



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Bachelorarbeit

Max Morten Schomann

**Ein Portal zur Visualisierung von Open Data
mit dem Play Framework**

*Fakultät Technik und Informatik
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science
Department of Computer Science*

Max Morten Schomann

**Ein Portal zur Visualisierung von Open Data
mit dem Play Framework**

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung

im Studiengang Bachelor of Science Angewandte Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Olaf Zukunft
Zweitgutachter: Prof. Dr. Philipp Jenke

Eingereicht am: 18. November 2015

Max Morten Schomann

Thema der Arbeit

Ein Portal zur Visualisierung von Open Data mit dem Play Framework

Stichworte

Offene Daten, Transparenzportal, Open Data Portal, Visualisierung, Play Framework

Kurzzusammenfassung

Diese Bachelorarbeit befasst sich mit frei verwendbaren offenen Daten, insbesondere des Hamburger Transparenzportals. Um die demokratische Meinungs- und Willensbildung zu fördern, wird ein verbesserter Zugang zu offenen Daten entworfen. Dazu wird ein Überblick über die Funktionen vorhandener Open Data Portale gegeben und daraus werden die Anforderungen an ein verbessertes Portal entwickelt. Das entwickelte Konzept sieht vor, die Benutzer die Inhalte anreichern zu lassen und bei Analysen zu unterstützen. Schwerpunkte bei der prototypischen Umsetzung mit dem Play Framework sind eine schnellere Erfassung der Inhalte und erstellbare Visualisierungen.

Max Morten Schomann

Title of the paper

A portal for visualizing open data with the Play Framework

Keywords

Open Data, Hamburg's Transparenzportal, Open Data Portal, Visualization, Play Framework

Abstract

This bachelor thesis deals with freely usable open data, in particular data from Hamburg's Transparenzportal. In order to encourage the democratic opinion-forming, an improved access to open data is developed. For this purpose an overview of the functions from existing open data portals is given and used to work out the necessary functions for an improved portal. The concept provides users with functions to enrich the content and to assist in analysis. Emphasis in the prototypical implementation with the Play Framework is a faster comprehension of content and creatable visualizations.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	1
1.2. Zielsetzung	2
1.3. Aufbau	3
2. Grundlagen	4
2.1. Visualisierung	4
2.1.1. Informationsvisualisierungen und Infografiken	4
2.1.2. Anforderungen	5
2.1.3. Wahrnehmungspsychologie	5
2.1.4. Gefahren	6
2.2. Visualisierungsarten für Daten	6
2.2.1. Word Cloud	6
2.2.2. Word Tree	7
2.2.3. Vorschaubilder	8
2.3. Web Technologien	9
2.3.1. JSON	9
2.3.2. REST	9
2.3.3. Play Framework	10
2.3.4. CKAN	11
2.4. Zusammenfassung	13
3. Open Data	14
3.1. Definition	14
3.2. Verwendung	15
3.3. Offene Daten und Visualisierungen	16
3.4. Open Data Portal	17
3.4.1. Funktionen	17
3.4.2. Engage	19
3.5. Offene Daten in Deutschland	20
3.5.1. Hamburger Transparenzgesetz	20
3.5.2. Hamburger Transparenzportal	20
3.6. Zusammenfassung	24
4. Analyse	25
4.1. Szenario	25

4.2.	Anwendungsfälle	26
4.3.	Anforderungen	33
4.3.1.	Funktionale Anforderungen	33
4.3.2.	Nichtfunktionale Anforderungen	35
4.4.	Ausgrenzungen	36
4.5.	Zusammenfassung	36
5.	Konzeption	37
5.1.	System	37
5.1.1.	Systemkomponenten	37
5.1.2.	Systemkontext	38
5.2.	Architektur	39
5.2.1.	Komponentenschnitt	39
5.2.2.	Komponenten	40
5.2.3.	Fachliche Komponenten	40
5.2.4.	Technische Komponenten	46
5.2.5.	Nichtfunktionalen Anforderungen	46
5.3.	Zusammenfassung	47
6.	Umsetzung	48
6.1.	Realisierungsumfang	48
6.2.	Realisierung	49
6.2.1.	Suche	49
6.2.2.	Dateien	50
6.2.3.	UserMetadaten	52
6.2.4.	Visualisierungen	52
6.2.5.	Konzeptionelle Anforderungen	54
6.3.	Evaluation	55
6.3.1.	Probleme	55
6.3.2.	Zielerfüllung	56
6.4.	Zusammenfassung	57
7.	Schluss	58
7.1.	Zusammenfassung	58
7.2.	Erweiterungen	59
7.3.	Fazit	60
A.	CD Inhalt	61
	Abbildungsverzeichnis	62
	Tabellenverzeichnis	63
	Literaturverzeichnis	64

1. Einleitung

Data is the new oil?

No: Data is the new soil.

[McC10, David McCandless, 2010]

1.1. Motivation

In den Aktenschränken der öffentlichen Verwaltung liegen große Mengen an Daten. Open Data umfasst als Schlagwort die Öffnung dieser Daten zu offenen Daten. Beispiele für Arten von offenen Daten sind Statistiken, Umweltdaten, Geodaten, Baupläne und öffentliche Beschlüsse, nicht dazu gehören personenbezogene Daten.

Die Öffnung erfolgt um Transparenz über die Tätigkeiten und Entscheidungsprozesse der öffentlichen Verwaltung herzustellen. Damit soll das Vertrauen der Bürger in die Politik erhöht werden. Zusätzlich sollen offene Daten Innovationen ermöglichen. Beispielsweise durch das Entdecken neuer Möglichkeiten die Daten zu verstehen und zu interpretieren. Das kann durch die Verknüpfung der offenen Daten mit anderen offenen Daten, internen Firmendaten oder privat erhobenen Daten erfolgen.

Im Juni 2012 wurde das Hamburgische Transparenzgesetz von der Hamburger Bürgerschaft verabschiedet. Zielsetzung des Gesetzes ist es:

durch ein umfassendes Informationsrecht die (...) vorhandenen Informationen unter Wahrung des Schutzes personenbezogener Daten unmittelbar der Allgemeinheit zugänglich zu machen und zu verbreiten, um über die bestehenden Informationsmöglichkeiten hinaus die demokratische Meinungs- und Willensbildung zu fördern und eine Kontrolle des staatlichen Handelns zu ermöglichen.

[Bü12, HmbTG § 1 Abs. 1]

Das Hamburger Transparenzportal als zentraler Zugang zu den freigegebenen Daten startete Ende 2014. Im April 2015 waren über 28.000 verschiedene Daten auffindbar. Die am häufigsten vertretenen Dateitypen sind mit Beispielen der Inhalte in Tabelle 1.1 aufgeführt.

1. Einleitung

Typ	Anzahl	Beispielinhalte
HTML	15.277	Links die hauptsächlich auf Informationssysteme der Bezirksversammlungen verweisen. Dort sind öffentliche Beschlüsse in Textform oder als PDF lesbar.
PDF	13.827	Baugenehmigungen, öffentliche Pläne und Verwaltungsvorschriften.
zip	1.375	Meistens Statistiken des Statistikamt Nord bestehend aus CSV, PDF und Excel Dateien.
Excel	1.313	Zumeist Statistiken des Statistikamt Nord.
CSV	1.032	Zum einen Statistiken zum anderen Messergebnisse, wie beispielsweise Radioaktivitätsmessungen.

Tabelle 1.1.: Inhalte des Hamburger Transparenzportals mit ihrer Anzahl am 15. April 2015

Das vorhandene Portal bietet hauptsächlich Möglichkeiten zum Suchen und Herunterladen einzelner Dokumente. Dies ist ideal um dem Bürger die Kontrolle des staatlichen Handelns zu ermöglichen. Die demokratische Meinungs- und Willensbildung wird auf dem Portal hingegen wenig gefördert, da der Bürger auf dem Portal nicht mit anderen Bürgern in Kontakt treten kann.

1.2. Zielsetzung

Ziel dieser Arbeit ist es, einen verbesserten Zugang zu den offenen Daten zu entwerfen. Einem technisch unerfahrenen Bürger soll die Bewertung, Analyse und Diskussion der vorhandenen Informationen erleichtert werden.

Es soll die Möglichkeit eröffnet werden, nicht nur die vorhandenen offenen Daten passiv zu konsumieren, sondern auch aktiv die offenen Daten anzureichern. Dazu sollen die Benutzer Inhalte bewerten, kommentieren, verbessern und diskutieren können. Zudem sollen die Benutzer bei der Analyse unterstützt werden, zum Beispiel durch erstellbare Visualisierungen. Außerdem soll die Zusammenarbeit mehrerer Benutzer und Benutzerinnen möglich sein.

Der erste Kontakt mit den offenen Daten Hamburgs besteht zur Zeit meistens darin, die Daten herunterzuladen und in einem externen Programm zu öffnen. Daher wird sich der, im Rahmen dieser Bachelorarbeit, zu erstellende Prototyp auf eine bessere Erkundung der Inhalte konzentrieren. Dazu ist angedacht, die gesuchten Inhalte, die als HTML, PDF, CSV oder Excel Dateien vorliegen nicht nur mit Textauszügen darzustellen, sondern bereits eine Vorschau auf die enthaltenen Inhalte anzuzeigen.

1.3. Aufbau

In Kapitel 1 wird die Relevanz und der Hintergrund des Themas erläutert. Zudem wird die Zielsetzung dieser Arbeit beschrieben. Darauf folgend wird das methodische Vorgehen vorgestellt.

Das folgende Kapitel 2 erläutert grundlegende, für diese Arbeit relevante, Begrifflichkeiten. Hierzu gehören Visualisierungsmöglichkeiten und eingesetzte Web Technologien.

Anschließend wird in Kapitel 3 eine Definition für Open Data vorgestellt. Darauf folgt ein Überblick über die Funktionen von Open Data Portalen. Abschließend wird das Hamburger Transparenzportal vorgestellt.

Es folgt die Analyse in Kapitel 4. Anhand eines Szenarios werden Anwendungsfälle entwickelt und aus diesen werden die Anforderungen abgeleitet.

In Kapitel 5 folgt die Konzeption. Hier werden die Erkenntnisse der ersten Kapitel mit den Anforderungen kombiniert und die daraus resultierende Architektur und die wichtigsten Komponenten werden vorgestellt.

Im Anschluss folgt die Beschreibung der Realisierung in Kapitel 6. Hier wird zuerst der Umfang der Umsetzung dargestellt. Darauf folgen die wichtigsten Schritte und Entscheidungen der Realisierung. Anschließend wird die resultierende Realisierung bewertet.

Abschließend wird in Kapitel 7 die Arbeit zusammengefasst, mögliche Erweiterungen werden vorgestellt und es wird ein Fazit gezogen.

2. Grundlagen

Dieses Kapitel stellt grundlegende, für diese Arbeit relevante, Begrifflichkeiten vor.

Zunächst werden Visualisierungen und deren Zweck vorgestellt. Anschließend werden Anforderungen an Visualisierungen beschrieben. Mögliche zu bedenkende Gefahren von Visualisierungen schließen den Abschnitt ab.

Darauf folgt eine Beschreibung von für diese Arbeit relevanten Visualisierungsarten. Dazu gehören die Word Cloud, Word Tree und eine Dokumentenvorschau Darstellung.

Abschließend werden in dieser Arbeit verwendete Web Technologien vorgestellt.

2.1. Visualisierung

Visualisierungen werden genutzt, um Daten in einer visuell erfassbaren Form darzustellen. Die Darstellung erfolgt meistens, um die Zusammenhänge von Daten besser verstehen und analysieren zu können. So sind zum Beispiel Trends und Ausreißer grafisch leicht zu entdecken.

Nach Zuiderwijk et al. [ZJD14] basieren Visualisierungen auf den Erkenntnissen verschiedener Forschungsrichtungen, wie Statistik, grafischem Design, Informatik, Psychologie und Mensch-Computer Interaktion. Weiter heißt es, Visualisierung sei ein Prozess, der das Sammeln, Sortieren, Filtern und Strukturieren von Daten umfasse.

Wichtige Interaktionstechniken für Visualisierungen sind laut Schoeneberg und Pein [SP14, Kapitel 14.3.2]: Vergleichen, Sortieren, Filtern, Hervorheben, Verdichten, Zoomen, Skalieren und Details auf Abruf.

2.1.1. Informationsvisualisierungen und Infografiken

Schoeneberg und Pein [SP14, Kapitel 14.2.2.1] definieren Informationsvisualisierungen als interaktive, mit Computer Unterstützung erstellte, visuelle Darstellungen. Ziel sei es, die menschliche Wahrnehmung zu unterstützen. Die Darstellungen sollen dabei helfen Daten besser zu verstehen und verarbeiten zu können.

Weiterhin bemerken die Autoren, dass in Informationsvisualisierungen abstrakte Daten dargestellt werden. Beispiele für abstrakte Daten sind unter anderem Texte, Prozesse und

Beziehungen. Diese hätten keine unmittelbaren Entsprechungen in der Natur. Um die abstrakten Daten wahrnehmbar zu machen, müssten daher in Visualisierungen den Daten Charakteristiken wie Form oder Farbe zugeordnet werden. Als Beispiel führen die Autoren den menschlichen Körper an und die Darstellung mit Hilfe der Computertomographie.

Darüber hinaus bemerken Schoeneberg und Pein [SP14, Kapitel 14.2.2.1] eine Abgrenzung von Infografiken zu Informationsvisualisierungen. So seien Infografiken manuell erstellte Darstellungen. Sie seien oft optisch ansprechend gestaltet, stellten jedoch meist nur wenige Informationen dar und müssten bei jeder Datenänderung manuell neu erstellt werden.

Informationsvisualisierungen hingegen ließen sich algorithmisch erstellen, sodass neue Daten automatisch eingefügt werden könnten. Durch die automatische Erzeugung könnten viele Daten angezeigt werden. Jedoch sind die Ergebnisse optisch oft verbesserungswürdig, betonen die Autoren.

In dieser Arbeit liegt der Hauptfokus auf Informationsvisualisierungen.

2.1.2. Anforderungen

Expressivität, Effektivität und Angemessenheit sind nach Schoeneberg und Pein [SP14, Kapitel 14.2.2.3] Anforderungen an Visualisierungen. Expressivität oder Ausdrucksfähigkeit sei grundlegend und bezeichne eine unverfälschte Wiedergabe der tatsächlich vorhandenen Informationen. Effektivität gebe Aufschluss über die Fähigkeit das enthaltene Wissen für den Betrachter leicht aufnehmbar darzustellen. Angemessenheit erfordere ein akzeptables Kosten-Nutzen Verhältnis bei der Erstellung einer Visualisierung.

Ergänzend zeigen die Autoren [SP14, Kapitel 14.2.2.2] zwei Verwendungszwecke von Visualisierungen auf. Zum einen sei dies die explorative Analyse, bei der noch keine Hypothese über die Eigenschaften der Daten vorhanden sei. Hier werden Visualisierungsmöglichkeiten gesucht, die die Eigenschaften der Daten beleuchten. Die konfirmative Analyse hingegen sei hypothesengetrieben. Ziel sei es Visualisierungsmöglichkeiten zu finden, welche die Hypothese überprüfbar machten.

2.1.3. Wahrnehmungspsychologie

Eine kurze Einführung in die Wahrnehmungspsychologie bieten Schoeneberg und Pein [SP14, Kapitel 14.2.3]. Dort wird der Ablauf von der unbewussten zur bewussten Wahrnehmung und schließlich zur Wissensspeicherung dargestellt. Relevant sind für diese Arbeit die dargestellten Grenzen der menschlichen Wahrnehmung und wie diese in Visualisierungen berücksichtigt werden sollten. So erläutern die Autoren, dass das Arbeitsgedächtnis nur drei bis fünf

Bedeutungseinheiten gleichzeitig im Bewusstsein halten kann. Als Folge, dieses „begrenzt aufnahmefähigen bewussten Arbeitsspeichers“ schlagen die Autoren vor, die Aufnahmeleistung durch bereits bekannte mentale Modelle zu reduzieren. So sollten Visualisierungen im Gedächtnis bekannte assoziative Erinnerungsmuster nutzen. Als Beispiel führen die Autoren eine einheitliche und konsequente Farbwahl an.

2.1.4. Gefahren

Weiterhin weisen die Autoren [SP14, Kapitel 14.2.2.4] auf Gefahren von Visualisierungen hin. Zum einen sei dies die Manipulierbarkeit des Betrachters, zum Beispiel bei nicht objektiven Informationen oder durch das Weglassen von Informationen. Zum anderen gebe es die Gefahr der Mehrdeutigkeit und Fehlinterpretation. Schließlich gebe es noch die Gefahr, dass durch eine zu hohe Konzentration auf die optische Darstellung das Ergebnis missverständlich sei oder die Informationen falsch wiedergegeben werden könnten.

2.2. Visualisierungsarten für Daten

Es gibt eine Vielzahl von Darstellungsarten. Im folgenden werden drei Darstellungsarten vorgestellt. Diese wurden im Hinblick auf die Zielsetzung, der besseren Erkundung von offenen Daten ausgewählt.

2.2.1. Word Cloud

In einer Word Cloud, Tag Cloud oder Schlagwortwolke werden eine Menge von Wörtern auf einer Fläche angezeigt. Wörter werden hervorgehoben, in dem sie zum Beispiel größer oder in einer anderen Farbe dargestellt werden. Hervorgehoben wird meistens die Häufigkeit eines Wortes.

Schoeneberg und Pein [SP14, Kapitel 14.3.4.3] heben hervor, dass Word Clouds für Analysewerkzeuge eingesetzt werden können. Der Grund sei ihre Einfachheit, Lesbarkeit und die Fähigkeit große und unstrukturierte Datenmengen darzustellen.

Als Problem beschreiben die Autoren [SP14, Kapitel 14.3.4.3], dass in der Word Cloud Darstellung der ursprüngliche semantische Kontext verloren geht. Als Lösung schlagen sie vor Details per Interaktionsmöglichkeit abrufbar zu machen. Eine Möglichkeit sehen sie in dem Einsatz von Word Trees. Diese könnten den semantischen Kontext anzeigen und so dabei helfen die Relevanz und Zusammenhänge der Wörterfolgen besser einzuschätzen.

