestehenden Analysesystems für die Messung des Einflusses auf den Entscheidungsprozess konzentrieren.

Die externe Validität ist einerseits durch den begrenzten Zeitraum und andererseits durch die Unmöglichkeit einer Auswertung ohne reale Kundendaten bedroht. Die Methoden wie die Befragung der Nutzer (Produktmanager) und die Systembeobachtung könnten unter diesen Bedingungen keine aussagekräftigen Daten liefern. Zudem ist die Wechselwirkung der Selektion die externe Validität gefährdet, weil es eine begrenzte und freiwillige Teilnahme am Testen des Systemes erfolgte.

Im Laufe der Arbeit stellte sich das Problem heraus, die passende Literatur zu finden. Entweder gab es keine relevante Information in den gefundenen Quellen oder die

aktuelle und thematisch passende Arbeiten waren nicht offen verfügbar. Diese Tatsache weist auf die unzureichende Grundlage und den Kenntnisstand in diesem Bereich hin und beeinträchtigt die externe Validität der Arbeit. Deswegen wurde viel Aufwand, besonders beim Konzept, darauf verwendet, die einzelnen Stücke der Informationen und Erkenntnisse in einem Modell zu kombinieren. Da die Arbeit sich mit mehreren Aspekten wie Datenerhebung, Metriken, Datenschutz, Visualisierung und Entscheidungsfindung beschäftigt, erforderte es auch viel Zeit, in jeder Thematik vertiefend den Stoff zu verarbeiten. Die Suche nach Literatur war nach dem vertiefenden Einblick effektiver, da die Suche nach konkreten Aspekten durchgeführt wurde.

Unter anderem könnten die Anwendbarkeit der Daten und Übertragbarkeit auf anderen Szenarien die externe Validität gefährdern, da das Konzept nur im Rahmen der Firma Janitza angewendet wurde. Weitere Risiken beziehen sich nach der internen Realisierung in der Firma und können die externe Validität beeinflussen. In der Realisierungsphase wurde die Migration der Daten, wenn sich die Struktur für die erfassten Daten ändert, nicht einbezogen. Ein mögliches Risiko besteht darin, dass für die Speicherung der Analytics Ergebnisse Firebase genutzt wird. Die Firma besitzt kostenpflichtig ein Abonnement, und somit sollten die Schreib-/Lesezugriffe auf die Kollektionen vernünftig genutzt werden. Wenn die Datenstruktur wechselt oder erweitert wird, werden nach dem Update des Interfaces nicht alle historischen Daten richtig verarbeitet. Dies erfordert eine Migration der Daten. Da zurzeit jedoch keine wirklichen Kundendaten verwendet werden, ist diese Aktion für die alten Daten nicht ganz sinnvoll. Für diese Arbeit wurden die Daten mit alter Struktur nicht gelöscht, sondern primitiv von einem Legacy-Interface zu den neuen Strukturen überführt, um eine mögliche Darstellung der historischen Werte anzeigen zu können.

Da das Ereignis-Tracking erst im Rahmen dieser Arbeit in der Firma umgesetzt wurde, befindet sich die Implementierung des Moduls nicht öffentlich im Hauptzweig. Dabei besteht das Risiko, dass die Anwendung weiterentwickelt wird und das Ereignis-Tracking Modul veraltet und inkompatibel bleibt. Die Schritte, um dies zu vermeiden und das Analysesystem weiter zu nutzen, werden im Abschnitt 5.4 beschrieben.

