

Location- & Context-based Surveys and Storytelling

LoCobSS - Abschlussbericht

Dr. Sebastian Meier | VISLAB.io

April 2021

Projektteam:

Sebastian Meier @ VISLAB.io, Projektleitung und Entwicklung

Fabian Dinklage, Visualisierung und Entwicklung

Katrin Glinka, Inhalte und Evaluation

Inhaltsverzeichnis

1. Zusammenfassung	2
1.1 Executive Summary	2
1.2 Projekterfolgskontrolle	5
2. Datenschutzkonforme Erhebung von Nutzer*innendaten	7
2.1 Datensparsamkeit & Datenvermeidung	7
2.2 Trennen von Datenbeständen	7
2.3 Privatsphärenschutz und das Erheben von Attributen	8
2.4 Möglichkeiten zur Erhebung sensibler Information	8
3. Analyse der Fragen: Unterstützung der Redakteur*innen	12
3.1 Testdaten	12
3.2 Hate Speech	12
3.3 Kategorisierung	13
3.4 Sentiment Analysis	18
4. Wissenschaftskommunikation	20
4.1 Daten-gestütztes Storytelling	20
4.2 Personalisierung und Relevanz	22
4.3 Umsetzung	25
4.4 Prototyp I - Klimawandel und Mobilität	25
4.5 Prototyp II - Klimawandelrisiken in Deutschland	30
5. Prototyp: Technische Dokumentation	36

1. Zusammenfassung

Das Forschungsvorhaben LoCobSS wurde im Rahmen “[Ideenwettbewerb\(s\) für innovative analoge und digitale Partizipationsformate und -technologien](#)” gefördert (Förderkennzeichen: [16IP103](#)).

Im **Wissenschaftsjahr 2022** soll ein großangelegtes Partizipationsverfahren durch das BMBF durchgeführt werden, um die Perspektive der Bürger*innen auf Wissenschaft und Forschung in Deutschland zu erfassen. Ein [ähnliches Projekt](#) gab es bereits 2018 in Belgien. Der Ideenwettbewerb soll innovative Konzepte für die Umsetzung des Prozesses in Deutschland generieren. Dabei sollen sich die Ideen auf drei Phasen fokussieren:

1. Aktivierung & Befragung der Bürger*innen
2. Analyse, Strukturierung und Auswertung der Fragen
3. Beantwortung und Nachnutzung der Fragen

Das Vorhaben LoCobSS konzentriert sich auf digitale Aspekte des Prozesses, mit einem spezifisch Fokus auf **privatsphären-konforme Customization der User-Experience**, sowohl für die **Bürger*innen** als auch die **Redakteur*innen** des Ministeriums. Um die entwickelten Methoden und Konzepte zu demonstrieren wurden eine Reihe an Prototypen entwickelt, welche in den folgenden Kapiteln vorgestellt werden.

1.1 Executive Summary

Im Rahmen von LoCobSS wurden eine Reihe Methoden entwickelt und evaluiert in Bezug auf einen Mehrwert für das Partizipationsverfahren. Die dabei entstandenen Softwarekomponenten können als prototypische Grundlagen für eine finale Implementation dienen (siehe [technische Dokumentation](#)). Doch noch viel wichtiger, sollen die folgenden Erkenntnisse Ausschreibungen und weitere Entwicklungen zum Partizipationsverfahren leiten und unterstützen:

1.1.1 Privatsphäre sollte höchste Priorität haben

Nicht zuletzt die Entwicklung der Corona-App haben gezeigt, dass das Bewusstsein in der Bevölkerung für das Thema Privatsphäre weiter zunimmt. Als öffentliche Institution sollte diesem Thema höchste Priorität gewidmet werden.

Der Schutz der Privatsphäre steht einer Erhebung statistischer Daten nicht im Wege. Es sollten dabei einige Grundsätze beachtet werden:

- Nur Daten erheben die tatsächlich *notwendig* sind (Datensparsamkeit & Datenvermeidung).
- Wenn Daten nicht in personenbezogener Form benötigt werden, User-Daten und statistische Daten (umkehrbar) trennen.
- Vermeiden von Dark-Design Patterns (z.B. opt-in statt opt-out).

Weitere Hilfestellungen finden Sie in [Kapitel 2](#).

1.1.2 Barrierearm und nutzerfreundlich

Um eine möglichst breite Gruppe der Bevölkerung am Verfahren zu beteiligen, sollten Barrieren niedrig gehalten werden:

- Einfache Sprache wählen
- Möglichst wenig Pflichtfelder
- Möglichst wenig Eingabebeschränken/-pflichten
- Keinen Registrierungszwang
- Mehrsprachigkeit unterstützen
- Den Nutzungsgewohnheiten und -ansprüchen entsprechen (z.B. Mobil-optimiert)

Weitere Hilfestellung hierzu im Prototypen zur Erfassung von Fragen in [Kapitel 2](#).

Kommentar: Dieses Vorhaben hat sich auf digitale Komponenten des Partizipationsverfahrens konzentriert. Für eine besonders inklusive Gestaltung des Verfahrens, sollten die digitalen Komponenten nur ein Teil der Aktivierungsstrategie darstellen und mit analogen Komponenten und einer breit aufgestellten Outreach-Kampagne kombiniert werden.

1.1.3 Auch an die Redakteur*innen denken

Während der Fokus klar auf die Bürger*innen liegt, sollte nicht vergessen werden, dass es auch ein gut ausgestattetes Team zur Redaktion der Inhalte benötigt. Das Verfahren in Belgien hat über 10.000 Fragen generiert. Entsprechend ist davon auszugehen, dass in Deutschland über 50.000 Fragen zu erwarten sind. Die eingehenden Fragen müssen validiert (Relevanz, Schimpfwörter, etc.) und kategorisiert werden. LoCobSS hat die Möglichkeiten der Automatisierung dieser Schritte auf Basis der Daten des belgischen Verfahrens evaluiert und zieht daraus folgende Schlüsse:

- Verfahren zur automatisierten Erkennung von Beschimpfungen, Hate-Speech, etc. erkennen diese bisher nur teilweise. Dies kann lediglich unterstützend angewandt werden und muss nach wie vor manuell gemacht werden.

- Sentimentanalysen zur Erfassung von Einstellungen zu bestimmten Themen funktionieren auch nur bedingt und sollten nicht für kritische Entscheidungen eingesetzt werden.
- Eine Unterstützung der Kategorisierung kann durch Verfahren der Inhaltsanalyse (z.B. word2vec ähnliche Verfahren) essentiell unterstützt werden. Hierzu haben wir einen funktionsfähigen Prototypen entwickelt.

Insgesamt lässt sich zu automatisierten Verfahren sagen, dass viele auf die englische Sprache hin optimiert sind. Durch die Übersetzung der Inhalte kann dies in vielen Bereich umgangen werden. Bei sprachspezifischen Problemen, wie z.B. Beleidigungen stoßen diese Ansätze aber an ihre Grenzen.

Details zum Prototypen zur Kategorisierung in [Kapitel 3](#).

1.1.4 Customization: Content-Recommendation

Durch Customization-Methoden kann die User-Experience personalisiert werden und so zu einer intensiveren Auseinandersetzung mit der Plattform führen. Um das explorieren der Inhalte auf der Plattform zu erleichtern und Serendipität zu fördern, sollte neben einer klassischen Suche auch Recommender-Funktionalitäten integriert werden. Hierbei unterscheidet man zwischen Content-based Recommendation und Collaborative Recommendation. Für letzteren Ansatz müssen die Interaktionen der Nutzer*innen erfasst werden. Aus Privatsphäregründen sollte dies nur mit bedacht durchgeführt werden. Im Rahmen dieses Vorhabens, haben wir einen prototypischen Content-based Recommender entwickelt, welcher auf den selben Prinzipien des Redaktionswerkzeugs zur Kategorisierung aufbaut.

Mehr zum Recommender in [Kapitel 3](#).

1.1.5 Customization: Wissenschaftsvermittlung

Klassische Wissenschaftsvermittlung ist traditionell statisch, unidirektional und textlastig. Im Rahmen von LoCobSS haben wir mit modernen Methoden des Daten-Journalismus und das data-driven Storytellings, neue Formate zur Vermittlung wissenschaftlicher Formate entwickelt, welche sich individuell personalisieren lassen. So wird die Lebenswelt der Leser*innen in die Kommunikation mit einbezogen und erlaubt den Leser*innen so direkte Bezüge herzustellen. Durch interaktive Elemente wird die Aufmerksamkeit der Nutzer*innen gebunden. Explorative Elemente laden zu einer tiefergehenden Auseinandersetzung mit der Materie ein. Die Personalisierung erfolgt ebenfalls mit Blick auf die Wahrung der Privatsphäre.

Eine detaillierte Erklärung der beiden Prototypen folgt in [Kapitel 4](#).