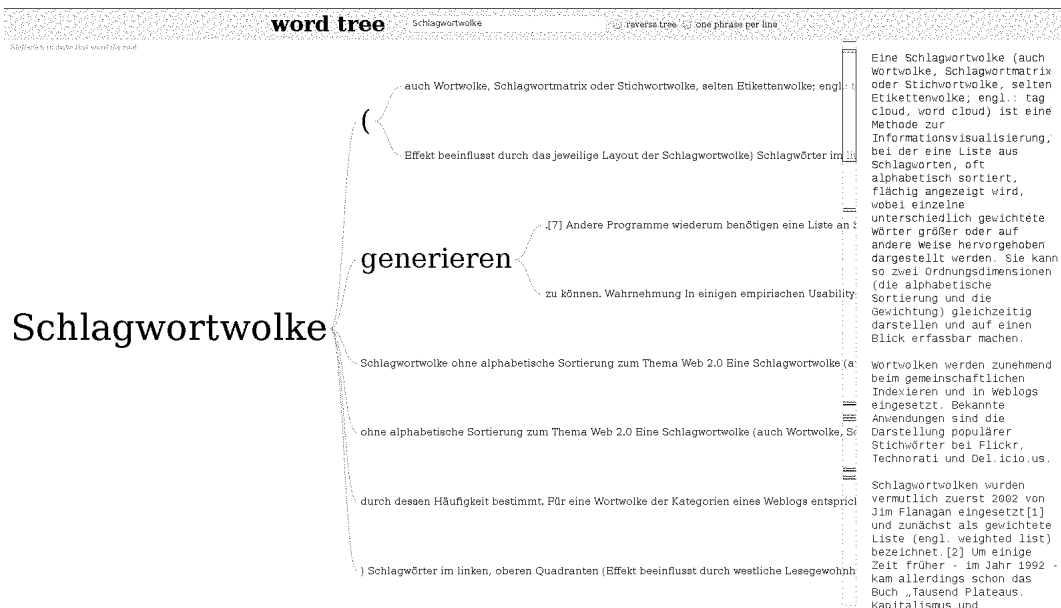


Abbildung 2.2.: Word Tree Beispiel [Dav15]

Schoeneberg und Pein [SP14, Kapitel 14.3.4.5] ergänzen, dass es sich anbietet Word Trees mit Word Clouds zu verbinden. So könne sich per Word Cloud eine Übersicht über den Text verschafft werden und per Word Tree die Benutzung der Wörter und deren Beziehungen besser analysiert werden.

2.2.3. Vorschaubilder

In [WFR⁺01] untersuchen Woodruff et al, ob Vorschaubilder von Suchergebnissen die Suchzeit reduzieren können. Dazu vergleichen sie die Suchzeit von getesteten Personen bei reinen Textergebnissen, Vorschaubildern und verbesserten Vorschaubildern. Letztgenannte wurden so modifiziert, dass vom Benutzer vorgegebene Worte in gelb markiert wurden. Die verbesserten Vorschaubilder erreichen laut den Autoren in ihrer Studie in allen Kategorien die beste Performance.

Eine ähnliche Technik wird bei der Suchmaschine Startpage¹ eingesetzt, dort können die Suchbegriffe auf der gefundenen Seite hervorgehoben werden.

¹<https://startpage.com/> - Zugriffsdatum: 31.08.2015

2.3. Web Technologien

Im nachfolgenden werden die später eingesetzten Web Technologien vorgestellt.

Dazu gehören, JSON zur Serialisierung von Daten, der REST Architekturstil, das Play Framework, sowie CKAN.

2.3.1. JSON

JavaScript Object Notation beziehungsweise JSON ist ein einfaches, von Programmiersprachen unabhängiges Textformat zur Serialisierung von strukturierten Daten.

Es basiert auf einem „object“ und einem „array“. Ein object ist eine ungeordnete Sammlung von Null oder mehr *Name:Wert* Paaren. Ein array ist eine geordnete Liste von Werten.

Der Name ist ein String. Die Werte können die primitiven Typen String, Nummer, Boolean oder null sein oder wiederum die strukturierten Typen object oder array. [Bra14, RFC 7159]

2.3.2. REST

Representational State Transfer bzw. REST ist ein Architekturstil.

REST Architekturen basieren auf Uniform Resource Identifiern (URI), die eindeutig eine Ressource identifizieren.

Die Ressourcen können mit den HTTP Methoden GET, POST, PUT und DELETE manipuliert werden. GET wird zum Abrufen von Ressourcen benutzt, PUT zum Ersetzen von Ressourcen, POST zum Erstellen eines neuen Eintrages und DELETE zum Löschen der angegebenen Ressource. Die anderen HTTP Methoden werden selten eingesetzt. Die Ressourcen werden häufig in Dateiformaten wie JSON oder XML repräsentiert.

REST ist eine Client-Server Architektur. Die Kommunikation läuft über eine Schnittstelle, dies dient der Entkoppelung. Zudem ist REST ein zustandsloses Protokoll, das heißt jede Anfrage des Clients muss alle Information beinhalten, die der Server zur Beantwortung benötigt. [LdK14, Kapitel 4.3]

2.3.3. Play Framework

Das Play Framework² ist ein seit 2007 entwickeltes Webframework.

Im Jahr 2009 wurde das Play Framework als Open Source Projekt unter der Apache Lizenz veröffentlicht. Es läuft auf der Java Virtual Maschine. Seit der Version 1.1 gibt es eine direkte Scala API.

2012 wurde für die Version 2.0 das Framework in Scala neu geschrieben. Es hat weiterhin eine Java und eine Scala API. [HBC13, LdK14]

Fähigkeiten

Play hat eine auf HTTP fokussierte API nach dem REST Stil. Die Browser der Benutzer sind die Clients. Sie kommunizieren ausschließlich über HTTP mit dem Server. Die Browser erhalten nur HTML, Javascript und CSS Programmcode. Jeglicher Scala Code wird auf dem Server ausgeführt.

Play übernimmt die typsichere Abbildung von zustandslosem HTTP auf das Objekt orientierte Model von Java, bzw. das Objekt orientierte und funktionale Model von Scala. Außerdem gibt es eine typsichere Template Syntax. Diese View Templates bestehen aus HTML Code und können ausführbaren Scala oder Java Code enthalten.

Außerdem gibt es eine integrierte Unterstützung zum Parsen von JSON.

Skalierbarkeit ist eins der Hauptfeatures des Frameworks. Es ist leicht möglich, skalierbare Web Anwendungen zu entwickeln. Ursprung dieser Skalierbarkeit sind das eingesetzte Prinzip der Zustandslosigkeit. Die Basis sind Scala und der Einsatz des Aktoren Frameworks Akka.

Integriert ist im Framework der JBoss Netty Webserver. Der Server ist nicht blockierend und kann eine Vielzahl von gleichzeitigen Anfragen mit einer kleinen Anzahl von Threads performant beantworten. [HBC13, LdK14]

Model-View-Controller

Das Play Framework basiert auf dem Model-View-Controller (MVC) Architektur Prinzip. [Sta09, Kapitel 7.6.1]

Zur Kommunikation mit den Clients wird HTTP genutzt. In der route.conf Konfigurationsdatei wird den Controller Methoden ein HTTP Pfad und eine HTTP Methode zugeordnet.

Für jede eingehende HTTP Anfrage wird in der Konfiguration der Routen nach einer passenden Controller Aktion gesucht. Der aufgerufenen Controller Code ruft, falls nötig, Methoden des Models auf. Das Model lädt oder verändert Daten aus dem Model bzw. einer Datenbank

²<https://www.playframework.com/> - Zugriffsdatum: 31.08.2015

und erzeugt eine Antwort. Diese Antwort wird vom Controller über ein View Template in HTML gewandelt und als HTTP Antwort zurückgesendet.

Es ist eine push basierende Architektur, das heißt der Controller bringt die Daten vom Model zum View. Das Zusammenspiel der MVC Komponenten ist in Abbildung 2.3 dargestellt. [HBC13, LdK14]

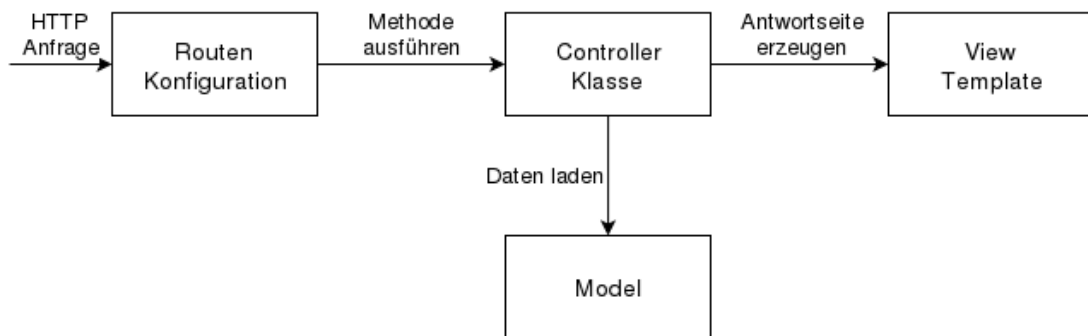


Abbildung 2.3.: Play Framework MVC nach [HBC13, page 18, Figure 2.2]

Einsatz

Produktiv eingesetzt wird das Play Framework unter anderem beim Sozialen Business Netzwerk LinkedIn und der Zeitung The Guardian. [HBC13, LdK14]

Das Play Framework wurde als Basis für das in dieser Bachelorarbeit entwickelte System gewählt, um die im Studium erworbenen Scala Kenntnisse zu vertiefen und einen Einblick in die Webentwicklung zu erhalten.

2.3.4. CKAN

CKAN³ ist eine Daten Management Lösung. CKAN wurde von der Open Knowledge Foundation als Open Source Software entwickelt.

Die Open Knowledge Foundation Deutschland ist ein gemeinnütziger Verein. Der Verein setzt sich für offene Daten, Transparenz und Partizipation bei öffentlichen Vorgängen ein.

CKAN ist der de facto Standard bei Open Data Portalen und wird bei vielen Portalen eingesetzt [Ass15d].

Seit 2014 wird die Weiterentwicklung von der CKAN Association betreut [Ass15b].

³<http://ckan.org/> - Zugriffsdatum: 31.08.2015

Funktionen

Laut der CKAN Webseite [Ass15c] bietet CKAN die folgenden Möglichkeiten:

- Veröffentlichung von Datensätzen
- Speicher für Roh- und Metadaten
- Suche
- API

Die Suche unterstützt eine Volltextsuche sowie eine unscharfe Suche, auch Fuzzy-Suche genannt. Zudem kann die Suche über Aspekte oder auf Aspekte eingeschränkt werden, wie z.B. Tags, Formate oder Lizenzen.

Metadaten

Alle veröffentlichten Datensätze haben eine eindeutige ID und verschiedene Metadaten. Standardmäßig gibt es die Metadatenfelder: *title*, *notes*, *tags*, *author* und *maintainer*. Zusätzlich gibt es die folgenden Felder: *groups* für Kategorien, *license* für einen Verweis auf die Lizenz der Daten und *url* welches auf weitere Informationen verlinken kann. Außerdem gibt es die *extra* Felder die z.B. die geographische Abdeckung des Datensatzes beschreiben können oder für projektspezifische Metadaten genutzt werden können.

Jeder Datensatz kann aus mehreren Ressourcen bestehen. Die Ressourcen werden jeweils mit einem Namen, der Größe, einem Dateiformat sowie einem Downloadlink versehen. [Ass15a]

Weitere Fähigkeiten

Für CKAN gibt es Erweiterungen und über Themes kann das Aussehen angepasst werden. CKAN kann Datensätze auf einer Karte darstellen, falls die geographische Abdeckung bekannt ist. Zudem können strukturierte Daten wie CSV oder Excel Dateien graphisch angezeigt werden. [Ass15c]

Die CKAN API [Ass15a] bietet viele der CKAN Funktionen an. Die API ist im Remote Procedure Call (RPC) Stil gehalten. Eine Suchanfrage mit dem Suchparameter „Gold“ hat beispielsweise die Form: http://demo.ckan.org/api/3/action/package_search?q=Gold. Der Server beantwortet die Anfrage des Clients mit JSON.

2.4. Zusammenfassung

Visualisierungen sind hilfreich um große Datenmengen schneller und besser zu verstehen. Informationsvisualisierungen stellen große Mengen an Daten bildlich dar, um die Auswertung zu erleichtern. Sie können bei der Analyse von offenen Daten helfen.

Bei der Auswahl und Erstellung muss auf die Expressivität, Effektivität und Angemessenheit geachtet werden. Zudem muss die eingeschränkte menschliche Wahrnehmung bedacht werden. Daher sind bereits bekannte Darstellungsarten gegenüber neuen und experimentellen vorzuziehen.

Für das Ziel, der besseren Erkundung von offenen Daten, sind die drei vorgestellten Darstellungsarten geeignet. Sie können die explorative Analyse des Benutzers unterstützen.

Word Clouds bieten durch ihre Fähigkeit große und unstrukturierte Datenmengen komprimiert darzustellen eine gute Übersicht über unterschiedlichste Daten. Als Ergänzung sind Word Trees sinnvoll, diese ermöglichen eine bessere Analyse des semantischen Kontextes.

Der Einsatz von verbesserten Vorschaubildern könnte hilfreich sein.

Bedacht werden muss, dass automatisiert erstellte Informationsvisualisierungen optisch oft verbesserungswürdig sind. Hier muss ein Kompromiss mit der Gefahr, der falschen Wiedergabe, durch eine zu hohe Konzentration auf die optische Darstellung, gefunden werden.

Für die spätere Umsetzung sind die Kenntnisse der CKAN Metadatenfelder und der CKAN API erforderlich. Bedacht werden muss die JSON Antwort der API. Außerdem muss der REST Stil für die Umsetzung mit dem Play Framework berücksichtigt werden. Zudem liefert das Model-View-Controller Prinzip die grundlegende Architektur der Umsetzung.

3. Open Data

Offene Daten sind Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können – maximal eingeschränkt durch Pflichten zur Quellennennung und „share-alike“. Davon ausgenommen sind personenbezogene Daten.

[e.V15, Open Knowledge Foundation Deutschland e.V., 2015]

Zentraler Gegenstand dieser Arbeit sind offene Daten.

In diesem Kapitel wird erläutert, was Open Data beziehungsweise offene Daten sind. Anschließend werden einige Projekte vorgestellt, die beispielhaft den Nutzen von offenen Daten erklären sollen.

Es folgt eine Übersicht über Probleme bei der Visualisierung von offenen Daten.

Darauf folgt ein Überblick über die Funktionen von Open Data Portalen. Im Anschluss wird die Entstehung des Hamburger Open Data Portals erläutert. Abschließend werden die Inhalte und Funktionen des Hamburger Transparenzportals vorgestellt.

3.1. Definition

Offene Daten sind Daten, die als modifizierbare und maschinenlesbare Daten verfügbar sind. Die Lizenz muss eine Verbreitung, Modifizierung und eine Kombination mit anderen Daten erlauben. Zusätzlich sollen die offene Daten von jedem genutzt werden können und dürfen, auch zu kommerziellen Zwecken. Außerdem sollen sie proaktiv, zeitnah und kostenlos bereitgestellt werden. [e.V15, DSH11, dI12]

Offene Daten umfassen nach Dietrich, Schulzki-Haddouti [DSH11] und der Open Knowledge Foundation [e.V15]:

- Geodaten, z.B. die Lage von Straßen, Grenzen und Gebäuden
- Finanzdaten, z.B. Haushalte
- Transportdaten, z.B. Fahrpläne oder Verkehrsinformationen

- Wetter- und Umweltdaten
- wissenschaftliche Daten, die in der Forschung erhoben werden
- Kulturdaten, z.B. von Museen oder Bibliotheken
- Statistiken
- Protokolle
- Gesetze und Verordnungen

Offene Daten können von jedem veröffentlicht werden. Quellen von offenen Daten sind laut Dietrich und Schulzki-Haddouti [DSH11] jedoch vor allem die internen Daten der öffentlichen Verwaltungen.

Daten die Betriebs- oder Geschäftsgeheimnisse enthalten oder die sicherheitsrelevant, datenschutzbedenklich oder personenbezogen sind fallen nicht unter offene Daten.

Ziel ist es, Innovationen zu ermöglichen, z.B. durch Analysen, Kombinationen von Daten oder die Entwicklung neuer Dienstleistungen. Offene Daten sind zudem ein Baustein des Open Government Prinzips. Open Government soll die Demokratie stärken. Grundlage ist die Transparenz über das Regierungs- und Verwaltungshandeln durch offene Daten. Das Ziel ist es, die Bürger bei politischen Prozessen aktiver mitgestalten zu lassen, um so die Demokratie zu stärken und die Politikverdrossenheit abzubauen. [dI12]

3.2. Verwendung

Eine Vielzahl von Projekten bauen auf offenen Daten auf.

Es gibt Projekte die Politik nachvollziehbar machen wollen. Die 2007 gestartete Webseite „OpenCongress“¹ ist dafür ein Beispiel. Diese beschäftigt sich mit der Gesetzgebung in den USA. Die Webseite ermöglicht es Gesetzesvorhaben, Abstimmungen und das Verhalten der Kongressabgeordneten nachzuvollziehen. Darüber hinaus informiert die Website über Spenden an Abgeordnete, um mögliche Korruption aufzudecken.

Andere Projekte wollen die Verwendung von Steuergeldern transparenter darstellen, wie z.B. „OffenerHaushalt“². Auf der Webseite werden die Haushaltspläne des Bundes, einiger Länder und einiger Kommunen als Treemap dargestellt. So lassen sich die Hauptposten leicht erkennen und per Klick lassen sich die Ausgaben weiter aufschlüsseln. Die Daten müssen dazu in einem bestimmten maschinenlesbaren Format vorliegen.

¹<http://www.opencongress.org/> - Zugriffsdatum: 31.08.2015

²<http://bund.offenerhaushalt.de/> - Zugriffsdatum: 31.08.2015

Ein weiteres Projekt, welches die gleiche zugrunde liegende Software OpenSpending nutzt, ist „Farmsubsidy“³. Dort wird die Verteilung der EU Agrarsubventionen dargestellt. Die Daten sind entweder direkt von den jeweiligen EU Mitgliedsstaaten veröffentlicht worden oder wurden über Informationsfreiheitsanfragen beantragt. Auch hier müssen die Daten manuell in ein bestimmtes einheitliches Format gebracht werden.