5.3 Weitere Ansätze

In dieser Arbeit wurde ein umfassendes und bereichsübergreifendes Analysesystem konzipiert, das verschiedene Aspekte wie Datenschutz, Datenerfassung, Metriken, Datenspeicherung, Visualisierung und Entscheidungsfindung berücksichtigt. In anderen Arbeiten könnten ähnliche Ansätze oder Ausarbeitungen zu einem einzelnen Aspekt des Themas separat gefunden werden. Beispielsweise setzte sich Ruppert in seiner Dissertation mit dem Einsatz von visuellen Analysesystemen für die Entscheidungsfin-

dung auseinander. Ähnlich wie in dieser Arbeit wurde die multikriterielle Analyse als Grundlage verwendet und durch den Ansatz der modellgetriebenen Entscheidungsunterstützungssysteme (DSS) umgesetzt. Darüber hinaus leitete er ein visuelles Modell für die Erkenntnisgewinnung ab, das Ähnlichkeiten mit dem im Abschnitt 3.6 vorgestellten Modell aufweist. Im Gegensatz zur aktuellen Arbeit wurden für die Validierung des visuellen Analysesystems Nutzertests als Auswertungsmethode eingesetzt. Die Entwurfsmethodik für das visuelle Analysesystem wurde erfolgreich auf sechs entscheidungsbezogene Szenarien angewendet. Insgesamt stellte die Arbeit von Ruppert verschiedene Komponenten visueller Analysesysteme vor, um textuelle und unstrukturierte Daten darzustellen. Ein Framework als ein gesamtes Analysesystem wurde jedoch nicht vorgestellt. (vgl. [Rup17])

Aus der qualitativen Studie von Akter konnte eine Literaturübersicht für die Konzipierung eines Analysesystems für die Entscheidungsfindung in der Industrie sowie ein sechs-schrittiger Prozess dafür abgeleitet werden (vgl. [Akt19]). Andere Arbeiten im Bereich Big Data wie [Isl17] und [Moh20] haben sich ebenfalls mit der Einteilung von Frameworks für die gesamte Datenverarbeitung auseinandergesetzt. Die im Konzept dieser Arbeit vorgestellten Schritte, die in den Schichten der Architektur aufgeteilt sind (siehe Abschnitt 3.1), spiegeln ein ähnliches Vorgehen wider, wie es in den genannten Arbeiten präsentiert wird. Allerdings bieten diese Arbeiten keine spezifischen Informationen zur Implementierung des Systems, sondern schlagen lediglich eine logische Aufteilung für ein Framework vor.

Wie bereits in den vorherigen Abschnitten erwähnt wurde, bleibt die Forschungsfrage FF4 zum Entscheidungsprozess zur Optimierung der Anwendung unbeantwortet. Dieser Aspekt könnte als Grundlage für zukünftige Studien dienen.

Die Forschungslücke dieser Arbeit liegt in der Messung und Validierung der Ergebnisse, um den Beitrag des entwickelten Analysesystems zur Entscheidungsfindung zu bewerten. Ein möglicher Ansatz für weitere Forschungen könnte im Vergleich des Beitrags zur Entscheidungsfindung zwischen dem entwickelten Analysesystem und einer Drittanbieterlösung bestehen. Des Weiteren könnten Nutzertests durchgeführt werden, um das Verhalten der Nutzer bei der Interaktion mit dem System und bei der Interpretation der Daten nachzuvollziehen. Anschließend könnten Befragungen der Nutzer und deren Feedback mithilfe eines Formulars gesammelt werden. Die Fragen und wichtigen Punkte zur Auswertung könnten in zukünftiger Arbeit weiter ausgearbeitet werden. Zudem ist es notwendig, eine passende Skala und Unterscheidung zu formulieren, um den Maßstab und Beitrag zur Entscheidungsfindung zu differenzieren.

5.4 Nächste Schritte

Aktuell ist das Analysesystem, wie im Abschnitt 4.6.4 erwähnt, mit einem Feature Toggle sowie einer begrenzten Anzahl von Nutzern im Betrieb. Ein solcher Einsatz des Systems ist für die Entscheidungsfindung und das Erreichen der Ziele in der Firma unzureichend. Um die Ergebnisse umfassender zu validieren, ist eine breitere Datenerhebung erforderlich. Dies könnte durch die Erweiterung der Nutzung des Analysesystems auf eine größere Anzahl von Nutzern erfolgen. Als erster Schritt könnte die Erweiterung der Nutzerliste auf alle internen Nutzer ausgerichtet werden. Zudem erfordert es die Zusammenführung auf Hauptzweige in der Entwicklung, um die Nutzung der Benutzeroberfläche nicht nur im Entwicklermodus zu ermöglichen. Danach könnten die Auslastung sowie die neuen Ergebnisse ausgewertet werden. Als nächstes könnte die weitere Veröffentlichung in die Produktion realistische und wertvolle Informationen über die Nutzer und deren Interaktion mit der Anwendung liefern.