Kommentar: Sollten solche Formate häufiger ausgerollt werden, würde es Sinn machen diese Konzepte zu standardisieren und entsprechende Frameworks zur Verfügung zu stellen, sodass neue Themen einfach und nachhaltig erschlossen werden können.

1.2 Projekterfolgskontrolle

1.2.1 Forschungsfragen

Wie lassen sich Konzepte lebensweltlicher Taxonomien (z.B. Lebenslagen) mit Taxonomien wissenschaftlicher Wissensrepräsentation verbinden?

Wie lassen sich die Fragen der Bürger*innen von den Wissenschaftler*innen praxisnah und forschungsrelevant auswerten?

Es ist mit einer extrem großen Zahl an Inhalten zu rechnen, die von den Redakteur*innen bearbeitet werden müssen. Mit dem Ziel Teile dieser Aufgaben zu unterstützen oder automatisieren zu können, wurden verschiedene verfahren des Maschinellen Lernens (ML) exploriert. Die erfolgreichsten Ergebnisse wurden im Bereich der Kategorisierung (Taxonomien) erzielt. Hierzu wurde ein kooperatives ML-System entwickelt, welches die Redakteur*innen bei ihrer Arbeit unterstützt und gleichzeitig den Bürger*innen erlaubt die Inhalte einfach zu durchsuchen. Siehe hierzu [Kapitel 3](#).

Wie lässt sich eine solche Erhebung mit dem Schutz der Privatsphäre der teilnehmenden Personen vereinen?

Diese Dokumentation enthält [Empfehlungen](#) zur Konzeption und Umsetzung von Umfragen mit Fokus auf Wahrung der Privatsphäre der Teilnehmer*innen. Für die Erhebung von abgeleiteten Attributen, wurde ein client-side classification Ansatz entwickelt. Die Empfehlungen und client-side classification wurden beispielhaft in einem [Survey Prototypen](#) implementiert.

Wie kann die Relevanz wissenschaftlicher Inhalte für die Bürger*innen erhöht werden?

Welche Potentiale ergeben sich durch die Erhebung des räumlichen Kontextes in Hinblick auf die individuelle Relevanz der Inhalte?

In Fortführung der Personalisierungsmethoden wurden zwei exemplarische Wissenschaftskommunikations-Anwendungen entwickelt, welche den räumlichen Kontext der Leser*innen nutzen, um das interaktive Storytelling zu personalisieren. Mehr hierzu in [Kapitel 4](#).

1.2.2 Arbeitspakete

1. **Konzeption:** Die inhaltliche Konzeption ist in diese Dokumentation eingeflossen. Die technische Dokumentation befindet sich in [Kapitel 5](#).
2. **Anwendungsbasis:** Um die einzelnen Komponenten möglichst realistisch demonstrieren und testen zu können, wurde eine web-basierte Anwendung entwickelt, welche in rudimentären Zügen die spätere Beteiligungsanwendung abbildet (Fragen eingeben, Fragen & Teilnehmer administrieren, Fragen kategorisieren), siehe [Kapitel 5](#).
3. **Kontextualisierung:** Um die algorithmischen Verfahren zu automatischen Verarbeitung des Inputs der Bürger*innen evaluieren zu können, wurden die Daten des belgischen Verfahrens als Basis genutzt. Die resultierenden Komponenten wurden modular als Microservices aufgesetzt, um eine spätere Nachnutzung möglichst einfach zu gestalten.
4. **Data-driven Storytelling:** Es wurden zwei interaktive Anwendungen zur Vermittlung wissenschaftlicher Erkenntnisse entwickelt. Ein Prototyp zum Thema Klimawandel und Mobilität und ein Prototyp zum Thema Klimawandelrisiken in Deutschland. Die Anwendungen basieren auf Daten und Informationen des Umweltbundesamts.
5. **Evaluation:** Die Evaluation hat sich in Zeiten von Corona sehr schwierig gestaltet. Besonders der Kontakt zu Expert*innen innerhalb der Ministerien (was in diesen Zeiten durchaus verständlich ist). Wir haben die Konzepte und Prototypen deshalb mit Expert*innen aus unserem Netzwerk evaluiert und das Feedback eingearbeitet.
6. **Dokumentation:** Die Dokumentation wurde als Web- und PDF-Version umgesetzt. Alle entwickelten Software-Komponenten wurden als quellöffentlicher Source Code zur Verfügung gestellt. Für Links zu den individuellen Repositorien, siehe [Kapitel 5](#).

2. Datenschutzkonforme Erhebung von Nutzer*innendaten

2.1 Datensparsamkeit & Datenvermeidung

Grundsätzlich sollte bei jeder Form von Datenerhebung nur die Daten gesammelt werden, die zwingend notwendig sind. Während dies auf den ersten Blick völlig offensichtlich erscheint, verfolgen viele die Strategie so viele Daten wie möglich zu erheben, um z.B. nachträglich noch nicht vorgesehene Analysen zu machen. Die essentiellen Daten zu identifizieren verlangt sich im vornherein mehr Gedanken über die Nachnutzung der Daten zu machen. Verfolgt man aber diesen Ansatz, kann man die Nutzer*innen sehr transparent darüber informieren, wozu ihre Daten genutzt werden. Gleichzeitig kann der individuelle Fußabdruck der Datensammlung verringert werden.

2.2 Trennen von Datenbeständen

Gleichzeitig sollte definiert werden, ob es zwingend notwendig ist statistische Daten wie z.B. demographische Daten mit den Nutzerdaten (z.B. Username, Email, Passwort) abzuspeichern. Im Falle eines potentiellen Datendiebstahls, könnten sonst kriminelle die Nutzerdaten, wie Email-Adressen, spezifischen demographischen Attributen zuordnen. Ein Beispiel im Kontext der Befragung der Bürger*innen wo es Sinn macht diese beiden Ebenen zu trennen:

Ziel: Man möchte im Laufe des Prozesses und nach Abschluss des Projektes sicherstellen, dass eine möglichst gute Stichprobe der Bevölkerung an der Umfrage teilgenommen haben.

Dimensionen: Geschlecht, Alter und räumliche Verortung

Abfragemethoden:

1. Geschlecht: Aus Inklusionsgründen sollte hier ein breites Spektrum an Möglichkeiten angeboten werden. Dies birgt aber die Gefahr, dass Informationen über eine sehr kleine Untergruppe der Bevölkerung erhoben werden. Hier muss abgewogen werden, wie fein die Optionen und wie gut der Schutz der Privatsphäre sein soll.
2. Alter: Für das gestellte Ziel sollte es ausreichen grobe Altersklassen zu bilden.
3. Räumliche Verordnung: Umso größer die räumliche Auflösung, umso höher der Schutz der Privatsphäre. In vielen Fällen wird es wahrscheinlich ausreichen Bundesländer oder z.B. Postleitzahlen zu erheben.

Speicherung: Sind wir nur daran interessiert, dass die Fragen von einer guten Stichprobe der Gesellschaft gestellt werden, könnte man die demographischen Daten komplett von den Fragen und den Nutzerdaten trennen. So könnte man nachher die demographischen Attribute

weder Fragen noch Personen zuordnen. Könnte aber trotzdem Aussagen über die Stichprobe treffen.

2.2.1 Implementierung

Im Prototypen haben wir die Trennung von demographischen Daten und Nutzerdaten exemplarisch implementiert. Siehe hierzu [Prototypen-Dokumentation](#).

2.3 Privatsphärenschatz und das Erheben von Attributen

Das vorsichtige Abwägen beim Erheben von personenbezogenen Attributen ist notwendig, da über diese Personen in der realen Welt identifiziert werden können. Dies kann zum einen passieren, weil es innerhalb der untersuchten Zielgruppe sehr kleine Untergruppen mit bestimmten Attributen gibt. Wenn in der Datenbank nun eine Datensatz vorliegt, welcher genau diese seltenen Attribute hat, können wir diesen Datensatz einer Person zuordnen (oder wissen zumindest das eine von N Personen diesen Datensatz erstellt hat). Neben einer direkten Zuordnung darf nicht vergessen werden, dass diverse Unternehmen bereits enorme Mengen persönlicher Daten gesammelt haben. Kombiniert man diese mit den neu erhobenen Daten, können möglicherweise auch Rückschlüsse auf Individuen generiert werden.

Um diese Gefahren ganz praktisch anhand einiger Beispiele zu demonstrieren, haben wir eine [interaktive Demo](#) entwickelt.

Fig. 2.3.1 Demo: Daten erheben & die Privatsphäre schützen

The screenshot shows a user interface for data aggregation. At the top, there is a text box with a hand icon and the text: "Kombiniere Attribute und erfahre wie groß die kleinste Gruppe ist." Below this are three buttons labeled "Attribute: Alter", "Geschlecht", and "Nationalität". Underneath the buttons, the text "Die kleinste Zahl an Fällen (1) gibt es in der Gruppe: Frauen: 65; und älter: Berlin" is displayed in green. At the bottom, a small note says "Datensatz: 11 Altersgruppen, 2 Geschlechter, 125 Nationalitäten".