Beispiel für eine kommunale Open Data Plattform in Deutschland ist „Frankfurt Gestalten“⁴. Dort können Bürger unter dem Motto „Bürger machen Stadt“ lokal-politische Entscheidungen in der Nachbarschaft verfolgen. Dazu werden Ortsbeiratsanträge aus dem Parlamentsinformationssystem der Stadt Frankfurt händisch einem Bezirk zugeordnet. Hinzu kommen noch Polizei- und Baustellenmeldungen, sowie lokale Nachrichten. Außerdem können die Bürger eigene Ideen einbringen und diskutieren.

In Hamburg gibt es das Projekt „Wir bauen Hamburg“⁵. Manuell werden hier Informationen zu Bebauungsplanverfahren der Stadt Hamburg aus dem Amtsblatt zusammengetragen.

Neben diesen Beispielen gibt es noch viele weitere Projekte die auf offenen Daten aufbauen. So gibt es eine Vielzahl von Darstellungen, die z.B. lokale Unfall-, Kriminalitäts-, Bildungs-, Luftqualitäts- oder Gesundheitsdaten darstellen. Die meisten Projekte konzentrieren sich auf wenige Quelldaten. Diese Darstellungen werden oft als manuell erstellte Visualisierungen präsentiert, also als Infografiken [siehe 2.1.1].

3.3. Offene Daten und Visualisierungen

Graves und Hender [GH13] stellen fest, dass ein Teil der potentiellen Nutzer von offenen Daten nicht in der Lage ist, die offenen Daten zu sammeln, zu bearbeiten, zusammenzuführen und zu verstehen. Die Hauptursache sehen die Autoren in einem fehlendem technischen Verständnis, fehlenden Programmierkenntnissen und mangelnden Datenmanagement Erfahrungen. Als Lösung schlagen die Autoren einfach zu erstellende Visualisierungen vor. Zudem zeigen sie auf, dass ein „explorierender Mechanismus zum Navigieren der Daten und der Metadaten in diesen Visualisierungen“ nötig ist.

Als ein Problem existierender Visualisierungslösungen beschreiben die Autoren, dass die Werkzeuge meist nur eine fertige Visualisierung ausgeben. Der Nutzer hätte so nicht die Möglichkeit zu erforschen, welche Daten, aus welchen Quellen und in welcher Weise die Daten verwendet wurden. Zudem müsse der Nutzer, falls er eine ähnliche Visualisierung mit kleinen Änderungen erstellen will, von vorne anfangen.

³<http://farmsubsidy.openspending.org/> - Zugriffsdatum: 31.08.2015

⁴<http://www.frankfurt-gestalten.de/> - Zugriffsdatum: 31.08.2015

⁵<http://www.wir-bauen-hamburg.de/> - Zugriffsdatum: 31.08.2015

Die Probleme bei Visualisierungen lassen sich laut den Autoren in drei Phasen einteilen:

- In der Erstellungsphase müssten verschiedene Daten mit verschiedenen Formaten zusammengebracht und bearbeitet werden.
- In der Erkundungsphase sei es problematisch zu erkennen, woher die Daten stammten und wie sie bearbeitet wurden.
- In der Wiederverwendungsphase sei es problematisch die zugrunde liegenden Daten anzupassen oder andere Aspekte zu ändern.

3.4. Open Data Portal

Open Data Portale stellen die Rohdaten der öffentlichen Stellen proaktiv, gebündelt und mit Metadaten zur Verfügung.

Die ersten Länder, die ihre Daten in dieser Form geöffnet haben, waren die USA und Großbritannien. US Präsident Barack Obama hat 2009 in einer seiner ersten Amtshandlung verfügt, Daten der Regierung in dem data.gov⁶ Portal zu veröffentlichen. In Großbritannien gibt es seit dem Januar 2010 das data.gov.uk⁷ Portal.

In Deutschland startete 2013 das GovData⁸ Portal einen Pilotbetrieb. Bund, Länder und Kommunen betreiben meist jeweils eigene IT Systeme und haben verschiedene Gesetze. GovData soll die dezentral vorgehaltenen Daten mit einer einheitlichen Metadatenstruktur auffindbar machen. Ende 2015 soll der Regelbetrieb starten. [uKG15b]

3.4.1. Funktionen

Die wichtigsten Elemente eines idealen Open Data Portals beschreiben Zuiderwijk et al. in einem Literaturüberblick [ZJD14].

Laut Zuiderwijk et al. bestehen Open Data Portale aus einem Zusammenspiel zwischen technischen und sozialen Systemen. Verschiedene Akteure, wie Gesetzgeber, Datenbereitsteller und Bürger, sowie verschiedene Werkzeuge müssten in einem Portal zusammengeführt werden.

Die Autoren fassen die in ihrem Überblick gesammelten Elemente eines idealen Open Data Portals in einer Abbildung [ZJD14, Fig. 2] zusammen. Tabelle 3.1 zeigt die Funktionen eines Open Data Portals für Datenbereitsteller, Tabelle 3.2 die Funktionen für Open Data Nutzer.

⁶<http://www.data.gov/> - Zugriffsdatum: 31.08.2015

⁷<http://data.gov.uk/> - Zugriffsdatum: 31.08.2015

⁸<https://www.govdata.de/> - Zugriffsdatum: 31.08.2015

3. Open Data

Daten erstellen oder sammeln
Daten speichern
Datensätze prüfen, zusammenstellen und sensitive Daten entfernen
Datensätze veröffentlichen
Anfragen nach Daten bearbeiten
Datensätze diskutieren
Analysieren, wie das Portal genutzt wird

Tabelle 3.1.: Open Data Portal Funktionen für Datenbereitsteller nach [ZJD14, Fig. 2]

Daten suchen
Anmelden
Sprache wählen, Metadaten übersetzen
Nach Datenlizenz suchen
Daten ansehen
Daten herunterladen
Daten anpassen, analysieren, erweitern, kombinieren und verbinden
Verbesserten Datensatz veröffentlichen und mit dem Originalen verbinden
Datensatz anfragen und andere Anfragen erfüllen
Tutorials nutzen
Daten visualisieren
Datensätze diskutieren
Verschiedene Aspekte der Datensätze bewerten

Tabelle 3.2.: Open Data Portal Funktionen für Datennutzer nach [ZJD14, Fig. 2]

Zusätzlich identifizieren sie vier Hauptelemente eines Open Data Portals:

1. Veröffentlichung der offenen Daten
2. Suchen und Betrachten der offenen Daten
3. Analysieren, Verbessern, Kombinieren, Verlinken und Visualisieren der Daten
4. Interpretation, Diskussion und Rückmeldung

Darüber hinaus erwähnen sie drei weitere hilfreiche Elemente eines Open Data Portals. Dazu zählen sie Tutorials, ein Qualitätsmanagement und mehrere Arten von Metadaten. Verschiedene Arten von Metadaten seien hilfreich, um Daten mit unterschiedlichen Metadaten Vokabulars besser verbinden zu können.

3.4.2. Engage

Alexopoulos et al [ALC14] stellten fest, dass die existierenden Open Data Portale nur selten mehr als die ersten zwei der vier Hauptelemente umfassen.

Um diese Lücke zu füllen, präsentieren die Autoren die Engage Plattform. Diese soll insbesondere „Web 2.0“ Funktionalitäten beinhalten. Dazu gehören Gruppenfunktionalitäten, um die Zusammenarbeit zu erleichtern. Außerdem Datenbearbeitungsmöglichkeiten, beispielsweise Daten- und Metadatenverbesserungsmöglichkeiten sowie das Hinzufügen, die Konvertierung und die Kombinationsmöglichkeit der Daten. Ergänzend sind eine verbesserte Datenmodellierung und ein verbesserter Umgang mit Metadaten vorgesehen. Dazu zählen auch verschiedene Metadaten Vokabulars.

Möglichkeiten zur Rückmeldung und Ideensammlung sollen vorhanden sein. Eine Bewertungsmöglichkeit der Datenqualität soll ebenfalls enthalten sein. Überdies soll das Einstellen von neuen Daten Teil der Engage Plattform sein. Ferner sollen Visualisierungsmöglichkeiten angeboten werden.

Auf dem umgesetzten Engage Portal⁹ waren im Sommer 2015 wenige aktuelle Daten. Datensätze werden manuell hoch geladen. Zu einem vorhandenen Datensatz können verbesserte Daten hoch geladen werden. Es wird empfohlen Daten manuell in einem externem Tool zu bearbeiten. Tabellenartige Datensätze sind filterbar und es lassen sich einfache Diagramme erstellen. Metadaten können verbessert werden. Bewertungen sind möglich. Es gibt ein Feedback Formular und es können Anfragen nach neuen Daten gestellt werden. Ein Wiki enthält einige Anleitungen und in Gruppen können sich Benutzer Nachrichten schicken. [Pro14]

⁹<http://www.engagedata.eu/> - Zugriffsdatum: 31.08.2015

3.5. Offene Daten in Deutschland

Vorläufer von offenen Daten waren die Informationsfreiheitsgesetze. Seit 2006 gibt es das Informationsfreiheitsgesetz in Deutschland. Es ermöglicht jedem Bürger Informationen von Bundesbehörden anzufragen. Einige Bundesländer haben eigene Informationsfreiheitsgesetze, Hamburg seit dem August 2006.

3.5.1. Hamburger Transparenzgesetz

Im Oktober 2011 veröffentlichte die Hamburger Volksinitiative „Transparenz schafft Vertrauen“ einen Gesetzesentwurf für ein Transparenzgesetz. Dieser Entwurf sollte die bisher nötigen Anfragen nach dem Informationsfreiheitsgesetz ablösen. Dazu sah der Entwurf eine umfassende Veröffentlichungspflicht interner Behördendaten vor. Bevor es zu einem Volksbegehren kam, übernahm die Hamburger Bürgerschaft den Entwurf des Transparenzgesetzes. Der Entwurf wurde im Juni 2012 verabschiedet und trat im Oktober 2012 in Kraft.

Ziel des Gesetzes ist es, die Kontrolle des staatlichen Handelns zu ermöglichen und die demokratische Meinungs- und Willensbildung zu fördern.

Eine Projektgruppe, bestehend aus Vertretern der Finanz-, Justiz- und Kulturbehörde, klärte die rechtlichen und organisatorischen Aspekte und begleitete die technische Umsetzung mit Hilfe der Firma Dataport, dem Hitec e.V. und dem Fraunhofer Institut. [LT14]

3.5.2. Hamburger Transparenzportal

Im Oktober 2014 startete das Hamburger Transparenzportal¹⁰ [LT14]. Es macht die Geheimhaltung zur erläuterungsbedürftigen Ausnahme und die Transparenz zur Regel.

Inhalte

Nach dem beschlossenen Hamburger Transparenzgesetz [Bü12, HmbTG § 3] werden folgende Informationsgegenstände auf dem Portal proaktiv veröffentlicht:

- Senatsbeschlüsse, Entscheidungen des Senats
- Mitteilungen des Senats an die Bürgerschaft, z.B. Anträge zum Beschluss von Gesetzen
- Beschlüsse, Protokolle und Anlagen aus öffentlichen Sitzungen, z.B. von den Bezirksversammlungen

¹⁰<http://transparenz.hamburg.de/> - Zugriffsdatum: 31.08.2015

3. Open Data

- Verträge der Daseinsvorsorge, die die Grundbedürfnisse der Bürger berühren, wie z.B. Verträge die die Wasser-, Energieversorgung, Bildungseinrichtungen oder die Datenverarbeitung für hoheitliche Tätigkeiten betreffen
- Pläne der Verwaltung, z.B. Organisations- oder Haushaltspläne
- Verwaltungsvorschriften
- Statistiken und Tätigkeitsberichte
- Gutachten und Studien
- Geodaten
- Umweltdaten, wie Messungen oder Erhebungen
- Baumkataster, ein Verzeichnis aller Stadtbäume
- öffentliche Pläne, z.B. Bebauungspläne, Flächennutzungspläne oder der Schulentwicklungsplan
- Baugenehmigungen, ausgenommen reine Wohnbebauung mit weniger als sechs Wohneinheiten
- Subventionen und Zuwendungen, über 1.000 Euro in einem Jahr
- wesentliche Unternehmensdaten städtischer Beteiligungen, z.B. die Vergütung der Geschäftsleitungen
- Verträge mit öffentlichem Interesse, wenn das Volumen über 100.000 Euro in einem Jahr beträgt und nur wenn die Veröffentlichung die wirtschaftlichen Interessen der Stadt nicht erheblich beeinträchtigt
- Dienstanweisungen

Zusätzlich gibt es eine weitere Klausel, die alle vergleichbaren Informationen von öffentlichem Interesse umfasst. Diese soll sicherstellen, dass zukünftige noch nicht abzusehende Informationen ebenfalls veröffentlicht werden. Weiterhin ist es möglich zusätzliche Informationen per Antrag anzufragen.

Die Informationen sollen nach dem Transparenzgesetz [Bü12, HmbTG § 10] unverzüglich, elektronisch, maschinell durchsuchbar und im Volltext veröffentlicht werden.

Die Nutzung und Verbreitung der Information soll frei seien. Dazu wurde die „Datenlizenz Deutschland – Namensnennung – Version 2.0“ gewählt [uKG15a]. Die Lizenz erlaubt die kommerzielle und nicht kommerzielle Nutzung. Es muss angezeigt werden, ob die Daten verändert wurden. Ein Quellenvermerk muss den originalen Datensatz verlinken, die Datenlizenz nennen und den Bereitsteller benennen.

Quellen

50 Liefersysteme, wie die Parlamentsdatenbank, sind automatisiert an das Portal angebunden. Zusätzlich ist eine Veröffentlichung von Daten auf dem Transparenzportal in den Geschäftsprozessen der Behörden integriert [LT14].

Ausgenommen von der Informationspflicht sind Organe der Rechtspflege, wie Gerichte und Strafverfolgungsbehörden, sowie öffentliche Stellen, bei der Wahrnehmung sicherheitsrelevanter Aufgaben und bei Vorgängen der Steuererhebung.

Öffentliche Belange sollen geschützt werden. So müssen keine Entwürfe und Verschluss-sachen veröffentlicht werden. Informationen, die die Beziehungen zu einem anderen Land betreffen oder die die innere Sicherheit nicht unerheblich gefährden müssen ebenfalls nicht veröffentlicht werden.

Personenbezogene Daten sind unkenntlich zu machen. Betriebs- und Geschäftsgeheimnisse sind nur zu veröffentlichen, wenn das Informationsinteresse das Geheimhaltungsinteresse überwiegt. [Bü12]

Formate

Die Informationen sollen in einem maschinenlesbaren Format vorliegen. Das Format soll verbreitet seien und ein offengelegter herstellerunabhängiger Standard sein.

Neben den häufig anzutreffenden Formaten aus Tabelle 1.1 - HTML, PDF, Excel, zip und CSV - gibt es vereinzelt Datensätze mit Ressourcen im ASCII, JPEG, RAR, TIFF, TXT oder XML Format. Zusätzlich gibt es noch die GML, WEB, WFS und WMS Formate.

Die Ressourcen im GML oder WEB Format werden meistens als Link auf ein bestehendes Angebot der Stadt Hamburg verwendet.

Das WFS (Web Feature Service) Format beschreibt eine Schnittstelle zum Zugriff auf die Geodaten des Geoinformationssystems der Stadt Hamburg. Beim Öffnen einer WFS Ressource wird das XML Schema der Schnittstelle angezeigt.

Das WMS (Web Map Service) Format beschreibt eine Schnittstelle zum Zugriff auf Auszüge von Landkarten des Landesbetriebs Geoinformation und Vermessung Hamburg. Bei WMS

3. Open Data

Ressourcen wird zusätzlich zu der XML Schnittstellenbeschreibung die entsprechende Karte auf dem Geoportal Hamburg verlinkt, dort kann die Karte online betrachtet werden.

Funktionalität

Das Portal selbst läuft auf 20 Server-Systemen in einer virtuellen Umgebung im Dataport Rechenzentrum. Als Betriebssystem dient Ubuntu. Neben den Softwarelösungen der Stadt läuft eine CKAN Instanz [siehe 2.3.4]. [LT14]

Das Transparenzportal ist im Sommer 2015 im Wesentlichen nur eine Suchmaschine. Bei den gefundenen Datensätzen können die Metadaten betrachtet werden und die enthaltenen Ressourcen heruntergeladen werden. Die Daten werden meist nicht in maschinenlesbaren Datenformaten bereitgestellt, sondern in unstrukturierten Dateiformaten wie PDF Dateien, welche nicht einfach auslesbar sind.

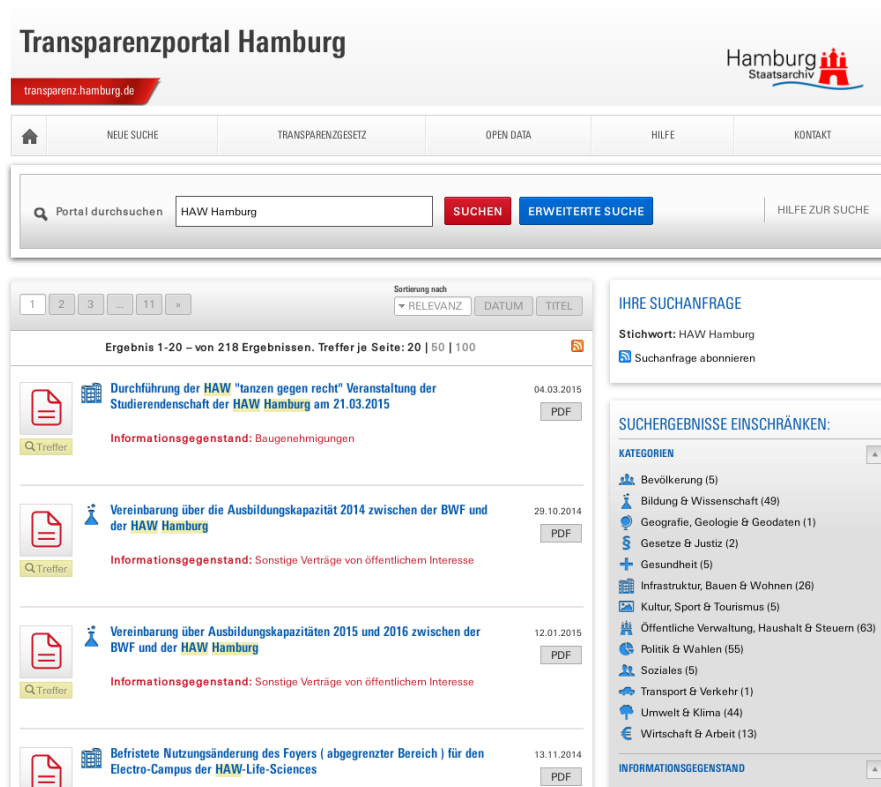


Abbildung 3.1.: Transparenzportal Hamburg [uHH15]

3.6. Zusammenfassung

Offene Daten sind frei verwendbare Daten.