5.5 Ausblick

Die Arbeit hat dazu beigetragen, ein Modell unter Verwendung von Datenerhebungsmethoden, datenschutzkonformer Datenverarbeitungsmethoden sowie einer übersichtlichen Visualisierung zu entwickeln. Diese Kenntnisse können in Zukunft dazu dienen, bei weiteren Forschungen Unterstützung zu bieten. Das entwickelte System bietet eine allgemeine Vorgehensweise für die Entwicklung eines internen Analysesystems an.

In zukünftigen Arbeiten könnte das vorgestellte Analysesystem weiter auf den Einsatz und die Integration in anderen Bereichen geprüft werden. Die Ergebnisse der Nutzung könnten ebenfalls in weiteren Forschungen beobachtet und ausgewertet werden.

Literaturverzeichnis

[Ade10] Aden, Timo: *Google Analytics: implementieren, interpretieren, profitieren*, Hanser, München, 2., aktualisierte und erw. aufl Aufl. (2010)

[Akt19] Akter, Shahriar; Bandara, Ruwan; Hani, Umme; Fosso Wamba, Samuel; Foropon, Cyril und Papadopoulos, Thanos: Analytics-based decision-making for service systems: A qualitative study and agenda for future research. *International Journal of Information Management* (2019), Bd. 48: S. 85–95, URL https://www.sciencedirect.com/science/article/pii/S0268401218312696

[Alb23] Alby, Tom: Popular, but hardly used: Has Google Analytics been to the detriment of Web Analytics? (2023): S. 304–311, URL https://dl.acm.org/doi/10.1145/3578503.3583601

[Alh16] Alhlou, Feras; Asif, Shiraz und Fettman, Eric: *Google Analytics Breakthrough: From Zero to Business Impact*, John Wiley & Sons (2016), google-Books-ID: jyQ6rgEACAAJ

[Amo22] Amort, Matthias; Arens, Stephan; Bruckmann, Jan-Friedrich; Fischer, Hans-Jörg; Fischer, Franz-Alois; Helfrich, Marcus; Schmittmann, Jens M. und Supernok-Kolbe, Marcel: *Nachhaltigkeit und Recht*, Fachmedien Recht und Wirtschaft (2022), google-Books-ID: IHOdEAAAQBAJ

[AS19] Ahmad Sabri, Ily Amalina; Man, Mustafa; Abu Bakar, Wan Aezwani Wan und Mohd Rose, Ahmad Nazari: Web Data Extraction Approach for Deep Web using WEIDJ. *Procedia Computer Science* (2019), Bd. 163: S. 417–426, URL https://www.sciencedirect.com/science/article/pii/S1877050919321635

[Aze19] Azeroual, Otmane; Saake, Gunter und Abuosba, Mohammad: ETL Best Practices for Data Quality Checks in RIS Databases. *Informatics* (2019), Bd. 6(1): S. 10, URL https://www.mdpi.com/2227-9709/6/1/10, number: 1 Publisher: Multidisciplinary Digital Publishing Institute

[Bre12] Brewer, Eric: CAP twelve years later: How the "rules"have changed (2012), Bd. 45(2): S. 23–29, conference Name: Computer

[Bur05] Burkhard, Remo Aslak: Towards a Framework and a Model for Knowledge Visualization: Synergies Between Information and Knowledge Visualization, in: Sigmar-Olaf Tergan und Tanja Keller (Herausgeber) *Knowledge and Information Visualization: Searching for Synergies*, Lecture Notes in

Computer Science, Springer, Berlin, Heidelberg (2005), S. 238–255, URL https://doi.org/10.1007/11510154_13