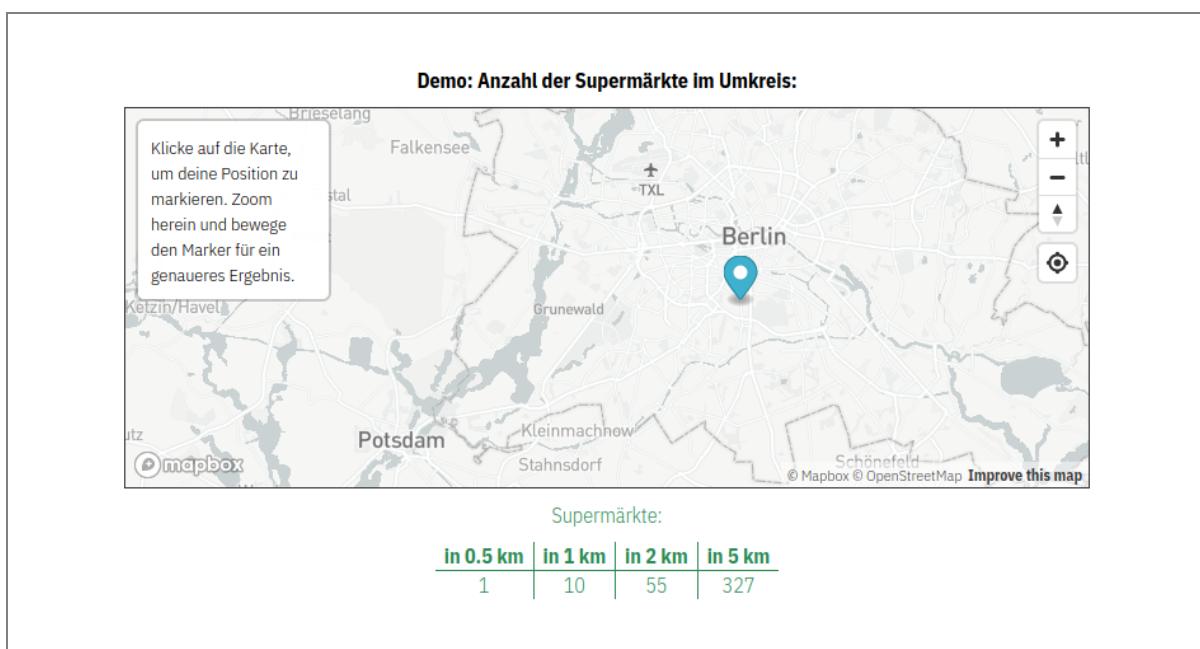
2.4 Möglichkeiten zur Erhebung sensibler Information

In manchen Fällen möchte man vllt. Daten erheben die sehr persönlich sind und eventuell die Privatsphäre von Individuen gefährden. Nehmen wir als Beispiel den exakten Wohnort der Person. Sollten wir tatsächlich konkrete Kontaktinformationen benötigen, z.B. zum

postalischen Kontakt, müssen die Daten selbstverständlich erhoben werden (aber auch hier sollte eruiert werden, ob z.B. ein Trennen von Daten Sinn macht). In einigen Fällen wird der Wohnort aber nur erhoben, weil man davon ausgehend andere Attribute ableiten möchte, z.B. wohnt jemand in einer Großstadt oder eher auf dem Land. In all solchen Fällen, in denen hochauflöste Attribute erhoben werden, um anschließend niedrig aufgelöste Attribute abzuleiten, empfiehlt es sich, dieses Ableiten direkt im Browser, auf Seiten der Nutzer*innen durchzuführen. Ist man z.B. daran interessiert, ob jemand in Großstadt oder auf dem Land wohnt, kann man eine einfache Postleitzahlabfrage integrieren und statt der Postleitzahl selber, das Attribut Stadt/Land zurückgeben. Solche **client-side classifications** sind für die Entwickler*innen komplexer zu implementieren, können aber so implementiert werden, dass diese für die Nutzer*innen keinen Unterschied in der Performance oder Interaktion darstellen.

Um diese Methode der **client-side classification** zu demonstrieren, haben wir eine [**interaktive Demo**](#) entwickelt.

Fig. 2.4.1 Demo: Client-side Classification



Implementierung

Im Prototypen haben wir die Trennung von demographischen Daten und Nutzerdaten exemplarisch implementiert. Zusätzlich haben wir die oben beschriebene client-side classification implementiert. Beispielhaft wird hierzu die eingegebene Postleitzahl genutzt, um direkt im Browser die [RegioStar Gem5](#) Klasse zu identifizieren (Metropole; Regiopole, Großstadt; Zentrale Stadt, Mittelstadt; Städtischer Raum; Kleinstädtischer / dörflicher Raum). Siehe hierzu [Prototypen-Dokumentation](#).

Fig. 2.4.2 Prototyp: Formular zum Stellen einer Frage

Frage stellen

Wir wollen eure Fragen sehen. Du kannst uns alles fragen. Versuche deine Frage möglichst präzise und konstruktiv zu formulieren. Alle Fragen werden erst überprüft, bevor sie auf der Seite veröffentlicht werden. Fragen die nicht mit unseren Richtlinien übereinstimmen werden gelöscht.

*Felder mit * sind Pflichtfelder. Alle anderen Felder sind optional.*

Deine Frage*

Weitere Hinweise zu deiner Frage

Dein Name*

Stell dich kurz vor:
Wir wollen, dass diese Umfrage ein möglichst diverses Bild der Gesellschaft wiederspiegelt, deshalb hilft es uns, etwas mehr über dich zu erfahren. Diese Informationen werden separat von deinen persönlichen Informationen gespeichert und können nicht mit dir in Verbindung gebracht werden.:

Postleitzahl **Dein Geschlecht** **Dein Alter**

In Kontakt bleiben?
Wenn du dich registrierst können wir dir eine Nachricht schicken, wenn es Neugkeiten zu deiner Frage gibt. Wir teilen deine Kontaktinformationen nicht mit Dritten und kontaktieren dich nur in Bezug auf dieses Projekt. Du kannst deinen Account jederzeit löschen.

Nein, Danke.

Ja, einen Account erstellen.

Ich bin bereits registriert

Um diese Seite abzusichern, müssen wir sicherstellen, dass du eine echte Person bist.
Es tut uns leid, aber Roboter dürfen nicht teilnehmen:

I'm not a robot  reCAPTCHA
Privacy - Terms

Frage hochladen

Weiterhin haben wir bei der Implementierung versucht sogenannte “Dark Design Patterns” zu vermeiden und “Ethical Design” Prinzipien zu verfolgen:

1. Nur die Felder zu Pflichtfeldern machen, die wirklich notwendig sind. Pflichtfelder können Personen auch davon abschrecken den Ausfüllprozess abzuschließen.
2. Möglichst barrierearme Anforderungen an die Eingaben stellen (z.B. mindest oder maximale Zeichenzahlen, Passwortvorgaben, etc.)
3. Transparent erklären wozu welche Daten erhoben werden und was damit geschieht. Zum Beispiel transparent erklären, warum demographische Daten erhoben werden und wie diese verarbeitet werden.
4. Opt-in und nicht Opt-out. Stattdessen sollten die Vorteile einer Registrierung deutlich gemacht werden und die Wahl der Nutzer*in überlassen werden.
5. Sicherheitsmaßnahmen und Privatsphäre müssen häufig abgewägt werden. Wenn wir Bürger*innen erlauben wollen Fragen zu stellen, ohne z.B. vorher ihre Email-Adresse zu validieren, müssen wir andere Maßnahmen integrieren, um illegale Aktivitäten auf der Seite zu verhindern. Wir haben als Beispielhafte Lösung [reCaptcha](#) implementiert. Wir haben explizit Version 2 genutzt, da die Privatsphäre bei dieser Variante besser geschützt wird.

3. Analyse der Fragen: Unterstützung der Redakteur*innen

Ein Großteil der Bemühungen rund um die Befragung der Bürger*innen stellt die Erhebung und Partizipation in den Mittelpunkt. Währenddessen sollte nicht vergessen werden, dass auch die Mitarbeiter*innen bei der Handhabung der eintreffenden Fragen unterstützt werden müssen. Wenn das Verfahren in [Belgien](#) als Grundlage genommen wird, sollten in Deutschland um die 50.000 Inhalte durch die Bürger*innen generiert werden. Diese Inhalte müssen überprüft, freigegeben und kategorisiert werden. Wenn jeder Frage nur eine Minute Aufmerksamkeit geschenkt wird, bedeutet dies über 100 Arbeitstage nur mit der reinen Administration der Inhalte.