In dieser Arbeit werden die offenen Daten der öffentlichen Verwaltungen betrachtet. Ziele bei der Öffnung von Daten sind Transparenz und Innovation.

Es gibt bereits einige Projekte die auf offenen Daten aufbauen. Einige haben das Ziel Politik nachvollziehbar darzustellen. Andere wollen die Lokalpolitik fördern. Infografiken bereiten ausgewählte Datensätze auf.

Ein Problem bei der Nutzung von offenen Daten ist, dass einigen der potenziellen Benutzer das entsprechende Know-how zur Nutzung fehlt. Für einen besseren Zugang sind einfach zu erstellende Visualisierungen nötig. Fertige Visualisierungen sollten zudem nachvollziehbar sein und die zugrunde liegenden Daten sollten anpassbar sein.

Quellen von offenen Daten sind Open Data Portale. Diese veröffentlichen die Daten und bieten Suchfunktionen. Ideal wäre es, wenn die Portale auch Analysen, Verbesserungen und Visualisierungen ermöglichen machen würden. Zusätzlich wünschenswert sind Möglichkeiten zur Interpretation und Diskussion.

Das Engage Portal ist ein erster Ansatz für ein verbessertes Open Data Portal.

Eine Volksinitiative gab den Anstoß für ein Hamburger Open Data Portal. Auf dem Hamburger Transparenzportal werden viele Daten proaktiv zur Verfügung gestellt. Allerdings ermöglicht das Portal nur eine Suche und das Herunterladen von Ressourcen.

4. Analyse

In diesem Kapitel werden anhand eines Nutzungsszenarios die konkreten Anforderungen an ein verbessertes Open Data Portal identifiziert und spezifiziert. Dazu werden die aus dem Szenario entwickelten Anwendungsfälle dargestellt. Im Anschluss werden aus den Anwendungsfällen die Anforderungen extrahiert.

4.1. Szenario

Ein Szenario ist eine Beschreibung der Interaktion des Benutzers mit dem System. Es dient dazu das Verhalten des Systems darzustellen und die Rollen der Benutzer zu beleuchten.

Das hier entwickelte Szenario basiert auf den Erkenntnissen des zweiten und des dritten Kapitels. Ziel ist es einen verbesserten Zugang zu offenen Daten zu finden, sodass mehr Bürger die offenen Daten nutzen.

Hauptbestandteil sind drei der vier Hauptelemente eines Open Data Portals aus Abschnitt 3.4.1: Suchen, Analysieren und Interpretieren. Das erste Element, der Veröffentlichungsprozess, wird nicht betrachtet.

Der in Abschnitt 3.3 als nötig erachtete Mechanismus zur explorierenden Navigation soll durch eine Visuelle Suche realisiert werden. Diese Visuelle Suche soll die Inhalte über eine Word Cloud Darstellung schneller erfassbar machen. Zur näheren Kontextanalyse soll eine Word Tree Darstellung auswählbar sein.

Allgemeines Szenario

Der Benutzer besucht das Portal um Informationen über ein bestimmtes Thema zu erhalten.

Anhand der Suchergebnisse entscheidet der Benutzer, welche Dokumente er sich intensiver anschauen möchte. Eventuell filtert er seine Suche nach bestimmten Kriterien, wie z.B. einer für ihn nutzbaren Lizenz.

Zusätzlich verwendet er die Visuelle Suche um einen ersten Blick auf die Inhalte der Suchergebnisse zu werfen. Nach einem ersten Einblick in die Daten, liest der Benutzer die vorhandenen Bewertungen des Datensatzes. Außerdem erfährt er was andere Benutzer von

dem Datensatz halten, ob die Daten vollständig sind, wofür sie nutzbar sind, wie die Qualität ist, ob es Eigenarten der Daten gibt und was andere Benutzer bereits mit den Daten erstellt haben. Möglicherweise gibt es auch Interpretationshilfen, falls der Datensatz in fachlichen Termen geschrieben ist die dem Benutzer nicht geläufig sind.

Anschließend beginnt der Benutzer mit der Analyse der Daten, die je nach Ziel des Benutzers unterschiedlich umfangreich ausfallen kann. Möglicherweise liest er das vorliegende Schriftstück oder er erstellt auf dem Portal bereits erste Visualisierungen.

Nachdem er die ersten eigenen Erkenntnisse gezogen hat, fängt er an die Daten auf dem Portal zu filtern, sortieren und bereinigen. Eventuell verbindet er die Daten mit selbst erhobenen Daten, Firmendaten oder anderen offenen Daten. Aus den kombinierten Daten erstellt er weitere Visualisierungen. Falls der Nutzer technische Erfahrungen oder Kenntnisse von Analysesoftware hat, kann er die Daten herunterladen und lokal bearbeiten.

Um andere an seinen Erkenntnissen teilhaben zu lassen bewertet der Nutzer die Daten und fasst seine Schlussfolgerungen zusammen. Zusätzlich kann er von den Rohdaten Verweise auf seine Ergebnisse, Visualisierungen, Analysen und selbst erhobene passende Daten setzen. Sollte er die Daten lokal bearbeitet haben, kann er diese Daten hoch laden und mit den Ursprungsdaten verbinden. Abschließend kann der Benutzer seine Ergebnisse über soziale Medien teilen.

Sollte der Benutzer zu wenige oder keine passenden Informationen finden, kann er den Portalbetreiber, die zuständige Stelle oder andere Benutzer nach entsprechenden Daten fragen.

Möchte der Nutzer mehrere Analysen verbinden, so hat er die Möglichkeit eigene Texte, Visualisierungen und Dokumente zu einer „Geschichte“ zu verbinden.

Zusätzlich bietet das Portal noch Interaktionsmöglichkeiten um die Benutzer zusammenzubringen, z.B. ein Forum für allgemeine Diskussionen oder Hilfestellungen, ein Wiki für Tutorials und ein Gruppenfindungssystem, um Ideengeber und Umsetzer oder Benutzer mit ähnlichen Interessen zusammenzubringen und deren Zusammenarbeit zu ermöglichen.

4.2. Anwendungsfälle

Aus dem Szenario lassen sich Benutzer und ein Portalbetreiber als Akteure identifizieren. Die Interaktionen dieser mit dem System sind die Anwendungsfälle des Systems.

Die beteiligten Akteure und die identifizierten Anwendungsfälle sind in dem Systemkontext in Abbildung 4.1 als Anwendungsfalldiagramm dargestellt.

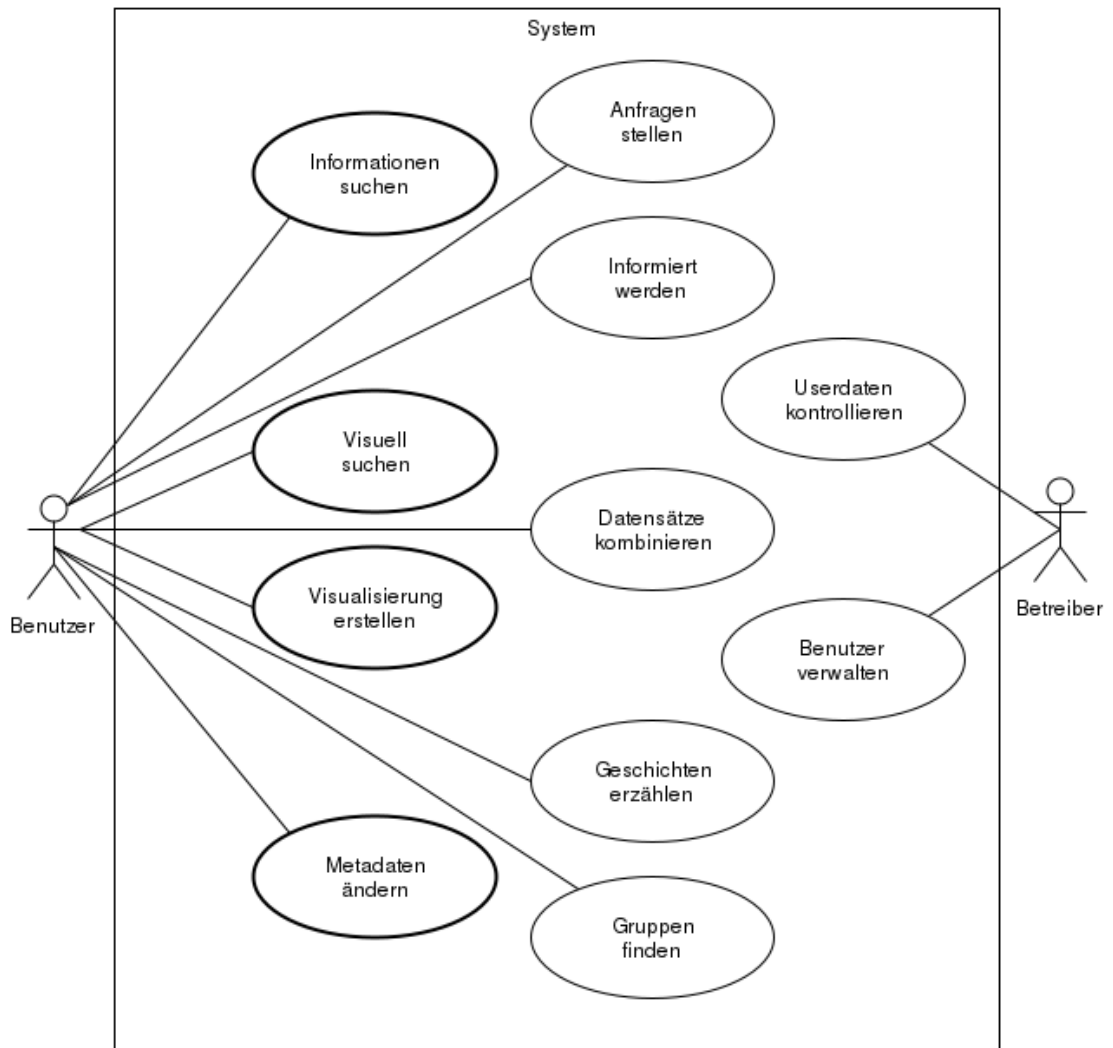


Abbildung 4.1.: Systemkontext

Im Folgenden werden die wichtigsten Anwendungsfälle genauer ausgeführt. Die weiteren Anwendungsfälle werden als User Story präsentiert. Um die Darstellung lesbarer zu gestalten werden mögliche technische Kommunikationsprobleme und Sicherheitsaspekte vernachlässigt.

4. Analyse

Titel	Informationen suchen
Ziel	Der Benutzer möchte sich über ein Thema informieren
Akteur	Bürger
Auslöser	Ein Benutzer sucht Informationen über ein Thema
Vorbedingung	Der Benutzer hat das Portal aufgerufen
Nachbedingung	Der Benutzer hat die gewünschten Informationen erhalten
Erfolgsszenario	<ol style="list-style-type: none"> 1. Der Benutzer gibt seine Suchbegriffe in das Suchfeld ein. 2. Das System liefert eine Liste von Suchtreffern zurück. Für jeden Treffer ist darin enthalten: eine Zusammenfassung des Inhaltes, der Herausgeber, Verweise auf anhängende Ressourcen und eine Kategorie und Tag Einordnung. 3. Der Benutzer wählt ein Trefferergebnis zur näheren Ansicht aus. 4. Das System zeigt alle verfügbaren Metadaten zu dem Ergebnis an. Für jede Ressource werden verschiedene Darstellungsmöglichkeiten angeboten. Zudem werden die Bewertungen, Notizen und Versionen von anderen Benutzern angezeigt. 5. Der Benutzer wählt eine Ressource zur Ansicht aus. 6. Das System zeigt die Ressource in der gewünschten Darstellung an, siehe auch Anwendungsfall „Visualisierung erstellen“.
Erweiterungen	<ol style="list-style-type: none"> 1. <ol style="list-style-type: none"> a) Der Benutzer filtert die Suche nach Kategorie, Dateiformat, Lizenz, Zeitraum und Ort. <ol style="list-style-type: none"> i. Das System berücksichtigt die Filteroptionen. b) Der Benutzer wählt die Visuelle Suche. <ol style="list-style-type: none"> i. Ausführung des Anwendungsfall „Visuell suchen“.
Fehlerfälle	<ol style="list-style-type: none"> 1. Das System findet keine passenden Informationen. Das System präsentiert eine Anfrage Maske, siehe User Stories.

Tabelle 4.1.: Anwendungsfall „Informationen suchen“

4. Analyse

Titel	Visuell suchen
Ziel	Der Benutzer möchte eine grafische Ansicht der Suchtreffer sehen
Akteur	Bürger
Auslöser	Der Benutzer sucht Informationen zu einem Thema
Vorbedingung	Der Benutzer hat das Portal aufgerufen
Nachbedingung	Der Benutzer hat die gewünschten Informationen erhalten
Erfolgsszenario	<ol style="list-style-type: none"> 1. Der Benutzer gibt seine Suchbegriffe in das Suchfeld ein und wählt die visuelle Suche. 2. Das System zeigt den Ressourceninhalt der Suchtreffer als Wort Cloud an. 3. Der Benutzer wählt einen Term einer Word Cloud aus. 4. Das System zeigt den dazugehörigen Word Tree mit dem selektierten Term an.
Erweiterungen	<ol style="list-style-type: none"> 2. <ol style="list-style-type: none"> a) Das System zeigt andere automatisch generierte Visualisierungen an. b) Das System zeigt eine Karte mit den Orten der gefundenen Ergebnissen an. 3. Der Benutzer wählt die normale Ergebnisansicht. <ol style="list-style-type: none"> a) Siehe Anwendungsfall „Informationen suchen“ Punkt 4.
Fehlerfälle	<ol style="list-style-type: none"> 1. Das System findet keine passenden Informationen. System präsentiert eine Anfrage Maske.

Tabelle 4.2.: Anwendungsfall „Visuell suchen“

4. Analyse

Titel	Visualisierung erstellen
Ziel	Der Benutzer will die Daten bearbeiten und Visualisierungen erstellen
Akteur	Bürger
Auslöser	Der Benutzer hat ein Datensatz gefunden, diesen möchte er grafisch analysieren
Vorbedingung	Der Datensatz liegt in einem darstellbaren Format vor
Nachbedingung	Der Benutzer hat eine Visualisierung erstellt
Erfolgsszenario	<ol style="list-style-type: none"> 1. Der Benutzer wählt den gewünschten Datensatz aus. 2. Das System stellt für CSV und Excel Dateien eine Tabellenansicht zur Verfügung. In dieser können die Daten sortiert, gefiltert und manipuliert werden. 3. Der Benutzer bearbeitet die Daten. 4. Das System stellt eine Ansicht zur Diagramm Erzeugung bereit. 5. Der Benutzer wählt die Art des zu erstellenden Diagramms und die dazugehörigen Parameter aus. 6. Das System stellt die Visualisierung dar und bietet die Möglichkeit diese in Sozialen Medien zu teilen.
Erweiterungen	<ol style="list-style-type: none"> 2. <ol style="list-style-type: none"> a) Das System bietet für PDF Dateien die Möglichkeiten Begriffe in der Datei zu markieren und die PDF Dateien als Bilddatei, PDF Datei oder auf dem Portal darzustellen. b) Das System bietet für zip Dateien die komprimierten Dateien zur Visualisierung an. c) Das System bietet für alle Dateiformate die Möglichkeit die Dateiinhalte auf dem Portal, sowie als Word Cloud und als Word Tree darzustellen. d) Das System zeigt bei mit Geodaten versehenen Datensätzen eine Kartenansicht. 6. Das System bietet an die geänderten Dateiinhalte und die Anzeigeparameter zu speichern und über einen bereitgestellten Link wieder zur Verfügung zu stellen.
Fehlerfälle	-

Tabelle 4.3.: Anwendungsfall „Visualisierung erstellen“

Titel	Metadaten ändern
Ziel	Der Benutzer verbessert die zu einem Datensatz gehörigen Informationen
Akteur	Bürger
Auslöser	Der Benutzer möchte die vorhandenen Metadaten verbessern oder neue Informationen hinzufügen
Vorbedingung	-
Nachbedingung	Das System bietet allen Benutzern die Möglichkeit die eingegebenen Änderungen zu betrachten
Erfolgsszenario	<ol style="list-style-type: none"> 1. Der Benutzer wählt einen Datensatz aus. 2. Das System bietet dem Benutzer die Möglichkeiten: <ul style="list-style-type: none"> • Metadaten anzuklicken, um in einem Änderungsfeld die Metadaten zu verbessern • Metadaten über ein Button hinzuzufügen • die Datenqualität auf einer Skala von 0 bis 10 zu bewerten • den Datensatz zu kommentieren, um z.B. auf Eigenarten der Daten hinzuweisen • Verlinkungen hinzuzufügen, auf ähnliche Daten oder Analysen des Datensatzes • neue Versionen der Daten hochzuladen, um Fehler zu bereinigen oder ein anderes Datenformat zur Verfügung zu stellen 3. Der Benutzer füllt die Felder aus die er speichern möchte. 4. Das System speichert die Daten und verlinkt sie vom Ursprungsdatensatz aus. Zusätzlich meldet das System dem Portalbetreiber das Feedback.
Erweiterungen	-
Fehlerfälle	-

Tabelle 4.4.: Anwendungsfall „Metadaten ändern“

User Stories

1. **Datensätze kombinieren:**

Ich, als Bürger, kann zwei Datensätze zu einer Visualisierung kombinieren.