[Can23] Canay, Özkan und Kocabiçak, Ümit: An innovative data collection method to eliminate the preprocessing phase in web usage mining. *Engineering Science and Technology, an International Journal* (2023), Bd. 40: S. 101360, URL https://www.sciencedirect.com/science/article/pii/S221509862300037X

[Che12] Chen; Chiang und Storey: Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly* (2012), Bd. 36(4): S. 1165, URL http://www.jstor.org/stable/10.2307/41703503

[Chr23] Chris, Cahill: The Future of Product Usage Data - 2023 Report. *Revenera Blog* (2023), URL https://www.revenera.com/blog/software-monetization/product-usage-data/

[Dah22] Dahr, Jasim Mohammed; Hamoud, Alaa Khalaf; Najm, Ihab Ahmed und Ahmed, Mohammed Imad: IMPLEMENTING SALES DECISION SUPPORT SYSTEM USING DATA MART BASED ON OLAP, KPI, AND DATA MINING APPROACHES (2022), Bd. 17

[Doc23] Docs, MDN Web: Navigator: sendBeacon() method - Web APIs | MDN (2023), URL https://developer.mozilla.org/en-US/docs/Web/API/Navigator/sendBeacon

[Fah11] Fahrner, Ulrich: *Die Explorative Datenanalyse als Lern- und Erkenntniswerkzeug*, doctoralthesis, Universität Augsburg (2011)

[Gau22] Gauer, Oliver: *Datenschutz Grundlagen*, HERDT-Verlag für Bildungsmedien GmbH, 1. ausgabe, januar 2022 Aufl. (2022)

[Gay18] Gayed, Jeremy: Open Source: Declarative Tracking for React Apps. *Medium* (2018), URL https://open.nytimes.com/introducing-react-tracking-declarative-tracking-for-react-apps-2c76706b

[GD14] Gkoulalas-Divanis, Aris; Loukides, Grigorios und Sun, Jimeng: Publishing data from electronic health records while preserving privacy: A survey of algorithms. *Journal of Biomedical Informatics* (2014), Bd. 50: S. 4–19, URL https://www.sciencedirect.com/science/article/pii/S1532046414001403

[Hal10] Haller, Heiko; Hartwig, Markus und Liedtke, Arne: *Google Analytics & Co: Methoden der Webanalyse professionell anwenden*, Addison-Wesley, München (2010)

[Has13] Hassler, Marco: *Web Analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren*, mitp, Heidelberg, 3. aufl Aufl. (2013)

[Hä08] Härting, Niko: Datenschutz im Internet: Wo bleibt der Personenbezug? *Computer und Recht* (2008), Bd. 24(11): S. 743–748, URL https://www.degruyter.com/document/doi/10.9785/ovs-cr-2008-743/html, publisher: Verlag Dr. Otto Schmidt

[Ima20] Imasheva, Baktagul; Azamat, Nakispekov; Sidelkovskiy, Andrey und Sidelkovskaya, Ainur: The Practice of Moving to Big Data on the Case of the NoSQL Database, Clickhouse (2020): S. 820–828

[Isl17] Islam, Osama; Alfakeeh, Ahmed und Nadeem, Farrukh: A Framework for Effective Big data Analytics for Decision Support Systems. *International Journal of Computer Networks And Applications* (2017), Bd. 4(5): S. 1, URL http://www.ijcna.org/Manuscripts/IJCNA-2017-0-13.pdf

[jan23] Energiemanagement (EnMS) - Janitza® (2023), URL https://www.janitza.de/loesungen/energiemanagement-enms.html

[Jay17] Jayamalini, K. und Ponnavaikko, M.: Research on web data mining concepts, techniques and applications (2017): S. 1–5, URL https://ieeexplore.ieee.org/abstract/document/8186676

[Jyo17] Jyothi, Padma; Bonthu, Sridevi und Prasanthi, B.V.: A Study on Raise of Web Analytics and its Benefits. *International Journal of Computer Sciences and Engineering* (2017), Bd. 5: S. 59–64