Gut kategorisierte und kuratierte Inhalte stellen auch gleichzeitig einen Mehrwert für die Nutzer*innen der Plattform dar. Deshalb sehen wir einen doppelten Nutzen darin, die Redakteur*innen bei ihrer Arbeit zu unterstützen. Bezuglich dieser Aufgaben haben die Organisator*innen andere Beteiligungsverfahren und öffentlicher Plattformen uns zwei primäre Aufgaben genannt die unterstützt werden könnten: Identifizieren von unangebrachten Inhalten und das Kategorisieren von Inhalten.

Im Rahmen des Vorhabens LoCobSS haben wir uns speziell mit Methoden der automatisierten Verarbeitung von Inhalten beschäftigt. Im Folgenden präsentieren wir von uns prototypisch implementierte Dienste und deren Evaluation. Die technische Dokumentation befindet sich in [Kapitel 5](#).

3.1 Testdaten

Um die folgenden Ansätze evaluieren zu können, wurden vergleichbare Testdaten benötigt. Zu diesem Zweck wurde ein Harvester entwickelt, welcher die Daten des [belgischen Beteiligungsverfahrens](#) aggregierte und strukturiert für dieses Vorhaben aufbereitete.

Für weitere Details siehe [Repository](#) des Harvesters.

3.2 Hate Speech

Bevor Eingaben der Bürger*innen über die Plattform veröffentlicht werden, sollte diese dahingehend überprüft werden ob diese den inhaltlichen Guidelines entsprechen. Beispiele für Inhalte die hierbei identifiziert werden sollten sind z.B. rassistische, sexistische, obszöne oder gewaltverherrlichende Inhalte. Eine große Herausforderung bei der Nutzung automatisierter Verfahren ist, dass die meisten Datensätze auf englische Sprache hin

optimiert sind. Gerade bei Thematiken wie Schimpfwörtern sind aber sprachliche Nuancen und z.B. auch spezielle Schreibweisen sehr wichtig. Hinzu kommen zweideutige Sätze, welche von algorithmen nur schwierig erkannt werden. Entsprechend zeigte sich in den Tests auch, dass eine Menge unangebrachte Inhalte von den Methoden nicht erkannt wurden. Deshalb kann diese Funktion (zumindest im aktuellen Implementationszustand) nur als Hinweis, nicht aber als automatisierte Filter eingesetzt werden.

3.2.1 Implementierte Verfahren

Um Ansätze für englische Sprache nutzen zu können, haben wir die Eingaben der Nutzer*innen automatisiert ins englische übersetzt.

- [ProfanityFilter](#) (python)
- [HateSonar](#) (python)

3.2.2 Verbesserung

Sollte man solch einen Ansatz weiterverfolgen sollen, müsste man gute deutsche Trainingsdaten und deutsche Vokabularen generieren. Während dies ein grundsätzlich ländliches Vorhaben wäre, welches sicherlich auch anderen Entwickler*innen und Forscher*innen zu Gute käme, ist es fraglich ob dies im Rahmen eines solchen Vorhabens finanziell angemessen wäre.

3.2.3 Alternativen

Die meisten großen Content-Plattformen nutzen sogenannte Micro-Job oder Micro-Task Plattformen. Bei denen Personen für Kleinstaufgaben bezahlte werden. Größere Unternehmen wie z.B. Facebook haben eigene Content-Moderators. Beide Methoden sind nur bedingt zu empfehlen. Da die Arbeiter*innen dieser Systeme häufig unter dem Mindestlohn bezahlt werden, für psychisch stark beanspruchende Arbeit. Dies sollte bei ähnlichen Ansätzen berücksichtigt werden. Empfehlenswerter wäre das schaffen von temporären Projektstellen, welche angemessene Löhne und Arbeitsbedingungen sicherstellen.

3.3 Kategorisierung

Um die eingehenden Inhalte analysieren zu können und nachnutzbar zu machen, müssen diese kategorisiert werden. Hierzu müssen aus den gesamten Inhalten übergreifende Thematiken abgeleitet werden und die Inhalte entsprechend mit Schlagworten (Tags) oder Kategorien (Taxonomien) versehen werden. Im Gegensatz zu Betrachtung und Bewertung individueller Inhalte, müssen bei diesem Problem viele Inhalte gleichzeitig miteinander verglichen werden. Nach Experimenten mit verschiedenen Natural Language Processing (NLP)

und Machine Learning (ML) Verfahren, haben wir uns hierbei für einen zweistufigen Prozess entschieden, der vielversprechende Ergebnisse liefert.

3.3.1 Vektorisieren von Text

Ein in der modernen Text-Analyse weit verbreitetes Verfahren, ist das Umwandeln von Texten in Vektoren. Das Verfahren basiert darauf, dass ein künstliches Neuronales Netzwerk auf einem großen Text-Korpus trainiert wird. Stark abstrahiert, lernt das Netzwerk dabei Strukturen und Ähnlichkeiten von Textbausteinen. Später kann das Netzwerk genutzt werden, um Textbausteinen einen mehrdimensionalen Vektor zuzuordnen. Über diese Vektoren können dann z.B. Vergleiche angestellt werden, um die Ähnlichkeit von Textbausteinen festzustellen. Hierbei gibt es Verfahren die für einzelne Wörter (z.B. [word2vec](#)) optimiert sind und andere die ganze Texte analysieren können. Wir haben für diesen Schritt den von Google Forscher*innen entwickelten [Universal Sentence Encoder](#) in [Version 4](#) genutzt.

Zur einfachen Analyse wurde ein [Service](#) entwickelt, welcher Texte in 512-dimensionale Vektoren umwandelt.

Der Vorteil dieses Verfahrens ist, dass die Berechnung der Ähnlichkeiten genauso mit sehr wenigen Inhalten funktioniert, als auch sehr vielen. So können schon während die ersten Inhalte eintreffen, erste Analysen durchgeführt werden.

3.3.2 Ableiten von Ähnlichkeit

Aufbauend auf den Text-Vektoren lässt sich eine Distanzmatrix erstellen, um schnell Ähnlichkeiten zwischen einzelnen Texten abzuleiten. So können z.B. Content-Recommendations basiernd auf den Ähnlichkeiten durchgeführt werden. Hierzu haben wir einen einfachen [Service](#) entwickelt, welcher zu einem nutzergenerierten Inhalt die ähnlichsten Inhalte zurückgibt.

3.3.3 Kategorien identifizieren und Inhalte kategorisieren

Im Laufe der Entwicklung dieser Module haben wir unterschiedliche Methoden aus den Bereichen des Topic-Modellings und des Clusterings getestet. Topic-Modelling beschäftigt sich im NLP mit dem Identifizieren von übergreifenden Themen innerhalb eines Textes. Clustering Methoden auf der anderen Seite versuchen ähnliche Datensätze zu identifizieren, in unserem Fall basierend auf den Werten der Vektoren. Beide Ansätze haben nicht zufriedenstellende Ergebnisse geliefert. Deshalb haben wir das Prinzip des Clusterings genommen und daraus einen kollaborativen Ansatz entwickelt, bei welchem Nutzer*innen und algorithmen gemeinsam versuchen passende Gruppen zu bilden.

Fig. 3.3.1 Prototyp: Interface zum clustern von Inhalten

Cluster

Neue Taxonomie ▾
Name für neue Taxonomie

Ausgangsfrage
Wie kann die Wissenschaft sicherstellen, dass es in Zukunft genügend Trinkwasser gibt?

Fragen im Cluster
 Wie kann die Wissenschaft sicherstellen, dass es in Zukunft genügend Trinkwasser gibt?

Taxonomie anwenden

Potentielle Kandidaten

Anzahl Kandidaten je Frage: 10 ▾

- Wie lösen wir den Trinkwassermangel?
- Wie lange wird es genug Trinkwasser auf der Welt geben? (Jana, 14 Jahre alt)
- Wird es einen Tag geben, an dem unsere Trinkwasserquelle zur Neige geht?
- Wie können wir verschmutztes Wasser wieder trinkbar machen?
- Wie können wir sicherstellen, dass jeder auf der Welt mit sauberem Trinkwasser versorgt wird?
- Wie weit kann die Wissenschaft gehen, um Lösungen zur Erkennung / Bekämpfung von Krankheiten zu finden?
- Werden wir in Zukunft genug Trinkwasser haben?
- Wie können wir auf dem Planeten genug Trinkwasser liefern?
- Wie können wir der exponentiell wachsenden Bevölkerung weiterhin ausreichend gesunde Nahrung und Trinkwasser garantieren?