2. **Gruppen finden:**

Ich, als Bürger, kann Bürger mit ähnlichen Interessen finden und mit ihnen gemeinsam an Projekten arbeiten.

3. **Informiert werden:**

Ich, als Bürger, werde informiert, wenn neue Informationen erscheinen die mich interessieren oder die meine Umgebung betreffen.

4. **Anfragen stellen:**

Ich, als Bürger, kann Anfragen stellen, ob bestimmte Daten veröffentlicht werden können.

5. **Geschichten erzählen:**

Ich, als Bürger, kann „Geschichten“ erzählen, indem ich mehrere Analyseergebnisse, Visualisierungen, Dokumente und eigene Inhalte miteinander verbinden kann.

6. **Userdaten kontrollieren:**

Ich, als Portalbetreiber, kann die von Benutzern generierten Inhalte kontrollieren und löschen. Zudem kann ich Benutzer warnen und sperren. Dies geschieht um mein Haftungsrisiko, für das Veröffentlichen von fremden Inhalten, zu minimieren.

7. **Benutzer verwalten:**

Ich, als Bürger, kann einen Benutzeraccount anlegen und meine Daten ändern. Der Benutzeraccount ermöglicht es mir Datensätze zu kommentieren, Daten hochzuladen, eigene Inhalte zu erstellen und an Gruppen teilzunehmen.

4.3. Anforderungen

Im Folgenden werden die Anforderungen an das System aufgestellt. Dies geschieht, um die einzelnen Funktionen des Systems präziser und besser überprüfbar darzustellen.

Die Anforderungen ergeben sich aus den einzelnen Schritten der Anwendungsfallsszenarios. Zur besseren Überprüfbarkeit sind die Anforderungen nummeriert.

4.3.1. Funktionale Anforderungen

Die Funktionalen Anforderungen beschreiben die Funktionalität des Systems, also welche Funktionen das System unterstützen muss. Sie sind nach den dazugehörigen Anwendungsfällen sortiert.

Anforderungen des Anwendungsfall „Informationen suchen“:

A-IS-1: Der Benutzer soll über eine Suchfunktion Datensätze finden können.

A-IS-2: Der Benutzer soll die Suche filtern können und zwar nach Kategorie, Dateiformat, Lizenz, Zeitraum und Ort.

A-IS-3: Das System zeigt die Suchtreffer mit einer Zusammenfassung des Inhaltes, dem Herausgeber und den dazugehörigen Ressourcen an.

A-IS-4: Das System zeigt in der Detailansicht: eine ausblendbare Metadatenübersicht, die vorhandenen Ressourcen mit verschiedenen Visualisierungsmöglichkeiten, Kommentare, Bewertungen, sowie Versionen anderer Benutzer.

Anforderungen des Anwendungsfall „Visuell suchen“:

A-VS-1: Das System muss die Ressourceninhalte der Suchtreffer als Word Cloud anzeigen können.

A-VS-2: Der Benutzer soll einen Term in der Word Cloud selektieren können. Dieser Begriff soll dann als Mittelpunkt eines Word Tree verwendet werden.

A-VS-3: Das System muss die Ressourcen als Word Tree darstellen können.

A-VS-4: Das System soll automatisch Visualisierungen erzeugen können.

Anforderungen des Anwendungsfall „Visualisierung erstellen“:

- A-V-1: Der Benutzer soll die Wahl zwischen verschiedenen Arten von einfach zu erstellenden Visualisierungen haben.
- A-V-2: Das System soll für die CSV und Excel Formate eine Tabellenansicht zur Verfügung stellen.
- A-V-3: Der Benutzer soll die bei der Visualisierung verwendeten Daten anpassen können (Sortieren, Filtern, Bearbeiten und Hinzufügen).
- A-V-4: Das System muss für die Daten in der Tabellenansicht verschiedene Visualisierungen anbieten, z.B. Balken-, Linien-, Säulen- oder Punktdiagramme mit veränderbaren Parametern.
- A-V-5: Das System soll PDF Dateien als Bild, PDF und auf dem Portal darstellen können.
- A-V-6: Das System soll zip Dateien entpacken können und dem Benutzer die Möglichkeit geben die komprimierten Dateien, wie die übrigen Formate, verwenden zu können.
- A-V-7: Das System soll für alle Inhalte die Möglichkeit bieten eine Vorschau auf den Inhalt zu generieren. Beispielsweise als Text mit hervorgehobenen Begriffen.
- A-V-8: Das System soll alle Inhalte als Word Cloud und Word Tree darstellen können.
- A-V-9: Das System soll zu Datensätzen mit Ortsangaben eine Karte anzeigen.
- A-V-10: Das System soll es ermöglichen Visualisierungen und die zugrunde liegenden Daten per URL aufzurufen. In dieser Ansicht soll es möglich sein, die Daten anzupassen und für eigene Visualisierungen zu nutzen.
- A-V-11: Das System soll es erlauben Diskussionen in sozialen Medien zu starten.

Anforderungen des Anwendungsfall „Metadaten ändern“:

- A-M-1: Der Benutzer soll Metadaten verbessern und hinzufügen können.
- A-M-2: Der Benutzer muss die Qualität der Datensätze bewerten können.
- A-M-3: Der Benutzer soll die Datensätze kommentieren können.
- A-M-4: Der Benutzer soll Datensätze mit Analysen oder ähnlichen Daten verlinken können.

4. Analyse

A-M-5: Der Benutzer soll Daten verbessern und hinzufügen können.

A-M-6: Das System muss die Benutzereingaben vom Ursprungsdatensatz aus sichtbar machen.

A-M-7: Das System soll die Benutzereingaben dem Betreiber zur Kontrolle vorlegen können.

Anforderungen der User Stories:

A-US-1: Der Benutzer soll mehrere Datensätze in einer Visualisierung zusammenführen können.

A-US-2: Der Benutzer soll mehrere Datensätze auf einer Karte gemeinsam darstellen können.

A-US-3: Das System soll Benutzer mit ähnlichen Interessen zusammenführen können.

A-US-4: Das System soll es Benutzern erlauben gemeinsam zu arbeiten.

A-US-5: Das System soll Benutzer über neue Datensätze benachrichtigen können. Möglich sind Benachrichtigungen für gewünschte Themengebiete und für neue Datensätze, die die Umgebung eines vorher definierten Ortes betreffen, z.B. des Wohnorts.

A-US-6: Das System soll die Benutzer dabei unterstützen Anfragen an Behörden nach neuen Daten zu stellen.

A-US-7: Der Benutzer soll „Geschichten“ erzeugen können, indem er verschiedene Dokumente, Visualisierungen, eigene Texte und Bilder verbinden kann.

A-US-8: Das System soll den Betreibern die Kontrolle der von den Benutzern erzeugten Daten ermöglichen.

A-US-9: Das System soll über eine Benutzerverwaltung verfügen.

4.3.2. Nichtfunktionale Anforderungen

Die aufgestellten nichtfunktionalen Anforderungen beschreiben nicht eine direkte Funktionalität des Systems, sondern welche Eigenschaften das System haben soll.

A-NF-1: Benutzerfreundlichkeit

Das System soll von der Zielgruppe intuitiv bedienbar sein und keine überfrachteten Oberflächen beinhalten. Ein Benutzer ohne Vorkenntnisse soll bei der ersten Nutzung innerhalb von 8 Minuten in der Lage sein die gewünschte Tätigkeit auszuführen.

4. Analyse

A-NF-2: Performance

Das System soll 75 Prozent der Anfragen in unter 500 ms beantworten, den Rest in unter 8 Sekunden.

A-NF-3: Erweiterbarkeit

Die Komponenten sollen lose gekoppelt seien. Funktionalitäten sollen leicht hinzugefügt und einfach ausgetauscht werden können.

A-NF-4: Play Framework

Das System soll mit dem Play Framework entwickelt werden, gemäß dem Titel dieser Bachelorarbeit.

4.4. Ausgrenzungen

Der interne Behördenablauf bei der Veröffentlichung von neuen offene Daten wird in dieser Arbeit nicht betrachtet.

Diese Arbeit basiert auf der Annahme einer verfügbaren CKAN API.

4.5. Zusammenfassung

In dem Kapitel wurde ein Szenario eines komfortableren Zugangs zu offenen Daten vorgestellt.

Die Hauptanwendungsfälle sind der Anwendungsfall „Informationen suchen“, der Anwendungsfall „Visuell suchen“, der Anwendungsfall „Visualisierung erstellen“, der Anwendungsfall „Metadaten ändern“ und das gemeinsame Arbeiten.

Abschließend wurde, um die Funktionen des Systems übersichtlicher darzustellen, aus den Anwendungsfällen die Anforderungen an das System herausgearbeitet.

5. Konzeption

Zur Umsetzung der Anforderungen ist ein Konzept hilfreich. Dieses Kapitel befasst sich mit der Architektur des verbesserten Open Data Portals.

Zuerst wird eine Übersicht über das System gegeben. Anschließend wird das System in Komponenten aufgeteilt und die resultierenden Komponenten werden vorgestellt.

5.1. System

Das System soll ein interaktives Online System werden. Die meisten Bürger haben Erfahrungen mit dieser Art von System und die bereits bestehenden Open Data Portale sind ähnlich umgesetzt. Zudem haben die meisten Bürger Zugang zu einem Web Browser.

Die Kontextabgrenzung 5.1 zeigt einen groben Überblick über das gesamte System.

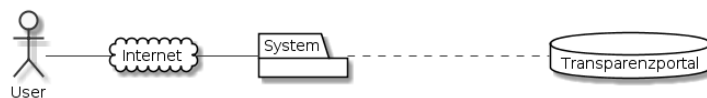


Abbildung 5.1.: Kontextabgrenzung

Der Benutzer greift über das Internet auf das System zu. Das System ist mit dem Hamburger Transparenzportal über die CKAN API verbunden.

5.1.1. Systemkomponenten

Aufbauend auf dem Abschnitt Open Data Portal Funktionen und der Beschreibung des Engage Portals wurden eine Forums Software und eine Wiki Lösung als in das Portal zu integrierende Software ausgemacht.

Forum und Wiki werden bei dem Engage Portal zur Unterstützung des gemeinsamen Arbeitens genutzt.

Ein Forum unterstützt die Zusammenführung von Benutzern mit ähnlichen Interessen, nach der Anforderung A-US-3. Zudem ermöglichen Foren eine Diskussion. Außerdem ermöglichen Foren Anfragen an Behörden und erfüllen somit die Anforderung A-US-6.

Ein Wiki kann für Tutorials genutzt werden. Außerdem können Analysen veröffentlicht und gemeinsam bearbeitet werden. Mit einem Wiki wird die Anforderung A-US-7, über das Verbinden verschiedener Inhalte zu „Geschichten“, ermöglicht.

Um das gemeinsame Arbeiten, nach A-US-4, weiter zu unterstützen wird zudem eine Etherpad Lösung integriert. Etherpads ermöglichen das gleichzeitige Arbeiten an einem gemeinsamen Dokument.

Forum, Wiki und Etherpad werden eingesetzt, da sie den meisten Benutzern bereits bekannt sind. Zudem haben sich die zugrunde liegenden Softwaresysteme in der Praxis bewährt und es gibt zahlreiche Einführungen in die Benutzung.

Um Daten zu speichern ist außerdem eine Datenbank erforderlich.

5.1.2. Systemkontext

In Abbildung 5.2 ist ein tieferer Einblick in das System dargestellt. Dort werden die einzelnen Softwarebestandteile des Systems gezeigt. Nachfolgend werden die einzelnen Bestandteile näher erläutert.

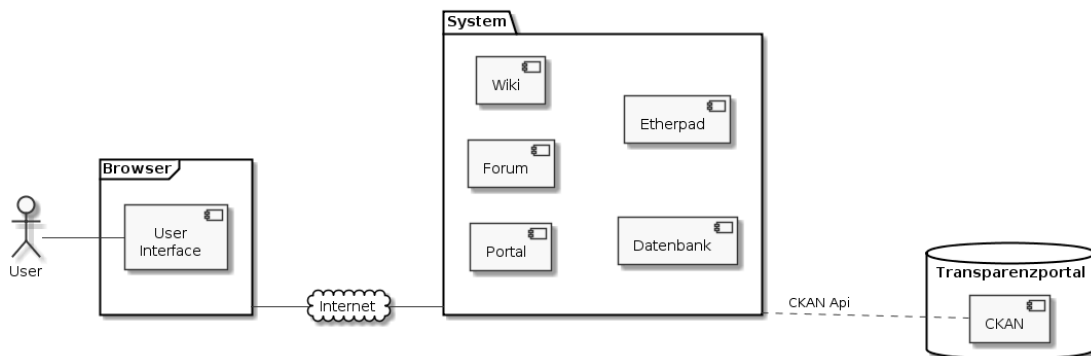


Abbildung 5.2.: Kontextdetails

User Der **Benutzer** greift über das **Internet** auf das **System** zu. Das **User Interface** des **Portals** ist in seinem **Browser** sichtbar. Dies ist seine hauptsächliche Interaktionsmöglichkeit mit dem **System**.

System Das **System** besteht aus einer Sammlung von Komponenten. Diese sind über eine gemeinsame Oberfläche ansprechbar.

Portal Das **Portal** bündelt alle Funktionen des **Systems** und verantwortet den Großteil der Anwendungsfälle.

Wiki Das **Wiki** stellt dem Benutzer die Möglichkeit bereit ein Tutorial zu erstellen. Zusätzlich können verschiedene Inhalte verbunden werden, um eine Geschichte zu erzählen. Außerdem können ausführliche Analysen bereitgestellt werden.

Forum Das **Forum** bietet die Möglichkeiten, Benutzer mit ähnlichen Interessen zu finden, in Gruppen Diskussionen zu führen und Anfragen nach neuen Daten zu stellen.

Etherpad Das **Etherpad** dient als zusätzliche Hilfe, um in Gruppen gemeinsam zu arbeiten.

Datenbank Die **Datenbank** speichert die vom Benutzer geänderten Metadaten und Daten.

CKAN Die **CKAN** Instanz des Hamburger **Transparenzportals** liefert über die **CKAN API** die offenen Daten.

5.2. Architektur

Die Architektur des Portals ist in mehrere Komponenten aufgeteilt. Die Komponenten dienen der Strukturierung des Konzeptes. Zugleich ermöglichen sie über ihre lose Kopplung die Erweiterbarkeit, nach A-NF-4. In diesem Abschnitt wird die Entstehung und die Funktionalität der einzelnen Softwarekomponenten und deren Zusammenspiel beschrieben.

5.2.1. Komponentenschnitt

Die entwickelten Anwendungsfälle dienten als erste Orientierung für die Unterteilung der Komponenten. So ergaben sich die fachlichen Komponenten Suche, UserMetadaten und Visualisierungen aus dem Anwendungsfall „Informationen suchen“, dem Anwendungsfall „Metadaten ändern“ und dem Anwendungsfall „Visualisierung erstellen“. Der Anwendungsfall „Visuell suchen“ lässt sich mit den Komponenten Suche und Visualisierungen abbilden. Er bildet daher keine eigene Komponente.

Aus den User Stories ließen sich die weiteren Komponenten Zusammenarbeit (User Stories 2-5), Kontrolle (User Story 6) und Benutzerverwaltung (User Story 7) herauslesen.

Anschließend wurden die aufgestellten Anforderungen als umzusetzende Operationen betrachtet und jeweils einer, fachlich am nächsten gelegenen, Komponente zugeordnet.

Abschließend wurden die einzelnen Operationen im Hinblick auf mögliche Hilfsfunktionen überprüft. Daraus ergab sich der Bedarf einer Persistenzkomponente. Zudem erschien es sinnvoll, die Schnittstelle zum CKAN Portal über einen CKAN Adapter zu abstrahieren. Zusätzlich ergab sich die Notwendigkeit, die offenen Daten für die Visualisierungen anpassen zu können. Diese Funktionalität wird in der Dateien Komponente gebündelt.

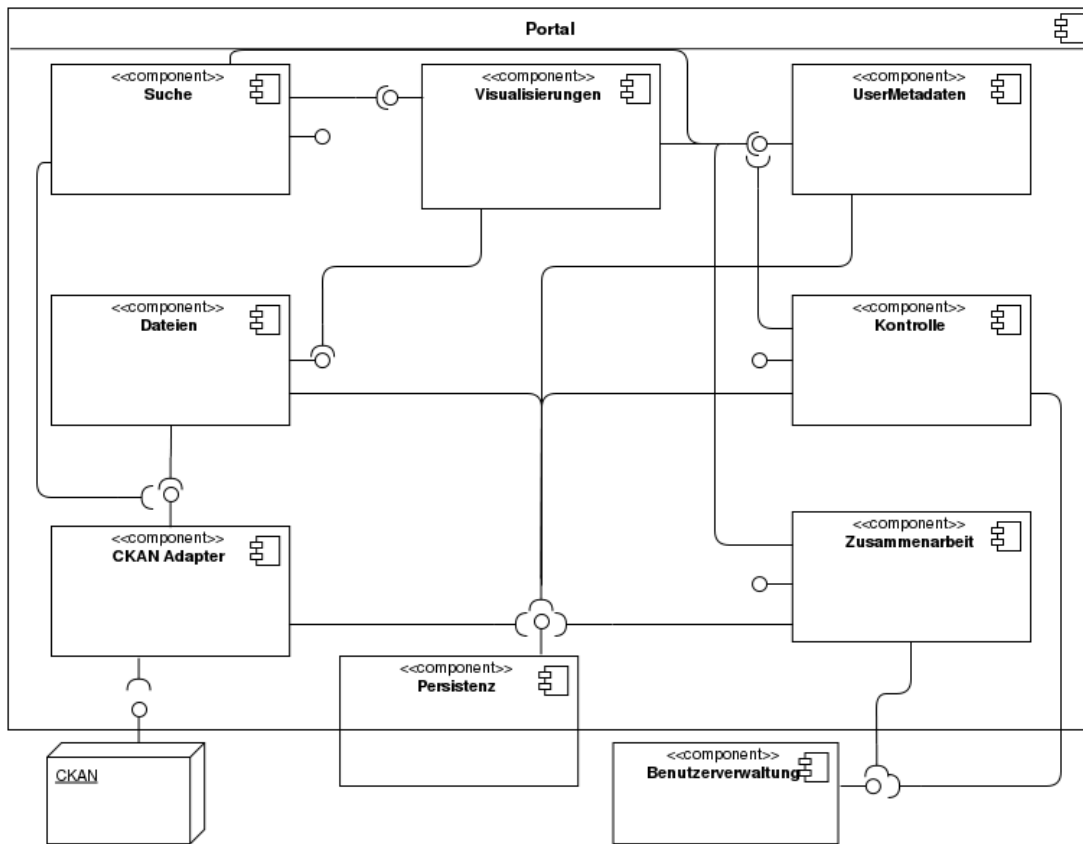


Abbildung 5.3.: Komponenten der Systemkomponente Portal

Das Komponentendiagramm 5.3 veranschaulicht die modulare Aufteilung des Systems.