[Kai10] Kaiser, Thomas: *Google Analytics: Erfolgskontrolle für Webseiten ; [so messen und optimieren Sie den Erfolg Ihrer Website ; so setzen Sie Google Analytics technisch perfekt und effektiv ein ; ermitteln Sie, woher Ihre Besucher kommen und was sie tun ; die besten Auswertungsstrategien für Ihre Webanalysedaten]*, Know-how ist blau, Franzis-Verl, Poing (2010)

[Kam20] Kamps, Ingo und Schetter, Daniel: Web-Analyse (Web-Analytics) – messen, analysieren und entscheiden, in: Ingo Kamps und Daniel Schetter (Herausgeber) *Performance Marketing: Der Wegweiser zu einem mess- und steuerbaren Online-Marketing – Einführung in Instrumente, Methoden und Technik*, Springer Fachmedien, Wiesbaden (2020), S. 165–189, URL https://doi.org/10.1007/978-3-658-30912-1_10

[Kei10] Keim, Daniel; Kohlhammer, Jörn; Ellis, Geoffrey und Mansmann, Florian: *Mastering the Information Age Solving Problems with Visual Analytics*, Eurographics Association (2010), URL https://diglib.eg.org:443/xmlui/handle/10.2312/14803, accepted: 2016-02-15T09:19:14Z

[Ket22] Ketchum, Russell: Prepare for the future with Google Analytics 4. *Google* (2022), URL https://blog.google/products/marketingplatform/analytics/prepare-for-future-with-google-analytics-4/

[Kik05] Kiker, Gregory A.; Bridges, Todd S.; Varghese, Arun; Seager, Thomas P. und Linkov, Igor: Application of multicriteria decision analysis in environmental decision making. *Integrated Environmental Assessment and Management* (2005), Bd. 1(2): S. 95–108, URL https://onlinelibrary.wiley.com/doi/abs/10.1897/IEAM_2004a-015.1, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1897/IEAM_2004a-015.1

[Kir20] Kirsh, Ilan und Joy, Mike: Splitting the Web Analytics Atom: From Page Metrics and KPIs to Sub-Page Metrics and KPIs (2020): S. 33–43, URL https://dl.acm.org/doi/10.1145/3405962.3405984

[Kit18] Kitchens, Brent; Dobolyi, David; Li, Jingjing und Abbasi, Ahmed: Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. *Journal of Management Information Systems* (2018), Bd. 35(2): S. 540–574, URL https://www.tandfonline.com/doi/full/10.1080/07421222.2018.1451957

[Kne21a] Kneuper, Ralf: *Datenschutz für Softwareentwicklung und IT: Eine praxisorientierte Einführung*, Springer, Berlin, Heidelberg (2021), URL https://link.springer.com/10.1007/978-3-662-63087-7

[Kne21b] Kneuper, Ralf: Technische und organisatorische Gestaltung des Datenschutzes, in: Ralf Kneuper (Herausgeber) *Datenschutz für Softwareentwicklung und IT: Eine praxisorientierte Einführung*, Springer, Berlin, Heidelberg (2021), S. 131–169, URL https://doi.org/10.1007/978-3-662-63087-7_6

[Kom16] Komorowski, Matthieu; Marshall, Dominic C.; Salciccioli, Justin D. und Crutain, Yves: Exploratory Data Analysis, in: MIT Critical Data (Herausgeber) *Secondary Analysis of Electronic Health Records*, Springer International Publishing, Cham (2016), S. 185–203, URL https://doi.org/10.1007/978-3-319-43742-2_15

[Kre22] Krebs, Heinz-Adalbert und Hagenweiler, Patricia: Verfahren zur Durchführung der Anonymisierung, in: Heinz-Adalbert Krebs und Patricia Hagenweiler (Herausgeber) *Datenanonymisierung im Kontext von Künstlicher Intelligenz und Big Data: Grundlagen – Elementare Techniken – Anwendung*, Springer Fachmedien, Wiesbaden (2022), S. 125–138, URL https://doi.org/10.1007/978-3-658-37588-1_9