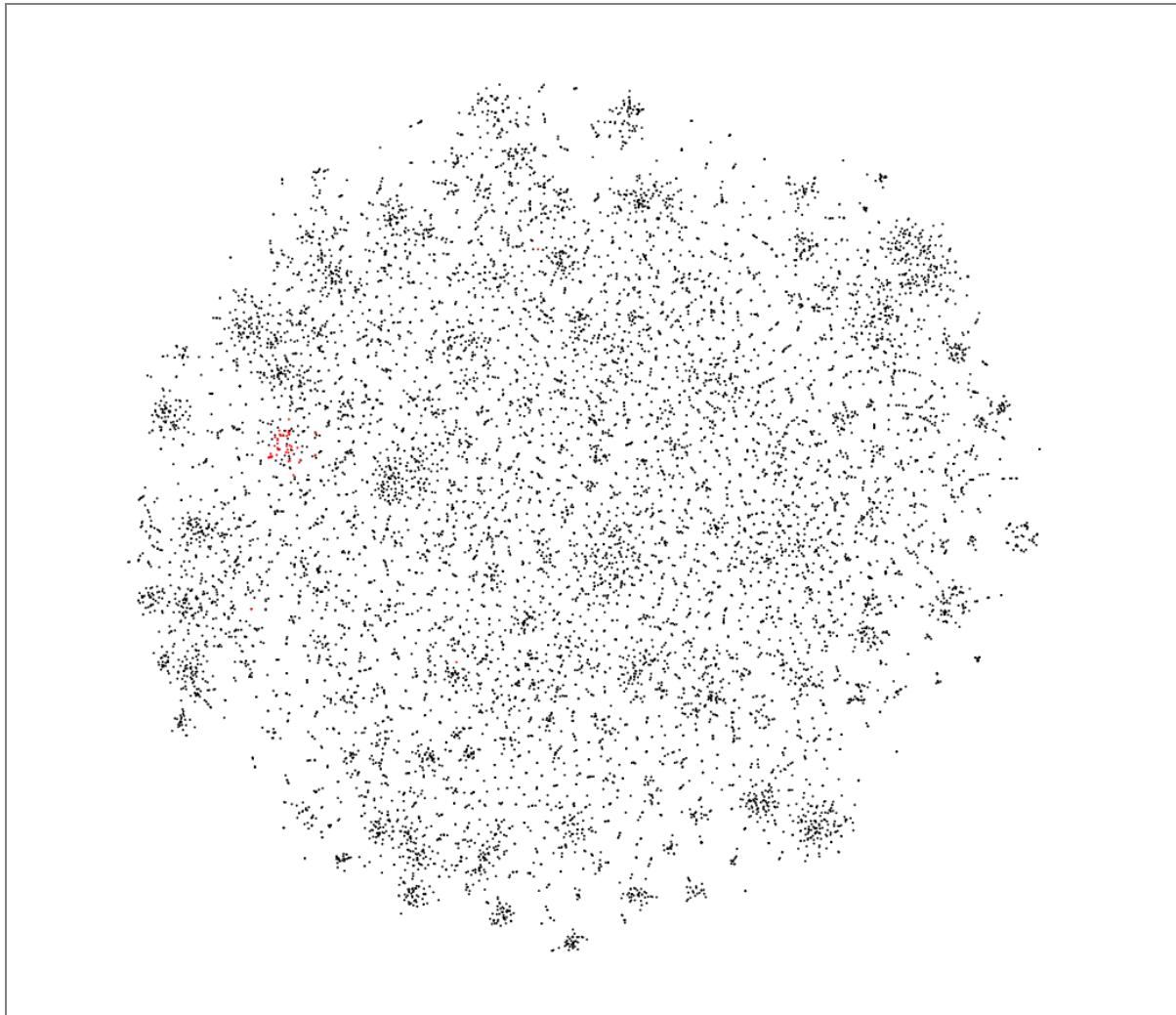
Auswahl zum Cluster hinzufügen und den Rest ausschließen

Ausgeschlossene Fragen

Ausgeschlossene Fragen zum Cluster hinzufügen

Die Redakteur*innen wählen dazu erst eine Inhalt als aus Ausgangspunkt aus (siehe Abbildung **Ausgangsfrage**). Das System sucht nun zu diesem Objekt ähnliche Inhalte heraus (siehe Abbildung **Potentielle Kandidaten**). Aus diesen Kandidaten können nun die ähnlichen Objekte ausgewählt werden. Diese werden dann dem aktuellen Cluster hinzugefügt. Alle nicht ausgewählten Kandidaten landen in der Liste ausgeschlossener Inhalte. Nachdem eine Auswahl getätigkt wurde, nutzt das System die neuen Objekte, um daraus weitere ähnliche Inhalte abzuleiten. So können mit nur wenigen Clicks hunderte Objekte in Kategorien kombiniert werden. Abschließend kann man der Sammlung einen Namen geben und so die Navigation und Suche für Besucher*innen der Seite unterstützen.

Fig. 3.3.2 Prototyp: Visualisierung der Ähnlichkeiten, in rot das Cluster zum Thema Trinkwasser



Als visuelles Feature werden die Inhalte der Plattform in einer zwei-dimensionalen Punktwolke abgebildet. Mit Hilfe einer Dimensionsreduktion ([t-SNE](#)) werden die Vektoren auf zwei Dimensionen reduziert. In dieser Wolke wird auch die aktuelle Auswahl des Cluster angezeigt. In der aktuellen Variante ist dies nur eine statische Visualisierung der ausgewählten Inhalte, zukünftig könnte man dies dahingehend erweitern, dass die Auswahl auch über die Visualisierung durchgeführt werden kann.

3.3.4 Nutzergenerierte Kategorien

Neben der Kategorisierung durch die Redakteur*innen haben wir ein weiteres Feature implementiert, was einen crowd-sourcing Ansatz bei der Kategorisierung verfolgt. Jede

Bürger*in die einen Inhalt auf der Plattform generiert, erhält nach dem Abschicken Vorschläge verwandter Inhalte. Die Bürger*innen können dann Feedback geben, ob diese Inhalte mit Ihrer Frage verwandt sind oder nicht.

Hinweis: Während die nutzerseitige Kategorisierung vollständig implementiert ist, fehlt noch eine Kombination mit dem System der Redakteur*innen. Die direkte Vektorisierung des Textes dauert im aktuellen Setup etwas länger. Aus Kostengründen geht der Klassifizierungsservice automatisch in einen Ruhezustand. Das "Erwachen" aus diesem Ruhezustand benötigt etwas Zeit. Bei einer finalen Integration könnte man dieses Delay auf ein Minimum reduzieren.

Fig. 3.3.3 Prototyp: Nutzergenerierte Ähnlichkeiten

Vielen Dank für das Erstellen deiner Frage. Unsere maschinellen Helfer haben einige Frage in unserer Sammlung gefunden, von denen sie glauben, dass sie mit deiner Frage verwandt sind. Du würdest uns sehr helfen, wenn du uns deine Meinung sagst, ob diese Fragen mit deiner verwandt sind:

Wird es in Zukunft genug Trinkwasser geben?

- Wie kann die Wissenschaft sicherstellen, dass es in Zukunft genügend Trinkwasser gibt?
nicht ähnlich ein bisschen ähnlich sehr ähnlich
- Wie lösen wir den Trinkwassermangel?
nicht ähnlich ein bisschen ähnlich sehr ähnlich

3.3.5 Content Recommendation

Ähnlich wie das Kategorisieren der Redakteur*innen, können die zugrundeliegenden Daten auch genutzt werden, um den Nutzer*innen der öffentlichen Seite Inhalte zu empfehlen. Diese Funktionalität ist bereits im aktuellen Prototypen integriert (siehe Abbildung unten). Wählt man eine Frage aus, werden ähnliche Inhalte angezeigt. Um Serendipität zu fördern, sind dies nicht nur die „ähnlichsten“ Inhalte sondern auch Inhalte die etwas weiter entfernt sind. Die Auswahl wird leicht randomisiert, sodass bei jedem Besuch andere ähnliche Inhalte angezeigt werden.

Fig. 3.3.4 Prototyp: Inhaltsempfehlungen

Wie kann die Wissenschaft sicherstellen, dass es in Zukunft genügend Trinkwasser gibt?

Das fällt mir in der wunderschönen Landschaft hier auf, dass alle Flüsse, die Sie hier überqueren oder an denen ich vorbeikomme, knochentrocken sind. Und dann kehren Sie zum Hotel zurück, wo ein Swimmingpool mit Hunderten von Kubikmetern Trinkwasser gefüllt ist. Der Kontrast könnte nicht größer sein.