5.2.2. Komponenten

Im Folgenden werden die Funktionen der Komponenten näher erläutert. Dabei wird jeweils dargestellt, welche Anforderung mit der jeweiligen Funktion erfüllt wird. Abschließend werden jeweils wichtige Methoden vorgestellt. Viele Operationen haben als Resultat Oberflächen ausgewiesen. Diese stellen die Schnittstelle zum Benutzer dar.

5.2.3. Fachliche Komponenten

Suche

Die Komponente Suche verantwortet den Anwendungsfall „Informationen suchen“ und realisiert gemeinsam mit der Komponente Visualisierungen den Anwendungsfall „Visuell suchen“.

Sie beinhaltet die Oberfläche der Suche, sowie die Suchergebnisanzeige (A-IS-1, A-IS-3). Die Suchergebnisse fragt sie vom CKAN Adapter ab. Dieser nutzt die CKAN Schnittstelle. Für die von Benutzern erstellten Inhalte wird die UserMetadaten Komponente befragt.

Um mit geographischen Metadaten versehene Inhalte darzustellen, nutzt die Komponente eine Kartendarstellung der Visualisierungskomponente (A-V-9).

Von der Ergebnisanzeige bietet sie für die vorhandenen Ressourcen die Visualisierungsoperationen der Visualisierungskomponente an (A-IS-4). Ebenso verweist sie auf die, von anderen Benutzern hoch geladenen, Daten der UserMetadaten Komponente (A-M-6).

Sie bietet Operationen zum Filtern der Suche (A-IS-2).

Für die Darstellung der Visuellen Suche nutzt die Komponente Operationen der Visualisierungskomponente. Insbesondere die Darstellung als Word Cloud und Word Tree (A-VS-1, A-VS-2, A-VS-3).

Suchergebnisliste *suche* (Begriff, Option<Filter>) sucht nach dem Suchbegriff mit einem optionalem Filter und stellt die Ergebnisse als Liste dar.

VisuelleSuchergebnisliste *sucheVisuell* (Begriff, Option<Filter>) sucht nach dem Begriff mit einem optionalem Filter und stellt die Ergebnisse visuell dar.

Detailseite *zeigeDetails* (ID) zeigt für das ausgewählte Ergebnis die vorhandenen Details an.

(A-IS-1, A-IS-2, A-IS-3, A-IS-4, A-M-6, A-VS-1, A-VS-2, A-VS-3 und A-V-9)

UserMetadaten

Die Komponente UserMetadaten ist zuständig für die von den Benutzern erstellten Inhalte. Im wesentlichen die Operationen des Anwendungsfall „Metadaten ändern“.

Sie enthält die Ansichten zur Eingabe von geänderten Metadaten, Bewertungen, Kommentaren, Analysen und weiterführenden Links (A-M-1, A-M-2, A-M-3, A-M-4). Von Benutzern verbesserte oder neue Daten nimmt sie entgegen (A-M-5).

Zudem ist die Komponente zuständig, die Userdaten in der Persistenz zu verwalten. Sie speichert die Userdaten mit der dazugehörigen Datensatz ID der CKAN Instanz. So können die Userdaten dem Datensatz der CKAN Instanz zugeordnet werden.

MetadatenAnzeige zeigeMetadaten (ID) zeigt die Metadaten zu dem Datensatz mit der gegebenen ID und verweist auf von Benutzern hinzugefügte Inhalte.

UserdatenAnzeige zeigeUserdaten (ID, User) zeigt die Userdaten des Users zu dem Datensatz mit der gegebenen ID.

UserdatenAnzeige erstelleNeueUserdaten (ID, User, neueUserdaten) füge die gegebenen neuen Userdaten zu dem Datensatz mit der gegebenen ID hinzu.

UserdatenAnzeige aendereUserdaten (ID, User, neueUserdaten) ändere die Userdaten des Datensatzes mit der gegebenen ID.

UserdatenAnzeige dateiHochladen (Datei, Option<ID>) nimmt neue oder verbesserte Dateien vom Benutzer entgegen und leitet auf die UserdatenAnzeige zur Eingabe von Metadaten.

Liste<Userdaten> sucheNachUserdaten (Begriff, Option<Filter>) sucht nach dem Begriff in den von den Benutzern erstellten Inhalten mit einem optionalem Filter und gibt die Ergebnisse als Liste zurück.

Liste<Userdaten> gibUserdaten (ID) gibt die zu der ID zugehörigen Userdaten als Liste zurück.

(A-M-1, A-M-2, A-M-3, A-M-4 und A-M-5)

Visualisierungen

Die Komponente Visualisierungen bietet verschiedene Tools zur grafischen Anzeige der Inhalte bzw. der Ressourcen eines Datensatzes an (A-V-1). Sie verantwortet den Anwendungsfall „Visualisierung erstellen“. Für die Komponente Suche stellt sie die Visualisierungen des Anwendungsfall „Visuell suchen“ zur Verfügung.

Die darzustellenden Daten lässt die Komponente von der Dateien Komponente aufbereiten.

Um die Benutzer in der Erstellungsphase [siehe 3.3] zu unterstützen bietet die Komponente eine Tabellenansicht für CSV und Excel Dateien (A-V-2). In der Ansicht können die Daten sortiert und gefiltert werden. Zusätzlich können Daten korrigiert und hinzugefügt werden (A-V-3). Darauf aufbauend bietet sie die Möglichkeit mehrere Datensätze zu kombinieren (A-US-1).

Aus den Tabellendaten können Diagramme generiert werden (A-V-4). Ziel ist es die konfirmative Analyse zu unterstützen, also Visualisierungsmöglichkeiten zu finden, welche eine Hypothese überprüfbar machen.

Die Diagramme können über einen speziellen Link verbreitet werden. Über diesen Link können die Ursprungsdaten des Diagramms aufgerufen werden. Und mit diesen Ursprungsdaten können neue, ähnliche Diagramme erzeugt werden (A-V-10). Dies soll die Erkundung und Wiederverwendung von Visualisierungen vereinfachen, siehe auch Abschnitt 3.3.

Andere Formate lässt die Komponente von der Dateienkomponente extrahieren und in ein verwendbares Format bringen (A-VS-4). Für die Darstellung bietet sie die in 2.2 beschriebenen Word Cloud und Word Tree Darstellungen (A-V-8).

Zusätzlich bietet die Komponente eine Kartenansicht. Auf dieser sind mehrere Dokumente auswählbar (A-US-2).

WordCloudAnzeige *zeigeWordCloud* (RessourcenID) zeigt eine WordCloud des Inhaltes der Ressource mit der gegebenen ID.

WordTreeAnzeige *zeigeWordTree* (RessourcenID, Wort) zeigt einen WordTree des Inhaltes der Ressource mit der gegebenen ID. Der Tree ist vor selektiert auf den gegebenen Begriff.

TabellenAnsicht *tabelle* (RessourcenID) zeigt ein Tool, in dem die Daten der gegebenen Ressource in einer Tabelle dargestellt sind. Die Daten können sortiert, gefiltert und geändert werden. Und es können verschiedene Diagramme erzeugt werden.

TabellenAnsicht *tabelle* (Liste<RessourcenID>) zeigt das Tabellentool mit den kombinierten Daten der angegebenen Ressourcen.

ShareID *teileDiagrammUndDaten* () speichert die vom Benutzer geänderten Einstellungen und Daten.

TabellenAnsicht *zeigeDiagrammMitUrsprungsdaten* (ShareID) zeigt das mit der ShareID hinterlegte Diagramm und die dazugehörigen Daten und Einstellungen in der Tabellenansicht an.

KartenAnsicht *karte* (List<ID>) zeigt eine Kartenansicht mit den Orten der gegebenen Datensätze hervorgehoben.

(A-V-1, A-V-2, A-V-3, A-V-4, A-V-8, A-V-10, A-VS-4, A-US-1 und A-US-2)

Dateien

Die Komponente Dateien kann verschiedene Dateiformate konvertieren und Informationen aus den Dateien extrahieren.

Für die PDF Bearbeitung bietet die Komponente Hervorhebungsmöglichkeiten und einen PDF zu Bild Konverter (A-V-5). Excel Dateien können in CSV Dateien umgewandelt werden und zip Dateien können entpackt werden (A-V-6).

Außerdem kann die Komponente aus Dateien lesbaren Text extrahieren und Inhalte darin hervorheben (A-V-7). Diese Funktion soll helfen zu evaluieren, ob die in Abschnitt 2.2.3 beschriebenen Vorschäbilder Benutzern helfen können, die Inhalte schneller einzuordnen und zu erfassen .

Zudem bietet die Komponente einen Download Assistenten. Dieser dient dazu die Ressourcen aus dem Internet herunterzuladen. Dabei nutzt er einen internen Cache.

FileID *holeDatei* (RessourcenID) überprüft ob die angegebene Ressource im Cache vorhanden ist. Falls nicht wird sie heruntergeladen und in den Cache eingetragen.

BildAnzeige *zeigeBild* (RessourcenID) zeigt die Ressource mit der gegebenen ID als Bild an.

BildAnzeige *zeigeBildMitHervorhebung* (RessourcenID, Begriff) zeigt die Ressource mit dem Begriff hervorgehoben als Bild an.

PDFAnzeige *zeigePDFMitHervorhebung* (RessourcenID, Begriff) zeigt das Dokument mit dem Begriff hervorgehoben als PDF an.

Inhaltstext *extrahiereText* (RessourcenID) extrahiert den Inhalt der Ressource als Text.

TextAnzeige *zeigeTextmitHervorhebung* (RessourcenID, Begriff) zeigt den Inhalt als Text mit dem Begriff hervorgehoben.

ZipAnzeige *entpacke* (RessourcenID) entpackt die angegebene Ressource.

FileID *konvertiereExcelNachCSV* (RessourcenID) konvertiert die angegebene Ressource in eine CSV Datei.

HTTPFile *liefereDatei* (FileID) liefert die angegebene Datei aus. Die Methode wird benötigt, zur Anforderung einer Datei von JavaScript Programmcode aus.

(A-V-5, A-V-6 und A-V-7)

Kontrolle

Die Komponente Kontrolle nutzt die Schnittstellen der Benutzerverwaltung, der UserMetadaten Komponente und der Persistenz Komponente. Ziel ist es, die Benutzerdaten dem Portalbetreiber zur Kontrolle anzuzeigen (A-M-7, A-US-8).

UserDaten zeigeUserDaten () zeigt alle Userdaten nach Datum sortiert an.

UserDaten loescheUserDaten (UserDatenID) löscht die angegebenen Userdaten.

UserDaten markiereUserDatenGeprueft (UserDatenID) markiert die Userdaten als geprüft.

(A-M-7 und A-US-8)

Zusammenarbeit

Die Komponente Zusammenarbeit bündelt die im Abschnitt Systemkomponenten erwähnten Softwarelösungen Forum, Wiki und Etherpad. Im Forum kann diskutiert werden, Gruppen mit ähnlichen Interessen können sich finden (A-US-3), Anfragen nach Daten können gestellt werden (A-US-6) und Feedback kann gegeben werden. Im Wiki können Tutorials, „Geschichten“ (A-US-7) oder Analysen präsentiert werden. Im Etherpad können die Benutzer gemeinsam an Projekten arbeiten (A-US-4). Dazu gibt es die Möglichkeit bei der Suchergebnisanzeige die gefundene Ressource in ein Etherpad oder Wiki zu importieren.

Zudem ist die Komponente dafür zuständig, dass die Benutzer Diskussionen in sozialen Netzwerken starten können. Dazu nutzt sie ein datenschutzfreundliches externes Tool nach dem Zwei-Klick-Verfahren (A-V-11).

Außerdem beinhaltet die Komponente einen Batch Job. Dieser überprüft einmal am Tag, ob neue Datensätze veröffentlicht wurden. Falls dies der Fall ist und Benutzer ihr Interesse an dieser Datenart bekundet haben, werden die entsprechenden Benutzer informiert (A-US-5).

Für Forum, Wiki, Batch Job und Etherpad muss eine Datenbank zur Verfügung gestellt werden.

Forum, Wiki, Etherpad und das Portal nutzen eine Benutzerverwaltung (A-US-9).

Um nicht die komplexen Sicherheitsanforderungen an eine Benutzerverwaltung selbst aufzustellen und bei einer Realisierung Sicherheitslücken einzubauen, wird auf eine externe Benutzerverwaltung zurückgegriffen. Diese wird hier nicht näher betrachtet, da dies ein Standardproblem der Informatik mit etablierten Lösungsansätzen ist.

(A-US-3, A-US-4, A-US-5, A-US-6, A-US-7, A-US-9, und A-V-11)

5.2.4. Technische Komponenten

CKAN Adapter

Der CKAN Adapter abstrahiert die Schnittstelle zu einer CKAN Instanz. Im Wesentlichen benutzt er die CKAN API zur Suche und konvertiert die erhaltene JSON Antwort in Scala Klassen.

Liste<MetaDaten> *suche* (Begriff, Option<Filter>) sucht nach dem Suchbegriff mit einem optionalem Filter und gibt eine Liste von Metadaten zurück.

Metadaten *zeigeDetails* (ID) liefert für das ausgewählte Ergebnis Details.

URL *gibURLZuRessource* (RessourcenID) gibt die Adresse der Ressource zurück.

Persistenz

Die Persistenz Komponente stellt für andere Komponenten Methoden bereit um Daten abzuspeichern und abzufragen.

dbID *erstelle* (Objekt) speichert ein gegebenes Objekt.

***loesche* (dbID)** lösche ein gegebenes Objekt.

Objekt *finde* (dbID) sucht ein Objekt nach der dbID.

List<Objekt> *sucheAlle* (Suchoptionen) sucht alle Objekte auf die die Suchoptionen zutreffen.

5.2.5. Nichtfunktionalen Anforderungen

Die erste nichtfunktionale Anforderung nach Benutzerfreundlichkeit wird durch eine einfache Benutzeroberfläche sichergestellt. Diese konzentriert sich auf das Wesentliche und ist somit intuitiv bedienbar.

Die zweite nichtfunktionale Anforderung nach Performance wird durch das Vermeiden rechenintensiver Operationen auf dem Server erfüllt.

Die Aufteilung in Komponenten erfüllt die dritte nichtfunktionale Anforderung nach Erweiterbarkeit.

Die vierte nichtfunktionale Anforderung ist der Einsatz des Play Frameworks. Dies beeinflusst die vorgestellte Architektur. Das Framework empfiehlt eine Einteilung der Klassen nach

dem Model-View-Controller Schema [siehe 2.3.3]. Dies wird empfohlen, um die Ablaufsteuerung, mit den Controllern, von der Datenhaltung, den Models, und der Darstellung, den Views, zu entkoppeln. So gibt es je ein controllers, ein models und ein views Paket. Dies hat zur Folge, dass die vorgestellten Komponenten nicht im Programmcode als Paket zusammengefasst werden können. Stattdessen erhält jede Komponente in den MVC Paketen jeweils eigene tiefer liegende Pakete.

Durch den Einsatz des Play Frameworks ist auch bedingt, dass die Komponentenschnittstellen auf dem REST Stil [siehe 2.3.2] basieren.

Die zweite nichtfunktionale Anforderung wird durch die einfache Skalierbarkeit des Play Frameworks weiter unterstützt.

Das Play Framework unterstützt auch die dritte nichtfunktionale Anforderung. Indem neue Routen definiert werden können, welche auf neue Komponenten verweisen können.

5.3. Zusammenfassung

In diesem Kapitel wurde erst ein Überblick über das System gegeben. Anschließend wurden Komponenten gebildet. Diese setzten die Anforderungen um. Darauf folgte eine Vorstellung der Komponenten Suche, UserMetadaten, Visualisierungen, Dateien, Kontrolle, CKAN Adapter, Persistenz und Zusammenarbeit.

6. Umsetzung

In diesem Kapitel wird die durchgeführte prototypische Realisierung des Systems beschrieben. Zu Beginn wird der Realisierungsumfang aufgezeigt. Im Anschluss werden die bei der Realisierung eingesetzten Technologien vorgestellt.

Darauf folgt eine Evaluation des Prototypen. Dabei werden mögliche Probleme und deren mögliche Lösung gezeigt. Abschließend wird die Arbeit im Hinblick auf ihre Zielsetzung bewertet.

6.1. Realisierungsumfang

Die prototypische Realisierung setzt das vorgestellte Konzept teilweise um. Aus Zeitgründen wurde nicht das gesamte Konzept umgesetzt.

Zentraler Bestandteil des Systems ist die Darstellung der in dem Hamburger Transparenzportal vorhandenen Ressourcen. Dazu wurde die Komponente Suche realisiert und die Schnittstelle zu der CKAN Instanz. Die Suchfilterung wurde für eine Darstellung der Konzepte als nicht elementar eingestuft und nicht umgesetzt.

Um die gefundenen Ressourcen für eine Darstellung aufbereiten zu können, wurde die Dateien Komponente umgesetzt.

Ein Hauptbestandteil des Prototypen ist es, aufzuzeigen wie die Inhalte des Transparenzportals besser dargestellt werden können. Zu diesem Zweck wurde die Komponente Visualisierungen mit einigen Darstellungen umgesetzt.

Verzichtet wurde auf das Teilen von Visualisierungen mit der möglichen Nachbearbeitung der verwendeten Daten und Einstellungen. Dieses Konzept ist ohne eine prototypische Realisierung ebenso verständlich und die Umsetzung hätte umfangreiche Arbeiten an einer der eingesetzten Bibliotheken bedurft.