[Kum20] Kumar, Vikas und Ogunmola, Gabriel Ayodeji: Web Analytics for Knowledge Creation: A Systematic Review of Tools, Techniques, and Practices. *International Journal of Cyber Behavior, Psychology and Learning* (2020), Bd. 10(1): S. 1–14, URL http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJCBPL.2020010101

[Kü20] Kühling, Jürgen: Gesundheitsdatenschutzrecht im Zeitalter von „Big Data“. *Datenschutz und Datensicherheit - DuD* (2020), Bd. 44(3): S. 182–188, URL https://doi.org/10.1007/s11623-020-1248-6

[Lau18] Laux, Helmut; Gillenkirch, Robert M. und Schenk-Mathes, Heike Y.: *Entscheidungstheorie*, Springer, Berlin, Heidelberg (2018), URL http://link.springer.com/10.1007/978-3-662-57818-6

[Lee17] Lee, Sukwon; Kim, Sung-Hee und Kwon, Bum Chul: VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* (2017), Bd. 23(1): S. 551–560, URL https://

ieeexplore.ieee.org/abstract/document/7539634, conference Name: IEEE Transactions on Visualization and Computer Graphics

[Lti20] Ltifi, Hela; Benmohamed, Emna; Kolski, Christophe und Ayed, Mounir Ben: Adapted Visual Analytics Process for Intelligent Decision-Making: Application in a Medical Context. *International Journal of Information Technology & Decision Making* (2020), URL https://www.worldscientific.com/worldscinet/ijitdm, publisher: World Scientific Publishing Company

[Luc19] Luckie, Matthew; Beverly, Robert; Koga, Ryan; Keys, Ken; Kroll, Joshua A. und Claffy, K: Network Hygiene, Incentives, and Regulation: Deployment of Source Address Validation in the Internet (2019): S. 465–480, URL https://dl.acm.org/doi/10.1145/3319535.3354232

[McG23] McGuirk, Mike: Performing web analytics with Google Analytics 4: a platform review. *Journal of Marketing Analytics* (2023), URL https://doi.org/10.1057/s41270-023-00244-4

[Mec17] Mechel, Christian: *Ökoeffizienzanalyse zum Vergleich heterogener Unternehmen*, Springer Fachmedien, Wiesbaden (2017), URL http://link.springer.com/10.1007/978-3-658-14692-4

[Mei15] Meixner, Oliver und Haas, Rainer: Wissensmanagement und Entscheidungstheorie. *scholars-Titel ohne Reihe* (2015), URL https://elibrary.utb.de/doi/book/10.24989/9783990304761

[Mil18] Miller, Kristen; Mosby, Danielle; Capan, Muge; Kowalski, Rebecca; Ratwani, Raj; Noaiseh, Yaman; Kraft, Rachel; Schwartz, Sanford; Weintraub, William S und Arnold, Ryan: Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. *Journal of the American Medical Informatics Association* (2018), Bd. 25(5): S. 585–592, URL https://academic.oup.com/jamia/article/25/5/585/4598305

[Moh20] Mohamed, Azlinah; Najafabadi, Maryam Khanian; Wah, Yap Bee; Zaman, Ezzatul Akmal Kamaru und Maskat, Ruhaila: The state of the art and taxonomy of big data analytics: view from new big data framework. *Artificial Intelligence Review* (2020), Bd. 53(2): S. 989–1037, URL https://doi.org/10.1007/s10462-019-09685-9

[Nam22] Nambiar, Athira und Mundra, Divyansh: An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing* (2022), Bd. 6(4): S. 132, URL https://www.mdpi.com/2504-2289/6/4/132, number: 4 Publisher: Multidisciplinary Digital Publishing Institute

[Nan19] Nanda, Adhish; Gupta, Swati und Vijrania, Meenu: A Comprehensive Survey of OLAP: Recent Trends (2019): S. 425–430, URL https://ieeexplore.ieee.org/abstract/document/8822203