👤 Conchita Barrera 20.02.2021

RELATED CONTENT

⭐ Atommüll ist das Problem der Zukunft. Was ist oder kann die Wissenschaft tun, um dieses Problem zu lösen? 20.02.2021

👤 Cathey Bolles

⭐ Warum entsalzen wir Wasser nicht in großem Umfang, um alle mit Trinkwasser zu versorgen? 20.02.2021

👤 Trish Brunkhorst

⭐ Wie können wir in Zukunft einen Mangel an Wasser, Nahrungsmitteln und anderen Rohstoffen beheben? 20.02.2021

👤 Kattie Walkins

⭐ Können wir auch in Zukunft genug Lebensmittel für alle produzieren, die den Klimawandel im Auge haben? 20.02.2021

👤 Bessie Sinquefield

3.4 Sentiment Analysis

Auch wenn diese Funktion nicht während unserer Recherchen als ein wichtiges Feature für eine Partizipationsplattform aufkam, haben wir die Funktion von automatisierten Sentiment Analysen im Rahmen der Entwicklung von Hate-Speech-Detection explored. Die Ergebnisse waren sehr durchmischt. Teils werden Einstellungen zu bestimmten Themen sehr gut erkannt, häufig werden Einstellungen aber als eher neutral vom System bewertet. Für zukünftige Entwicklungen haben wir auch hierzu einen [Service](#) entwickelt, welcher entsprechende Sentiment-Werte für einen eingegebenen Text zurückgibt.

4. Wissenschaftskommunikation

Auf Benutzer*innen zugeschnittenes, datengestütztes Storytelling im Web

Klassische Wissenschaftskommunikation ist nach wie vor statisch, unidirektional und textlastig. Die strukturelle Nähe zu akademischen Publikationen macht diese nur schwer zugängig für viele interessierte Bürger*innen. Alternativen zu diesen Kommunikationspraktiken finden sich beispielsweise im Bereich des Daten-Journalismus oder der Daten- und Informationsvisualisierung. Dort werden Ansätze entwickelt und genutzt, welche den aktuellen Seh- und Nutzungsgewohnheiten der Leser*innen entsprechen. In diesem Teilabschnitt des LoCobSS Vorhabens wurden zwei exemplarische Anwendungen entwickelt, welche aufzeigen sollen, wie man mit diesen Methoden des daten-gestützten Storytellings auch wissenschaftliche Inhalte bürgerlich vermitteln kann. Darüber hinaus wurden Methoden der Personalisierung entwickelt, um es den Leser*innen zu erleichtern Bezüge zwischen den wissenschaftlichen Ergebnissen und der eigenen Lebenswelt herzustellen.

4.1 Daten-gestütztes Storytelling

Die zunehmende Verfügbarkeit von Daten, barrieararme Werkzeuge zur Erstellung von Visualisierungen und Bemühungen von Akteur*innen wie Datenjournalist*innen, haben in den letzten Jahren dazu beigetragen, dass Visualisierungen von Daten und Informationen nicht mehr nur in der Fachliteratur zu finden sind, sondern Einzug in die Massenmedien gehalten haben. Besonders hervorzuheben für ihre herausragende datenjournalistische Arbeit auf diesem Gebiet sind z.B. folgende Plattformen:

Deutschland

- [Zeit-Online](#)
- [Spiegel Online](#)
- [Tagesspiegel](#)
- [Morgenpost](#)

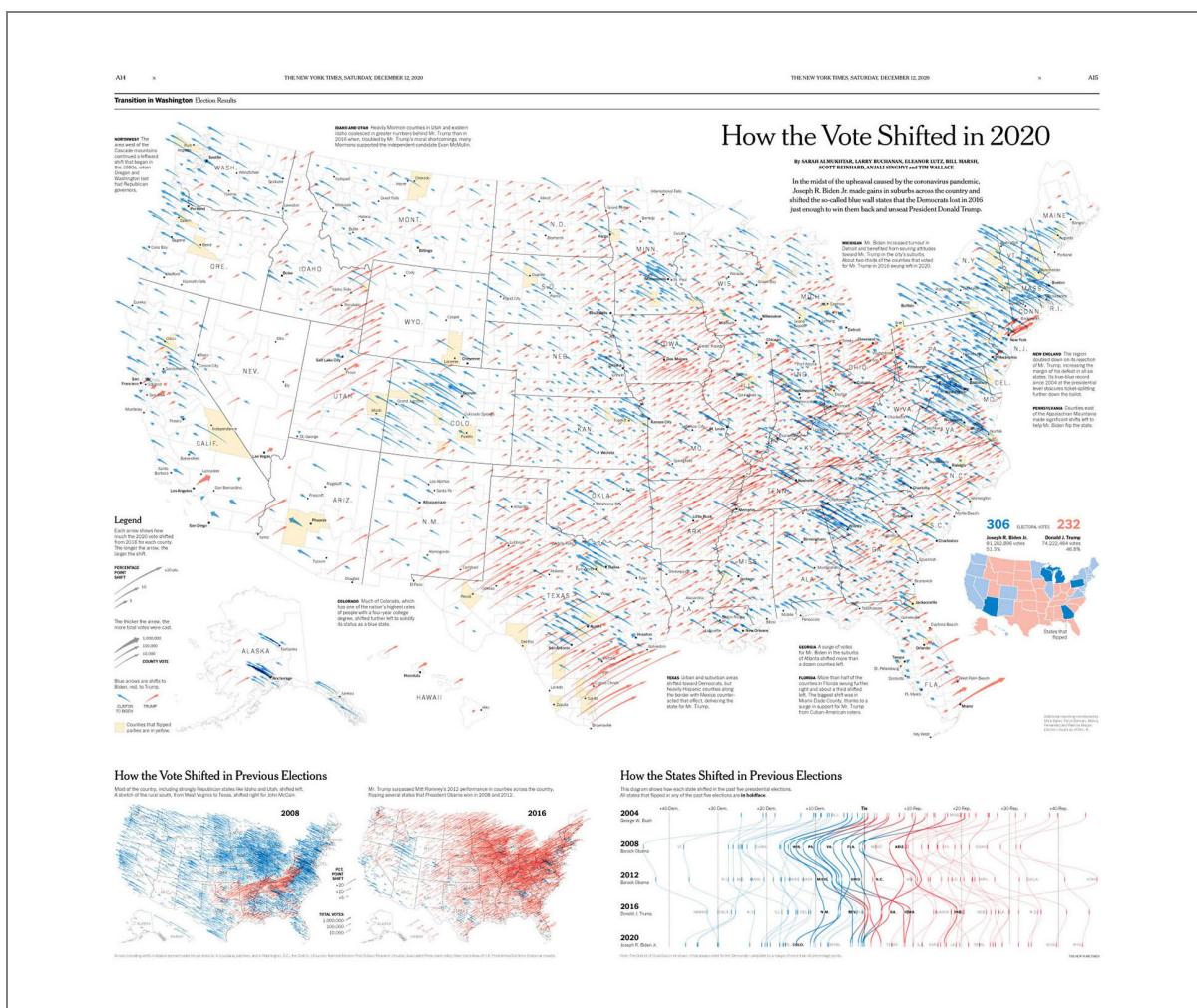
Europa

- [NZZ](#)
- [The Guardian](#)

Weltweit

- [New York Times](#)
- [Washington Post](#)
- [The Pudding](#)
- [Financial Times](#)
- [FiveThirtyEight](#)

Fig. 4.1.1 New York Times: Verschiebungen bei den Wahlen 2020



Wie in der beispielhaften Abbildung aus der New York Times oben zu sehen, gehen die Formen von Visualisierungen mittlerweile weit über einfache Balken- und Liniendiagramme hinaus und bieten komplexe Darstellungen und Analysemöglichkeiten. Diese komplexen Formen der Visualisierungen sollten aber mit Bedacht genutzt werden, da die Zunahme solcher

Darstellung nicht zwingend Hand-in-Hand geht mit der Visual-Literacy in der breiten Leserschaft. Weshalb wir in den folgenden Prototypen versucht haben die visuelle und informationelle Komplexität möglichst gering zu halten.

Weiterführende Lektüre zum Thema Datenjournalismus:

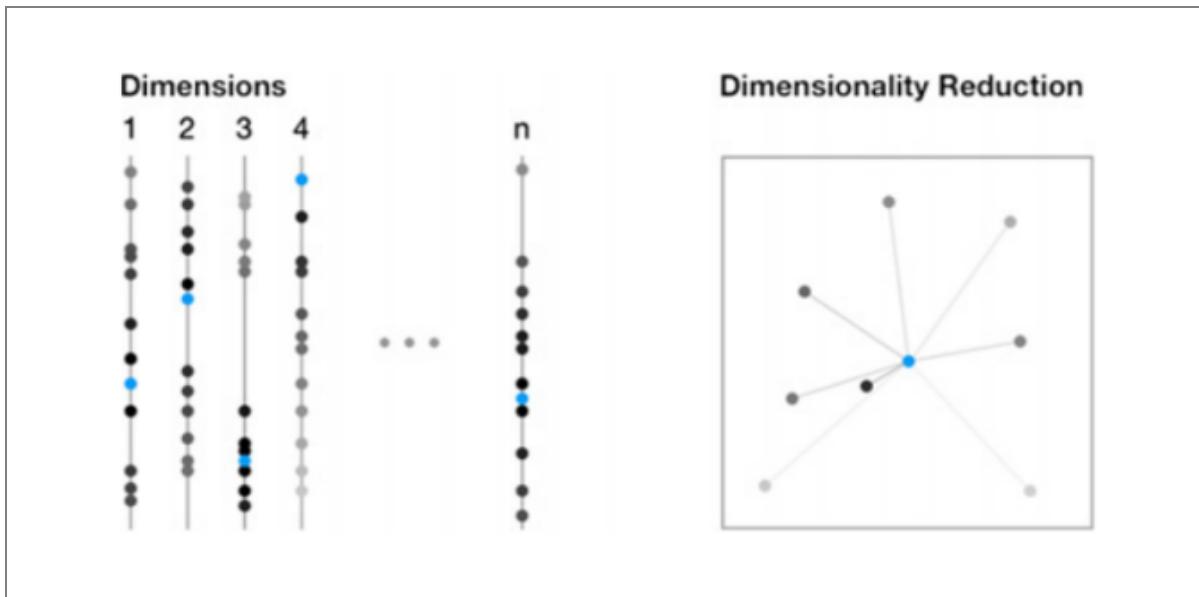
- [Segel & Heer \(2010\) Narrative Visualization: Telling Stories with Data](#)
- [Literaturübersicht im Bereich Datenvisualisierung](#)
- [Datajournalism Handbook](#)
- [Anton Zhiyanov \(2021\) DataViz Guidelines](#)

4.2 Personalisierung und Relevanz

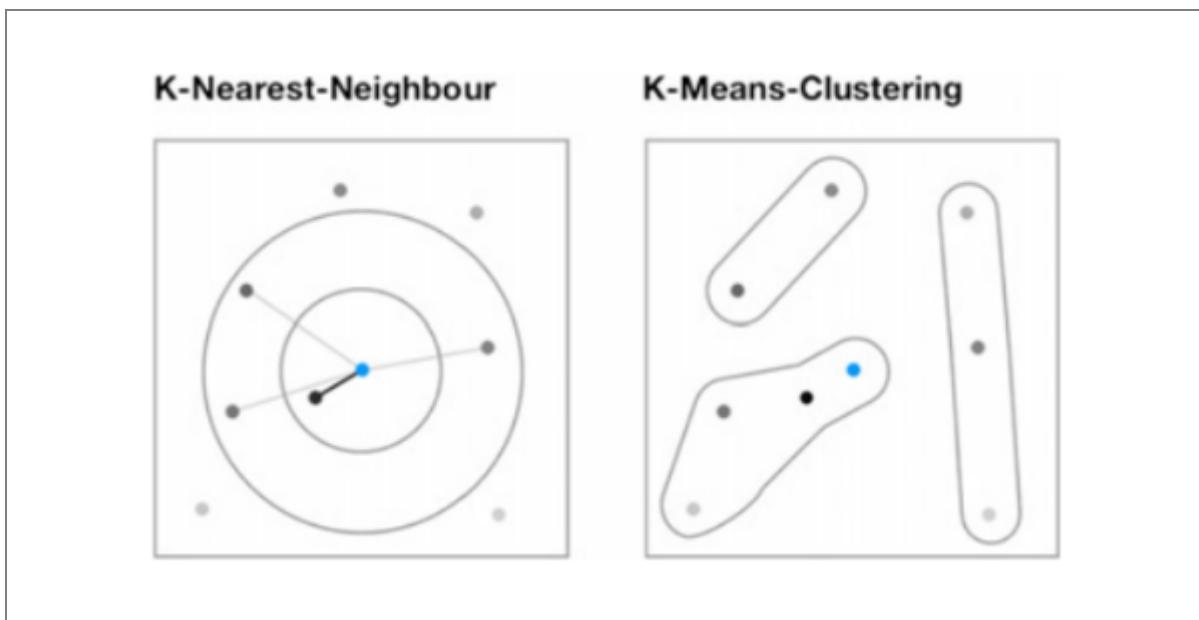
Besonders wenn es um eher abstrakte wissenschaftliche Phänomene geht, ist es wichtig den Leser*innen Bezüge zur eigenen Lebenswelt aufzuzeigen, um die Relevanz dieser Thematiken deutlich zu machen. Beim Beispiel der Klimawandelkommunikation, welche wir in unseren Prototypen thematisiert haben, wird dies besonders deutlich. Im Gegensatz zum lokalen Wetter, beschreibt der Klimawandel langfristige und großräumliche Veränderungen. Hierdurch kann es für Bürger*innen schwierig sein Bezüge zum eigenen Leben und Handeln herzustellen. Die Herausforderung ist also diese Verbindungen herzustellen, umso persönlicher die Bezüge, umso einfacher die Verbindung zur Lebenswelt der Bürger*innen.

Den folgenden Prototypen liegt folgendes Modell zu Grunde (siehe Meier & Glinka 2017): 1) es gibt einen Datenraum, welcher für die Visualisierung genutzt wird (z.B. Klimazonen in Deutschland), 2) es gibt Leser*innen die bestimmte Eigenschaften haben (Datenattribute, wie z.B. der Wohnort). Auf diesem Modell aufbauend, kann man die abstrahierte datengestützte Repräsentation einer Leser*in mit dem Datenraum abgleichen und verschiedene Operationen durchführen:

Bei mehrdimensionen Datenräumen kann eine Dimensionsreduktion genutzt werden, um möglichst ähnliche Datenobjekte im Vergleich zur Leser*in zu identifizieren, um so einen Einstieg in den größeren Datenraum zu erleichtern.

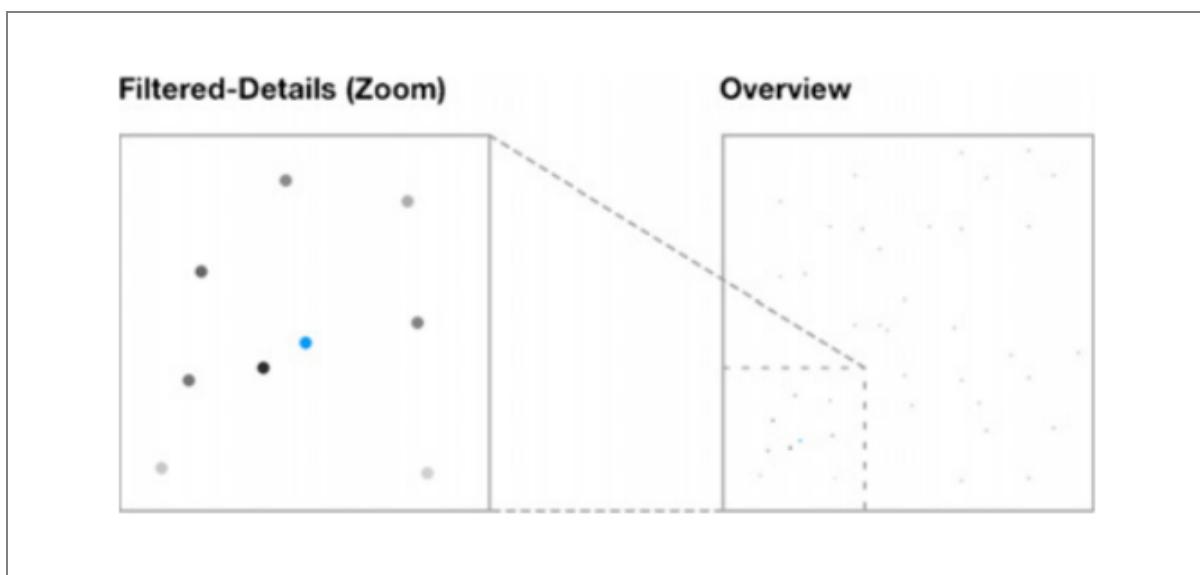
Fig. 4.2.1 Personalisierung: Dimensionsreduktion (blau Leser*in)

Ähnliche Prinzipien zumindest identifizieren ähnlicher Datenpunkte oder Gruppen ähnlicher Datenpunkte lassen sich auch mit anderen Verfahren durchführen, wie z.B. KNN oder KMC.

Fig. 4.2.2 Personalisierung: K-Nearest-Neighbour und K-Means-Clustering (blau Leser*in)

Genauso können die Attribute der Leser*in auch als Filter genutzt werden, um den Datenraum herunterzubrechen und zu verkleinern.

Fig. 4.2.3 Personalisierung: Filteransatz



In den beiden Prototypen wurde dieses abstrakte Modell anschaulich umgesetzt:

- 1. Klimawandel und Mobilität:** In dem Artikel soll das Thema CO2-Fußabdruck vermittelt werden. Die Leser*innen geben ihre Postleitzahl, ihr primäres Verkehrsmittel, sowie ihre tägliche Reisedistanz an. Dies ist das individuelle Datenobjekt. Dieses Datenobjekt wird nun genutzt, um CO2-Ziele im Rahmen der Klimawandelbekämpfung darzustellen. Statt allgemeiner Darstellungen, werden die Auswirkung von CO2-Reduktion direkt in der heimischen Postleitzahl der Leser*innen dargestellt. Weitere Details siehe unten.
- 2. Klimawandelrisiken:** In diesem Artikel sollen Klimawandelauswirkungen in Deutschland kommuniziert werden. Die Leser*innen geben hierzu ihre Postleitzahl ein. Über die Postleitzahl werden die Auswirkungen herausgefiltert, welche für diese Region Deutschlands zutreffend sind (Filter). Diese werden dann für die angegebene Postleitzahl lokal visualisiert. Weitere Details siehe unten.