Auf die Umsetzung einer Kartenansicht wurde verzichtet, da nur sehr wenige Inhalte des Transparenzportals mit Geodaten versehen sind.

Die UserMetadaten Komponente wurde im Umfang reduziert umgesetzt. Eine erfolgreiche Umsetzung dieses Konzeptes bräuchte eine kritische Masse von Benutzern, welche das Feature

nutzen müssten. Nur mit genug Inhalten ließe sich zeigen, wie die Benutzer sich gegenseitig helfen können. Diese kritische Masse wird mit diesem Prototyp nicht erreicht werden. Daher wird auf eine ausführliche Realisierung verzichtet.

Da die Komponente Persistenz auf Grund der vorgenommenen Einschränkung kein zentraler Bestandteil mehr ist, wurde hier auf eine umfangreiche Realisierung verzichtet. Zudem ist das Persistenzproblem ein oft gelöstes Problem der Informatik mit einer Vielzahl an vorhandenen Lösungen.

Bei der Komponente Kontrolle wird aus dem gleichen Grund auf eine Realisierung verzichtet und auf die integrierten Zugriffsmöglichkeiten der eingesetzten Datenbank verwiesen.

Die Komponente Zusammenarbeit wurde nicht umgesetzt, da sie im Wesentlichen aus mehreren Softwarestandardlösungen besteht. Auch hier lässt sich der volle Umfang des Konzeptes nicht in diesem Prototyp ohne involvierte Benutzer zeigen.

Bei der optischen Gestaltung des Prototypen wurde Wert auf eine schlichte Gestaltung gelegt, um die Konzepte ohne ablenkende grafische Spielereien zeigen zu können.

6.2. Realisierung

In diesem Abschnitt wird die Umsetzung der einzelnen Komponenten beschrieben. Insbesondere werden Lösungsalternativen dargestellt und die eingesetzten Bibliotheken vorgestellt. Die jeweils folgenden Tabellen verdeutlichen welche Anforderung in dem Konzept in welcher Komponente realisiert werden soll und ob diese Funktionalität im Prototypen umgesetzt wurde (+: Ja, -: Nein, T: teilweise).

6.2.1. Suche

Die Suche Komponente stellt als zentralen Einstieg ein Suchfeld zur Verfügung. Zur Suche wird die definierte CKAN Instanz befragt. Diese antwortet im JSON Format. Hervorhebenswert ist hier der Einsatz der vom Play Framework zur Verfügung gestellten Reader für JSON. Diese prüfen die Antwort auf Konsistenz und helfen dabei die JSON Werte in Scala Objekte und Werte umzusetzen. Dazu wurde das Metadatenschema im `models.MetadatenCKAN` Paket definiert. Das allgemeine Schema wurde im CKAN Kapitel 2.3.4 vorgestellt. Zusätzlich wurden die spezifischen Erweiterungen des Hamburger Transparenzportals berücksichtigt¹.

¹<http://transparenz.hamburg.de/contentblob/4354384/data/mdm-schema1-6.pdf> - Zugriffsdatum: 08.06.2015

6. Umsetzung

Anforderung	K	U
A-IS-1 (Suchen und Finden von Datensätzen)	+	+
A-IS-2 (Suche filtern)	+	-
A-IS-3 (Suchtrefferzusammenfassung und Ressourcen)	+	+
A-IS-4 (Detailansicht)	+	T
A-M-6 (Benutzereingaben sichtbar)	+	T
A-VS-1 (Inhalte als Word Cloud)	+	+
A-VS-2 (Word Tree von Word Cloud)	+	+
A-VS-3 (Inhalte als Word Tree)	+	+
A-V-9 (Kartenansicht)	+	-

Tabelle 6.1.: Umsetzung der Anforderungen in der Komponente Suche

6.2.2. Dateien

Die Komponente Dateien ist mit mehreren externen Bibliotheken umgesetzt. Dabei wurden mehrere Open Source Lösungen evaluiert.

Extraktion

Apache Tika² ist ein Projekt der Apache Software Foundation. Die Apache Software Foundation unterstützt Entwickler bei der Erstellung und Pflege von Open Source Software.

Apache Tika wurde gewählt, da es aus über 1000 verschiedenen Dateiformaten Metadaten und den Inhalt extrahieren kann. Tika ist verantwortlich den Inhalt aus den verschiedenen Ressourcentypen des Transparenzportal zu extrahieren.

zip Format

Zum Entpacken der zip Dateien wird das java.util.zip Paket genutzt. Dabei werden auch Hilfsmethoden der Commons IO genutzt.

Apache Commons³ ist ein Projekt der Apache Software Foundation. Fokus des Projektes ist es wiederverwendbare Java Komponenten zu Pflegen. Die Commons IO⁴ ist ein Teil des Projektes. Ziel der Bibliothek ist es Funktionalitäten bereitzustellen die der Unterstützung von Input Output Funktionen dienen.

²<https://tika.apache.org/> - Zugriffsdatum: 24.09.2015

³<https://commons.apache.org/> - Zugriffsdatum: 24.09.2015

⁴<https://commons.apache.org/proper/commons-io/> - Zugriffsdatum: 24.09.2015

PDF Format

Für die Manipulation von PDF Dateien wurden mehrere Bibliotheken getestet.

Zuerst wurde PDF Clown⁵ evaluiert. Diese Open Source Bibliothek soll ein Universalwerkzeug zur Manipulation von PDF Dokumenten sein. Bei der Markierung von Inhalten gab es jedoch Probleme mit Umlauten und die Textstruktur wurde nicht immer zuverlässig erkannt.

Daher wird eine andere Bibliothek eingesetzt. Die Apache PDFBox⁶ Bibliothek ist ein Projekt der Apache Software Foundation. PDFBox erlaubt das Arbeiten mit PDF Dateien. In diesem System wird die Bibliothek zur Hervorhebung von Begriffen in PDF Dateien verwendet.

Bei der Umwandlung von PDF Dokumenten in besser anzeigbare Bilder produziert PDFBox jedoch nur weiße Seiten. Daher wird für die PDF zu Bild Umwandlung eine weitere Bibliothek genutzt.

Ghostscript ist ein Open Source Softwarepaket zur Darstellung von PDF und Postscript Dateien auf Druckern und Bildschirmen. Mit Ghost4J⁷ kann die Ghostscript API von Java Code aus genutzt werden. Nachteilig ist, dass auf dem Server des Portals Ghostscript installiert sein muss. Ghostscript ist jedoch sowohl für Linux als auch Windows frei verfügbar. Anzumerken ist außerdem, dass für den produktiven Einsatz die Multi Threading Fähigkeiten genauer evaluiert werden müssen.

Excel Format

Zur Konvertierung von Excel Dateien in besser verwendbare CSV Dateien wurde wieder ein Projekt der Apache Software Foundation eingesetzt. Das Apache POI⁸ Projekt hat es sich zur Aufgabe gemacht Microsoft Office Formate über eine Java API manipulierbar zu machen.

Als Cache des Ressourcen Download Assistenten wurde das vorhandene interne Dateisystem des Servers gewählt.

Anforderung	K	U
A-V-5 (PDF Dateien als Bild, PDF und auf dem Portal)	+	+
A-V-6 (zip Dateien entpackbar)	+	+
A-V-7 (Vorschau für alle Inhalte)	+	+

Tabelle 6.2.: Umsetzung der Anforderungen in der Komponente Dateien

⁵<http://pdfclown.org/> - Zugriffsdatum: 24.09.2015

⁶<https://pdfbox.apache.org/> - Zugriffsdatum: 24.09.2015

⁷<http://www.ghost4j.org/index.html> - Zugriffsdatum: 24.09.2015

⁸<http://poi.apache.org/> - Zugriffsdatum: 24.09.2015

6.2.3. UserMetadaten

Für die Umsetzung der UserMetadaten Komponente wird eine Datenbank benötigt. Um flexibel die verschiedenen User Inhalte speichern zu können und da die Möglichkeit besteht Dateien in der Datenbank zu speichern, wurde die dokumentenorientierte MongoDB⁹ Datenbank gewählt. Es bestanden außerdem bereits Erfahrungen mit dem Einsatz dieser Datenbank.

Als Treiber für den Datenbank Zugriff wurde der Reactive Mongo¹⁰ Treiber gewählt. Dieser ist für den Einsatz mit dem Play Framework gedacht und die Dokumentation war am aussagekräftigsten von den evaluierten Alternativen.

Aus Kompatibilitätsgründen war es dabei nötig die Play Version von der aktuellsten Version auf den 2.2er Zweig downzugraden.

Zum Laufen benötigt das System eine lokal laufende MongoDB Instanz auf Port 27017 mit einer apertus Datenbank. Im produktiven Einsatz müsste die Datenbank noch abgesichert werden.

Anforderung	K	U
A-M-1 (Metadaten verbessern)	+	T
A-M-2 (Datenqualität bewerten)	+	+
A-M-3 (Datensätze kommentieren)	+	+
A-M-4 (Analysen und ähnliche Daten verlinken)	+	+
A-M-5 (Daten verbessern)	+	T

Tabelle 6.3.: Umsetzung der Anforderungen in der Komponente UserMetadaten

6.2.4. Visualisierungen

Für den Prototypen wurde Wert auf eine einfache Einbindung der Visualisierungen gelegt.

Für den Word Tree und die Word Cloud wird auf das Google Charts Projekt¹¹ zurückgegriffen. Die Visualisierungen sind einfach und kostenfrei zu nutzen. Sie bestehen aus einzubindendem JavaScript Code. Andere evaluierte Bibliotheken hatten meist nur eine Word Cloud Darstellung und keine Word Tree Darstellung. Alternativ hätten die Darstellungen manuell programmiert werden müssen.

⁹<https://www.mongodb.org/> - Zugriffsdatum: 24.09.2015

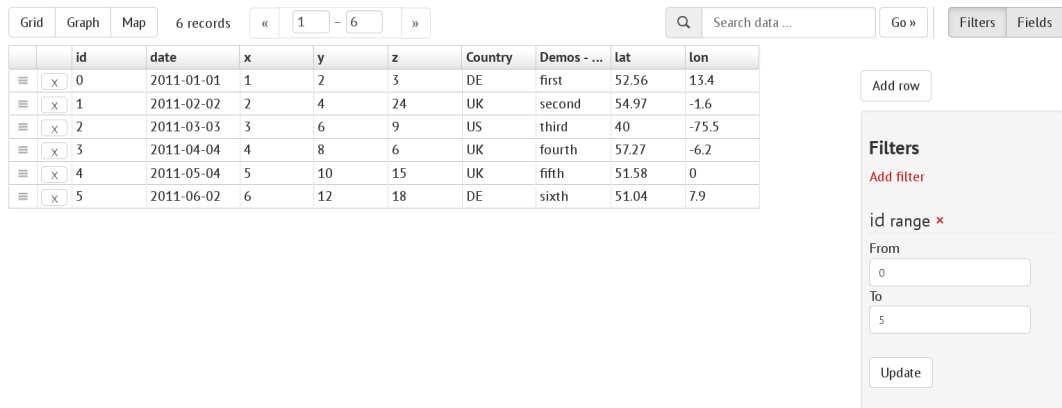
¹⁰<http://reactivemongo.org/> - Zugriffsdatum: 24.09.2015

¹¹<https://developers.google.com/chart/> - Zugriffsdatum: 24.09.2015

6. Umsetzung

Für die Word Cloud Darstellung wurde eine Erweiterung des VisApi Teams gewählt. Die TermCloud¹² Darstellung hat den Vorteil, dass dort den Schlagworten ein Link zugewiesen werden kann. Dieser Link verweist auf die Word Tree Darstellung.

Für die Tabellenansicht wird das Recline.js¹³ Projekt eingesetzt. Recline ist ein Open Source Projekt der Open Knowledge Foundation. Es bietet eine Tabellenansicht zur Manipulation, Sortierung und Filterung. Zudem können Diagramme erstellt werden.



	id	date	x	y	z	Country	Demos - ...	lat	lon	
≡	x	0	2011-01-01	1	2	3	DE	first	52.56	13.4
≡	x	1	2011-02-02	2	4	24	UK	second	54.97	-1.6
≡	x	2	2011-03-03	3	6	9	US	third	40	-75.5
≡	x	3	2011-04-04	4	8	6	UK	fourth	57.27	-6.2
≡	x	4	2011-05-04	5	10	15	UK	fifth	51.58	0
≡	x	5	2011-06-02	6	12	18	DE	sixth	51.04	7.9

Abbildung 6.1.: Recline.js Tabellenansicht [Lab15]

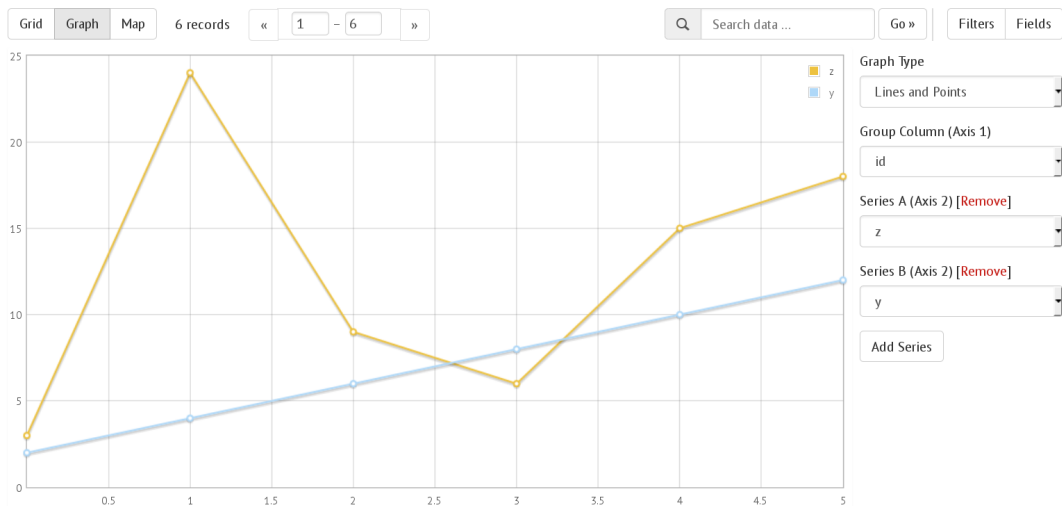


Abbildung 6.2.: Recline.js Diagramm [Lab15]

¹²<http://visapi-gadgets.googlecode.com/svn/trunk/termcloud/doc.html> - Zugriffsdatum: 24.09.2015

¹³<http://okfnlabs.org/recline/> - Zugriffsdatum: 24.09.2015

6. Umsetzung

Recline erfüllt die Anforderungen nach Angemessenheit und Effektivität [siehe 2.1.2]. Visualisierungen sind einfach zu erstellen und durch die verschiedenen erstellbaren Diagramme kann das Wissen leicht aufnehmbar dargestellt werden. Die Expressivität hingegen ist immer zu einem großen Teil von den Entscheidungen des Benutzers abhängig.

Anforderung	K	U
A-V-1 (erstellbare Visualisierungen)	+	+
A-V-2 (Tabellenansicht für CSV und Excel Dateien)	+	+
A-V-3 (Visualisierungsdaten anpassbar)	+	+
A-V-4 (Verschiedene Diagramme mit einstellbaren Parametern)	+	+
A-V-8 (alle Inhalte als Word Cloud und Word Tree darstellbar)	+	+
A-V-10 (Visualisierungen und zugrunde liegenden Daten per URL aufrufbar)	+	-
A-VS-4 (Visualisierungen automatisch erzeugbar)	+	T
A-US-1 (Datensätze kombinierbar)	+	T
A-US-2 (Datensätze auf Karte darstellbar)	+	-

Tabelle 6.4.: Umsetzung der Anforderungen in der Komponente Visualisierungen

6.2.5. Konzeptionelle Anforderungen

Die Komponenten Zusammenarbeit und Kontrolle wurden aus Zeitgründen nicht umgesetzt.

Die nichtfunktionalen Anforderungen wurden im Konzept bedacht und in der Umsetzung berücksichtigt.

Anforderung	K	U
A-US-3 (Benutzer mit ähnlichen Interessen verbinden)	+	-
A-US-4 (Gemeinsam arbeiten)	+	-
A-US-5 (Benachrichtigungen über neue Datensätze)	+	-
A-US-6 (Anfragen unterstützen)	+	-
A-US-7 („Geschichten“ erzählen)	+	-
A-US-9 (Benutzerverwaltung)	+	-
A-V-11 (Diskussionen in sozialen Medien starten)	+	-

Tabelle 6.5.: Umsetzung der Anforderungen in der Komponente Zusammenarbeit

6. Umsetzung

Anforderung	K	U
A-M-7 (Benutzereingaben kontrollierbar)	+	T
A-US-8 (Kontrolle von Betreiber)	+	-

Tabelle 6.6.: Umsetzung der Anforderungen in der Komponente Kontrolle

Anforderung	K	U
A-NF-1 (Benutzerfreundlichkeit)	+	T
A-NF-2 (Performance)	+	T
A-NF-3 (Erweiterbarkeit)	+	T
A-NF-4 (Play Framework)	+	+

Tabelle 6.7.: Umsetzung der nichtfunktionalen Anforderungen

6.3. Evaluation

Im Folgenden werden mögliche Probleme der Umsetzung und des Konzeptes aufgezeigt. Anschließend werden mögliche Lösungen vorgestellt.

Zum Abschluss wird evaluiert, inwieweit die Umsetzung das Ziel dieser Arbeit erfüllt.

6.3.1. Probleme

Bei der Umsetzung gab es kleinere Probleme. Kenntnisse dieser Probleme können weiterführende Arbeiten erleichtern. Die Probleme lassen sich in drei Kategorien einteilen:

Probleme bei offenen Daten

Ein nicht in diesem Umfang erwartetes Problemfeld sind die vorhandenen Metadaten und die uneinheitlichen Datenformate.

So hat fast jede der veröffentlichenden Stellen ihr eigenes Format für CSV Dateien. Eine weitere externe Bibliothek könnte hier Abhilfe schaffen. Alternativ sind von den Benutzern erstellte Importvorlagen denkbar.