[Nar08] Narayanan, Arvind und Shmatikov, Vitaly: Robust De-anonymization of Large Sparse Datasets (2008): S. 111–125, URL https://ieeexplore.ieee.org/abstract/document/4531148, iSSN: 2375-1207

[Pal20] Palanisamy, Sowndarya und SuvithaVani, P.: A survey on RDBMS and NoSQL Databases MySQL vs MongoDB (2020): S. 1–7, URL https://ieeexplore.ieee.org/abstract/document/9104047, iSSN: 2329-7190

[Pal21] Palomino, Fryda; Paz, Freddy und Moquillaza, Arturo: Web Analytics for User Experience: A Systematic Literature Review (2021): S. 312–326

[Pap19] Papp, Stefan; Weidinger, Wolfgang; Meir-Huber, Mario; Ortner, Bernhard; Langs, Georg und Wazir, Rania: Handbuch Data Science, in: *Handbuch Data Science*, Carl Hanser Verlag GmbH & Co. KG (2019), S. 1–15, URL https://www.hanser-elibrary.com/doi/10.3139/9783446459755.fm

[Pat13] Patel, Neil: How Netflix Uses Analytics To Select Movies, Create Content, & Make Multimillion Dollar Decisions. *Neil Patel* (2013), URL https://neilpatel.com/blog/how-netflix-uses-analytics/

[Pet22] Petrlic, Ronald; Sorge, Christoph und Ziebarth, Wolfgang: *Datenschutz: Einführung in technischen Datenschutz, Datenschutzrecht und angewandte Kryptographie*, Springer Fachmedien, Wiesbaden (2022), URL https://link.springer.com/10.1007/978-3-658-39097-6

[Pop18] Popovič, Aleš; Hackney, Ray; Tassabehji, Rana und Castelli, Mauro: The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers* (2018), Bd. 20(2): S. 209–222, URL https://doi.org/10.1007/s10796-016-9720-4

[Pow02] Power, Daniel: Decision Support Systems: Concepts and Resources for Managers. *Faculty Book Gallery* (2002), URL https://scholarworks.uni.edu/facbook/67

[Res21] Research und Markets: Global Marketing Analytics Markets to 2026: Monitoring and Reacting to Shifting Customer Preferences (2021), URL https://www.prnewswire.com/news-releases/global-marketing-analytics-markets-to-2026-monitoring-and-reacting-to-s
html

[Roy20] Roy, Rita und Giduturi, Apparao: Survey on Pre-Processing Web Log Files in Web Usage Mining. *Int. J. Adv. Sci. Technol.* (2020)

[Rup17] Ruppert, Tobias: Visual Analytics to Support Evidence-Based Decision Making (2017)

[Saa88] Saaty, Thomas L.: What is the Analytic Hierarchy Process? (1988): S. 109–121

[Sac16] Sacha, Dominik; Senaratne, Hansi; Kwon, Bum Chul; Ellis, Geoffrey und Keim, Daniel A.: The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*

(2016), Bd. 22(1): S. 240–249, URL http://ieeexplore.ieee.org/document/7192716/

[Sch15] Schäfer, Axel und Schöttker-Königer, Thomas: Deskriptive Statistik: Beschreiben, Ordnen, Zusammenfassen – so verschaffe ich mir einen Überblick meiner Daten, in: Axel Schäfer und Thomas Schöttker-Königer (Herausgeber) *Statistik und quantitative Methoden für Gesundheitsfachberufe*, Springer, Berlin, Heidelberg (2015), S. 27–63, URL https://doi.org/10.1007/978-3-662-45519-7_3

[Sch20] Schwitter, Nicole und Liebe, Ulf: Going Digital: Web data collection using Twitter as an example (2020)

[See23] Seenivasan, Dhamotharan: ETL (Extract, Transform, Load) Best Practices. *International Journal of Computer Trends and Technology* (2023), Bd. 71: S. 40–44

[Sit17] Sitanggang, Imas Sukaesih; Ginanjar, Asep Rahmat; Syukur, Muhamad; Trisminingsih, Rina und Khotimah, Husnul: Integration of spatial online analytical processing for agricultural commodities with OpenLayers (2017): S. 167–170, URL http://ieeexplore.ieee.org/document/8167127/