Da nur wenige Inhalte mit Geodaten versehen sind, wurde von deren Nutzung Abstand genommen. Auch hier wäre eine von dem Benutzer vorgenommene Zuweisung denkbar, beispielsweise über eine Karte. Ebenso könnte versucht werden automatisiert Straßen- oder Ortsnamen auszulesen. Ideal wäre es jedoch beim Veröffentlichungsprozess der Inhalte die Wichtigkeit der Geodaten hervorzuheben oder die Eingabe zu vereinfachen.

Einen Hauptteil der vorhandenen Ressourcen wird im PDF Format bereitgestellt. Für die Archivierung von Dokumenten ist dies ein geeignetes Format. Für die maschinelle Verarbeitung

ist dieses Format jedoch nicht optimal. Es konnte keine zufriedenstellende Bibliothek zur Manipulation gefunden werden. Zumal die Dokumente zum Teil eingescannt wurden und mit handschriftlichen Notizen versehen sind.

Um die Kontrolle des staatlichen Handelns zu ermöglichen sind die Inhalte ideal. Um jedoch die Dokumente zugänglicher machen zu können, wären offenere Formate besser geeignet.

Technische Probleme

Die Umsetzung wurde mit der neusten erschienenen Play Version begonnen. Diese Version beinhaltet noch einige kleinere Fehler eingebaut. Zudem waren die meisten Dokumentationen noch nicht an die Neuerungen der Version angepasst. Genauso wenig waren externe Projekte, wie der eingesetzte Datenbank Treiber, an die Version angepasst.

Realisierungsproblem

Beim produktiven Einsatz ist zu bedenken, dass der Erfolg der Systemteile mit Benutzerinteraktion stark von einer großen Nutzerbasis abhängt.

6.3.2. Zielerfüllung

Ziel dieser Arbeit ist es einen verbesserten Zugang zu offenen Daten zu finden.

Mehr Bürger sollen die offenen Daten nutzen können und die Zusammenarbeit soll gefördert werden. Zudem sollen die offenen Daten besser dargestellt und auffindbar werden.

Dazu wurde in der Literatur der Stand der Forschung recherchiert. Die gesammelten Ideen und eigene Überlegungen wurden in der Analyse verbunden.

Das Konzept erfüllt das Ziel, indem es eine mögliche Architektur eines verbesserten Portals aufzeigt. Es ist vorgesehen, dass die Bürger bei der Bewertung, Analyse und Interpretation unterstützt werden. So haben die Bürger die Möglichkeit offene Daten anzureichern und nicht nur die Option die offenen Daten passiv zu konsumieren.

Die prototypische Umsetzung setzt Teile des Konzeptes um. Die Umsetzung konzentriert sich auf das derzeitige Hauptproblem von Open Data Portalen. So werden im Prototypen Möglichkeiten getestet, über visuelle Darstellungen die Inhalte der offenen Daten schneller zu erfassen.

Die Vorschau auf die Inhalte ist ein erster Ansatz. Die Word Cloud Darstellung kann zusammen mit der Word Tree Darstellung die Inhalte einiger Daten schneller erfassbar darstellen. Erstellbare Diagramme helfen vorhandene Messergebnisse oder Statistiken schneller zu verstehen.

6. Umsetzung

In der Tabelle 6.8 und der Tabelle 6.9 ist dargestellt inwieweit die Anwendungsfälle und die User Stories umgesetzt wurden.

Ein verbesserter Zugang zu offenen Daten wurde also konzeptionell vorgestellt. Zudem wurde mit dem erstellten Portal zur Visualisierung von Open Data mit dem Play Framework der Grundstein für ein verbessertes Open Data Portal gelegt.

Anwendungsfall	K	U
Anwendungsfall „Informationen suchen“	+	+
Anwendungsfall „Visuell suchen“	+	+
Anwendungsfall „Visualisierung erstellen“	+	+
Anwendungsfall „Metadaten ändern“	+	T

Tabelle 6.8.: Umsetzung der Anwendungsfälle

User Story	K	U
1. Datensätze kombinieren	+	T
2. Gruppen finden	+	-
3. Informiert werden	+	-
4. Anfragen stellen	+	-
5. Geschichten erzählen	+	-
6. Userdaten kontrollieren	+	T
7. Benutzer verwalten	+	-

Tabelle 6.9.: Umsetzung der User Stories

6.4. Zusammenfassung

In diesem Kapitel wurde die prototypische Realisierung beschrieben. Dazu wurde der Realisierungsumfang verdeutlicht. Anschließend wurden einige Besonderheiten und die eingesetzten Bibliotheken präsentiert.

Zum Abschluss wurden einige Probleme und deren mögliche Lösungen erwähnt und die Resultate dieser Arbeit mit der Zielsetzung der Arbeit verglichen.

7. Schluss

Ein Portal zur Visualisierung von Open Data mit dem Play Framework wurde entworfen und umgesetzt.

Abschließend wird die Arbeit zusammengefasst, mögliche Erweiterungen der Arbeit werden vorgestellt und es wird ein Fazit gezogen.

7.1. Zusammenfassung

Wie können offene Daten besser dargestellt und auffindbar werden, sodass mehr Bürger die offenen Daten nutzen und wie können die Bürger bei der Analyse unterstützt werden. Das waren die zentralen Fragestellungen dieser Arbeit.

Dazu wurde in der Einleitung das Themenfeld vorgestellt. Anschließend wurden die Probleme des vorhandenen Hamburger Transparenzportals dargestellt. So bietet das Portal nur Zugang zu den vorhandenen Daten, eine weiterführende Analyse ist auf dem Portal nicht möglich.

Im Grundlagen Kapitel wurden Begriffe und die verwendeten Techniken Word Cloud, Word Tree und Dokumentenvorschau vorgestellt. Anschließend wurde das eingesetzte Play Framework und weitere eingesetzte Webtechnologien vorgestellt, wie CKAN als zentrale Schnittstelle zum vorhandenen Open Data Portal.

Im folgenden Open Data Kapitel wurden offene Daten als frei verwendbare Daten definiert. Anschließend wurden die verschiedenen Arten von offenen Daten vorgestellt, diese umfassen vor allem staatlich erhobene Daten. Open Data Portale wurden eingeführt mit ihren vier Hauptfunktionen: Veröffentlichung, Suche, Analyse und Diskussion. Zum Abschluss wurde das Hamburger Transparenzportal und seine Inhalte präsentiert.

In dem Kapitel Analyse wurde anhand eines Szenarios die Hauptanwendungsfälle eines verbesserten Portals herausgearbeitet. Dazu gehören, der Anwendungsfall „Informationen suchen“, der Anwendungsfall „Visuell suchen“, der Anwendungsfall „Visualisierung erstellen“, der Anwendungsfall „Metadaten ändern“ und das gemeinsame Arbeiten. Aus diesen wurden die Anforderungen an das System abgeleitet.

Anschließend wurde in dem Kapitel Konzeption eine Architektur vorgestellt. Zur Umsetzung der Anforderungen basiert diese auf den Komponenten: Suche, UserMetadaten, Visualisierungen, Dateien, Kontrolle, CKAN Adapter, Persistenz und Zusammenarbeit.

Abschließend wurde im Kapitel Umsetzung der Umfang der prototypischen Realisierung angegeben. Darauf folgte eine Beschreibung der Umsetzung sowie der dabei getroffenen Entscheidungen. Am Ende wurden mögliche Probleme dargestellt und die Arbeit im Hinblick auf ihre Zielsetzung überprüft.

7.2. Erweiterungen

Im Folgenden werden einige mögliche Erweiterungen der Arbeit vorgestellt.

Relativ einfach ließe sich das System um mehr Inhalte erweitern, dazu müsste die Suche bzw. der CKAN Adapter auf CKAN Instanzen anderer Open Data Portale erweitert werden. Dann wäre es zum Beispiel einfacher, ähnliche Daten verschiedener Gebiete zu vergleichen. Allerdings gibt es zumindest in Deutschland bislang nicht so umfangreiche Open Data Portale wie das Hamburger Transparenzportal. Zu bedenken sind außerdem die eventuellen Unterschiede der vorhandenen Metadaten und deren Bedeutung.

Die Inhalte der Daten sind meist in einer sehr formellen Sprache gehalten. Umgangssprachlich wird dies auch Behördendeutsch genannt. Um das Verständnis einiger Daten zu fördern würde es sich anbieten eine Art Wörterbuch aufzubauen.

Für einen produktiven Einsatz wäre es empfehlenswert, die grafische Gestaltung zu optimieren. So könnten andere Visualisierungstools genutzt werden oder die Visualisierungen neu programmiert werden. Alternativ gibt es inzwischen mehrere CKAN Erweiterungen die für das Hamburger Transparenzportal evaluiert werden können.

Zur Überprüfung der in dieser Arbeit verwendeten Konzepte sind weitere Untersuchungen nötig. So ist zu prüfen, ob die Inhalte schneller und besser verständlich aufbereitet werden konnten. Zusätzlich sind Usability Tests erforderlich. Diese Tests sollten von einer externen unabhängigen Person durchgeführt werden.

Für einen produktiven Einsatz ist ebenfalls eine Überprüfung der Sicherheit nötig, insbesondere im Hinblick auf die von Benutzern hoch geladenen Dateien und Daten.

Um die Datenanalysefähigkeiten zu verbessern, könnte überprüft werden, ob Teile einer Business Intelligence Software, wie Tableau oder dem Waikato Environment for Knowledge Analysis Tool, dazu geeignet und integrierbar sind. Der Fokus sollte dabei auf einfache Benutzbarkeit gelegt werden. Zusätzlich sollten auch Technologien die unter dem Schlagwort Big Data stehen geprüft werden.

7.3. Fazit

Offene Daten ermöglichen eine Kontrolle des staatlichen Handelns und können die demokratische Meinungs- und Willensbildung fördern.

Bislang wurde nur ein Teil des Potentials von offenen Daten genutzt. Ein verbesserter Zugang zu den offenen Daten kann mehr Bürger dazu animieren die offenen Daten zu nutzen.

Gemeinsam können die Bürger und Bürgerinnen die Demokratie stärken. Überdies können sie die Schätze aus den Aktenschranken der öffentlichen Verwaltung heben.

A. CD Inhalt

Dieser Arbeit liegt eine CD mit folgendem Inhalt bei:

- **Quellcode/**
- **apertusregimen-1.0.zip**
- **BA_Schomann.pdf**
- **BA_SchomannEclipseProject.zip**
- **ReadMe.txt**

Abbildungsverzeichnis

2.1. Word Cloud Beispiel	7
2.2. Word Tree Beispiel	8
2.3. Play Framework MVC	11
3.1. Transparenzportal Hamburg	23
4.1. Systemkontext	27
5.1. Kontextabgrenzung	37
5.2. Kontextdetails	38
5.3. Komponenten	40
6.1. Recline.js Tabellenansicht	53
6.2. Recline.js Diagramm	53

Tabellenverzeichnis

1.1. Inhalte des Hamburger Transparenzportals	2
3.1. Open Data Portal Funktionen für Datenbereitsteller	18
3.2. Open Data Portal Funktionen für Datennutzer	18
4.1. Anwendungsfall „Informationen suchen“	28
4.2. Anwendungsfall „Visuell suchen“	29
4.3. Anwendungsfall „Visualisierung erstellen“	30
4.4. Anwendungsfall „Metadaten ändern“	31
6.1. Umsetzung der Anforderungen in der Komponente Suche	50
6.2. Umsetzung der Anforderungen in der Komponente Dateien	51
6.3. Umsetzung der Anforderungen in der Komponente UserMetadaten	52
6.4. Umsetzung der Anforderungen in der Komponente Visualisierungen	54
6.5. Umsetzung der Anforderungen in der Komponente Zusammenarbeit	54
6.6. Umsetzung der Anforderungen in der Komponente Kontrolle	55
6.7. Umsetzung der nichtfunktionalen Anforderungen	55
6.8. Umsetzung der Anwendungsfälle	57
6.9. Umsetzung der User Stories	57

Literaturverzeichnis

- [ALC14] Charalampos Alexopoulos, Euripidis Loukis, and Yannis Charalabidis. A Platform for Closing the Open Data Feedback Loop based on Web2.0 functionality. *JeDEM*, 6(1):62–68, 2014.
- [Ass15a] CKAN Association. CKAN API, 2015. <http://docs.ckan.org/en/latest/api/index.html> - Zugriffsdatum: 03.08.2015.
- [Ass15b] CKAN Association. CKAN Association – Managing and Overseeing CKAN, 2015. <http://ckan.org/about/association/> - Zugriffsdatum: 03.08.2015.
- [Ass15c] CKAN Association. CKAN Feature Tour, 2015. <http://ckan.org/features/> - Zugriffsdatum: 03.08.2015.
- [Ass15d] CKAN Association. CKAN instances around the world, 2015. <http://ckan.org/instances/> - Zugriffsdatum: 31.08.2015.
- [AZC⁺14] Charalampos Alexopoulos, Anneke Zuiderwijk, Yannis Charapabidis, Euripidis Loukis, and Marijn Janssen. Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media. In *Electronic Government*, pages 230–241. Springer, 2014.
- [Bra14] Tim Bray. The JavaScript Object Notation (JSON) Data Interchange Format. RFC 7159, RFC Editor, March 2014. <https://tools.ietf.org/html/rfc7159> - Zugriffsdatum: 03.08.2015.
- [Bü12] Hamburgische Bürgerschaft. Hamburgisches Transparenzgesetz (HmbTG), 2012. <http://transparenz.hamburg.de/das-hmbtg/> - Zugriffsdatum: 03.08.2015.
- [CLA14] Yannis Charalabidis, Euripides Loukis, and Charalampos Alexopoulos. Evaluating Second Generation Open Government Data Infrastructures Using Value Models. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 2114–2126. IEEE, 2014.

- [Dav15] Jason Davies. Word Tree zu Schlagwortwolke, 2015. <https://www.jasondavies.com/wordtree/?source=179d1d34253dfc13ec5d1c9bcd3bd308> - Zugriffsdatum: 19.10.2015.
- [dG15] Kevin de Groote. Word Cloud zu Schlagwortwolke, 2015. <http://worditout.com/word-cloud/1256419> - Zugriffsdatum: 16.11.2015.
- [dl12] Bundesministerium des Innern. Eine Studie zu Open Government in Deutschland, 2012. http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.html - Zugriffsdatum: 03.08.2015.
- [DSH11] Daniel Dietrich and Christiane Schulzki-Haddouti. Dossier Open Data, 2011. <http://www.bpb.de/gesellschaft/medien/opendata/> - Zugriffsdatum: 03.08.2015.
- [e.V15] Open Knowledge Foundation Deutschland e.V. Offene Daten, 2015. <http://okfn.de/themen/offene-daten/> - Zugriffsdatum: 03.08.2015.
- [GH13] Alvaro Graves and James Hendler. Visualization tools for Open Government Data. In *Proceedings of the 14th Annual International Conference on Digital Government Research*, pages 136–145. ACM, 2013.
- [HBC13] Peter Hilton, Erik Bakker, and Francisco Canedo. *Play for Scala: Covers Play 2*. Manning Publications Co., 2013.
- [KK11] Rianne Kaptein and Jaap Kamps. *Word Clouds of Multiple Search Results*. 2011. http://www.researchgate.net/profile/Jaap_Kamps/publication/221309308_Word_Clouds_of_Multiple_Search_Results/links/02e7e52b4007644a1f000000.pdf - Zugriffsdatum: 12.05.2015.
- [Lab15] Open Knowledge Labs. Recline.js Multiview Demo, 2015. <http://okfnlabs.org/recline/demos/multiview/> - Zugriffsdatum: 29.10.2015.
- [LdK14] Nicolas Leroux and Sietse de Kaper. *Play for Java*. Manning Publications Co., 2014.
- [LT14] Jens Lattmann and Anina Trautermann. Hamburg öffnet die Aktenschränke. *Datareport*, 2014(3):18–21, 2014.
- [McC10] David McCandless. TED talk: The beauty of data visualization, 2010. http://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization - Zugriffsdatum: 24.03.2015.

- [Pro14] ENGAGE Project. Introduction to the ENGAGE platform, 2014. http://www.engagedata.eu/wiki/introduction_to_the_ENGAGE_platform/ - Zugriffsdatum: 03.08.2015.
- [SP14] Klaus-Peter Schoeneberg and Jennifer Pein. Entscheidungsfindung mit Big Data-Einsatz fortschrittlicher Visualisierungsmöglichkeiten zur Komplexitätsbeherrschung betriebswirtschaftlicher Sachverhalte im Unternehmen. In *Komplexitätsmanagement in Unternehmen*, pages 309–354. Springer, 2014.
- [Sta09] Gernot Starke. *Effektive Software-Architekturen*. Hanser, 2009.
- [uHH15] Freie und Hansestadt Hamburg. Transparenzportal Hamburg, 2015. http://suche.transparenz.hamburg.de/?q=HAW+Hamburg&sort=score+desc&title_sort+asc - Zugriffsdatum: 19.10.2015.
- [uKG15a] Geschäfts und Koordinierungsstelle GovData. Datenlizenz Deutschland – Namensnennung – Version 2.0, 2015. <https://www.govdata.de/dl-de/by-2-0> - Zugriffsdatum: 03.08.2015.
- [uKG15b] Geschäfts und Koordinierungsstelle GovData. Fragen und Antworten zu GovData, 2015. <https://www.govdata.de/faq> - Zugriffsdatum: 03.08.2015.
- [WFR⁺01] Allison Woodruff, Andrew Faulring, Ruth Rosenholtz, Julie Morrision, and Peter Pirolli. Using thumbnails to search the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 198–205. ACM, 2001.
- [WV08] Martin Wattenberg and Fernanda B. Viégas. The Word Tree, an Interactive Visual Concordance. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1221–1228, 2008.
- [ZJD14] Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1):17–33, 2014.
- [ZJJ13] Anneke Zuiderwijk, Marijn Janssen, and Keith Jeffery. Towards an e-infrastructure to support the provision and use of open data. In *Conference for E-Democracy and Open Government*, page 259, 2013.
- [ZJP13] Anneke Zuiderwijk, Marijn Janssen, and Armin Parnia. The complementarity of open data infrastructures: an analysis of functionalities. In *Proceedings of the 14th Annual International Conference on Digital Government Research*, pages 166–171. ACM, 2013.

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, 18. November 2015

Max Morten Schomann