[SR19] Sanchez-Rola, Iskander; Dell'Amico, Matteo; Kotzias, Platon; Balzarotti, Davide; Bilge, Leyla; Vervier, Pierre-Antoine und Santos, Igor: Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control (2019): S. 340–351, URL https://dl.acm.org/doi/10.1145/3321705.3329806

[Sta99] Stahlknecht, Peter und Hasenkamp, Ulrich: *Einführung in die Wirtschaftsinformatik*, Springer-Lehrbuch, Springer Berlin Heidelberg, Berlin, Heidelberg (1999), URL http://link.springer.com/10.1007/978-3-662-06903-5

[Ste46] Stevens, S. S.: On the Theory of Scales of Measurement. *Science* (1946), Bd. 103(2684): S. 677–680, URL https://www.science.org/doi/10.1126/science.103.2684.677

[Sto82] Stodolsky, David: Steven L. Alter: Decision support systems: Current practice and continuing challenges. Reading, Massachusetts: Addison-Wesley Publishing Co., 1980, 316 pp. *Behavioral Science* (1982), Bd. 27(1): S. 91–92, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830270109, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bs.3830270109

[Swe02] Sweeney, Latanya: k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY | International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (2002), URL https://www.worldscientific.com/doi/abs/10.1142/S0218488502001648

[Thi20] Thies, Laura Friederike; Jandt, Silke; Knote, Robin und Söllner, Matthias: Konfliktäre Anforderungen an smarte persönliche Assistenten. *Datenschutz und Datensicherheit - DuD* (2020), Bd. 44(9): S. 573–578, URL https://doi.org/10.1007/s11623-020-1327-8

[Upp22] Uppal, Tanya; Srivastava, Saumitya und Saini, Kavita: Web Development Framework : Future Trends (2022): S. 2181–2184, URL https://ieeexplore.ieee.org/abstract/document/10074105

[Vet23] Vettor, Robert: Vergleich der relationalen und NoSQL-Daten - .NET (2023), URL https://learn.microsoft.com/de-de/dotnet/architecture/cloud-native/relational-vs-nosql-data

[Wai09] Waisberg, Daniel und Kaushik, Avinash: Web Analytics 2.0: Empowering Customer Centricity (2009)

[Wan21a] Wang, Thomas: You May Not Know Beacon (2021), URL https://xgwang.me/posts/you-may-not-know-beacon/#data-size-limit

[Wan21b] Wang, Xin und Xing, Yujuan: Research on Web Log Data Mining Technology Based on Optimized Clustering Analysis Algorithm (2021): S. 6–11, URL https://ieeexplore.ieee.org/abstract/document/9694105

[Web15] Weber, Jonathan: Collecting Data from Mobile Apps, in: Jonathan Weber (Herausgeber) *Practical Google Analytics and Google Tag Manager for Developers*, Apress, Berkeley, CA (2015), S. 221–230, URL https://doi.org/10.1007/978-1-4842-0265-4_13

[Wit08] Witt, Bernhard C.: *Datenschutz kompakt und verständlich*, Vieweg, Wiesbaden (2008), URL http://link.springer.com/10.1007/978-3-8348-9442-7

[Wit23] Witt, Erik: Unload Beacon Reliability: Benchmarking Strategies for Minimal Data Loss. *Speed Kit* (2023), URL https://speedkit.com/blog/unload-beacon-reliability-benchmarking-strategies-for-minimal-data-loss

[Zul22] Zulhusni, Muhammad: IBM: Tech trends that continue to play a crucial role in 2023. *Tech Wire Asia* (2022), URL https://techwireasia.com/2022/12/ibm-tech-continues-to-play-a-crucial-role-in-2023/

[Zum12] Zumstein, Darius: Web Analytics (2012), URL https://sonar.ch/global/documents/302498

Abbildungsverzeichnis

Tabellenverzeichnis

Listings

.

A Anhang 1

B Anhang 2