Weiterführende Literatur zur Personalisierung von Visualisierungen:

- Meier, S., Glinka, K. (2018) Data-driven Personal Cartographic Perspectives – An Overview Of Applied, Artistic, And Academic Visualization Projects For Egocentric Retrospective Analysis Of Personal Spatio-Temporal Behavior Cartograms, *Kartographische Nachrichten*, (13/8).
- Meier, S., Glinka, K. (2017) The Individual in the Data, the Aspect of Personal Relevance in

Designing Casual Data Visualisations. i-com Journal of Interactive Media. Special issue: Human-Computer Interaction in Geovisualization. De Gruyter.16(3).

4.3 Umsetzung

Bei der Umsetzung wurde die in [Kapitel 2](#) beschriebene client-side classification genutzt, in Kombination mit vorberechneten statischen Datenextrakten. Hierdurch konnte eine privatsphären-konforme Implementation der folgenden Prototypen sichergestellt werden. Während der Interaktion mit den Visualisierungen werden keine Daten über die Nutzer*innen gespeichert.

Beide Prototypen nutzen das sogenannte Scrolltelling-Konzept. Die Benutzer*innen müssen zur Interaktion einfach nur scrollen. Dadurch werden Animationen und Interaktionen ausgelöst. Diese niedrigschwellige Form der Interaktion ist sehr intuitiv und funktioniert sowohl auf Desktop als auch auf mobilen Endgeräten. Die Nutzer*innen können durch das Scrolling, die Geschwindigkeit der Erzählung selber beeinflussen und steuern.

Für weitere technische Details der Umsetzung, siehe [Kapitel 5](#).

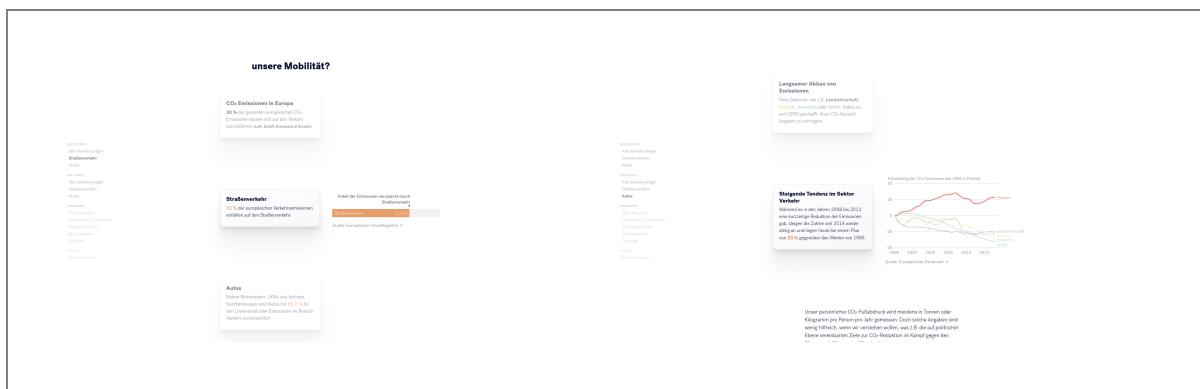
4.4 Prototyp I - Klimawandel und Mobilität

Um den Klimawandel zu verlangsamen, muss der Ausstoß schädlicher Treibhausgase reduziert werden. CO₂ ist unter den Treibhausgasen einer der wichtigsten Treiber des Klimawandels. 30% der CO₂-Emissionen in Europa werden durch den Verkehr verursacht. PKW sind für mehr als 13% der CO-Emissionen in Europa verantwortlich. Während wir als Bürger*innen nur indirekten Einfluss auf viele Formen der CO₂-Produktion haben, können wir sehr direkten Einfluss auf unsere tägliche Mobilität nehmen und damit einen Beitrag leisten.

In diesem ersten Prototypen sollen verschiedene Szenarien geplanter CO₂-Reduktion und deren Auswirkungen auf unsere Mobilität aufgezeigt werden.

Fig. 4.4.1 Prototyp Mobilität: Intro

Hierzu wird den Leser*innen zuerst aufgezeigt, welche Bedeutung dem PKW in Bezug auf CO2-Emissionen zukommt. Hierzu werden animierte Graphen genutzt, welche mit kurzen Absätzen unterstützt werden. Die Informationen werden in kleine Einheiten aufgebrochen und in Kombination aus Text und Visualisierung vermittelt.

Fig. 4.4.2 Prototyp Mobilität: Allgemeine Einführung in die Thematik

Dieser erste Abschnitt ist noch sehr allgemein gehalten. Als Einstieg in den personalisierten Bereich müssen die Leser*innen ein paar Informationen über sich preisgeben. Diese werden anschließend genutzt, um die Erzählung anzupassen.

Fig. 4.4.3 Prototyp Mobilität: Interface zur Erfassung der persönlichen Angaben

Privatsphäre geht vor: Die von dir hier angegebenen Informationen werden von uns nicht gespeichert. Diese Seite nutzt keine Cookie- oder Tracking-Technologien.

Welcher Reisetyp bist du?
Wähle ein Fortbewegungsmittel.

EMISSIONEN
Alle Verkehrsträger
Straßenverkehr
Autos

SEKTÖREN
Alle Verkehrsträger
Straßenverkehr
Autos

STÄDTE
Den Reiseweg
Deutschland Durchschnitt
Klimazonen 2050
Mobilitätsmix
Das Rad

POSTLEITZAHLEN
Autos
Autos ohne Motor

Wie weit fährst du täglich?
Wähle die Strecke, die du im Durchschnitt zurücklegst.

5 10 25 50

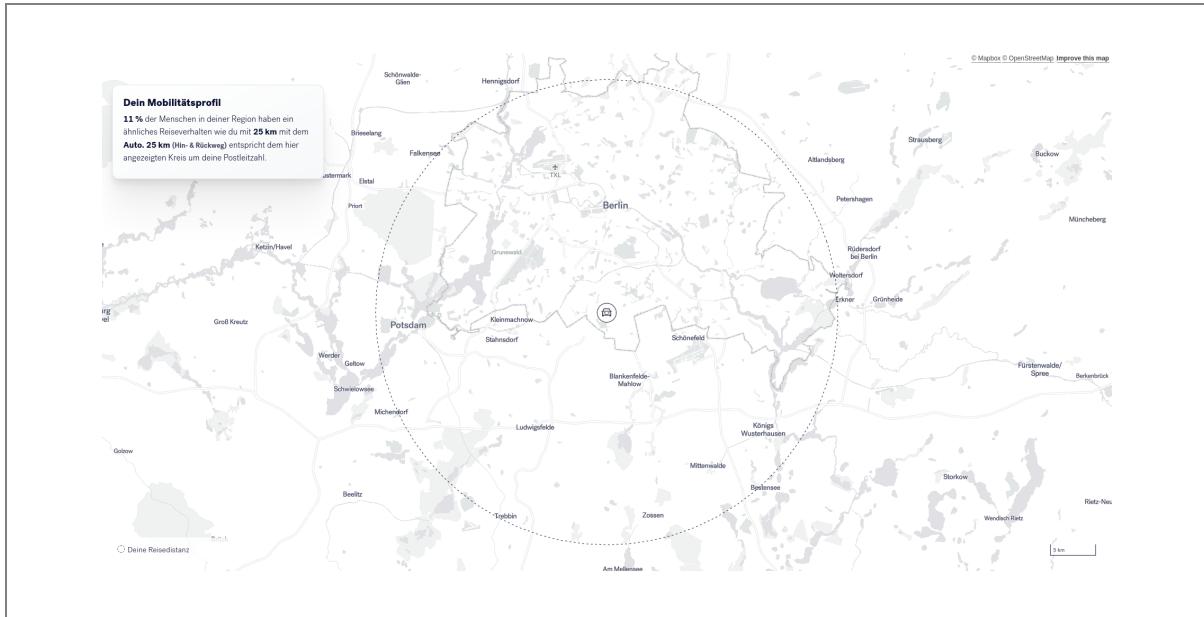
In welchem Raum bist du unterwegs?
Gib deine Postleitzahl ein und bestätige.
Oder fahre mit einer zufälligen Postleitzahl fort.

Postleitzahl
z.B. 10115 (Berlin)

Bestätigen Zufällige Postleitzahl

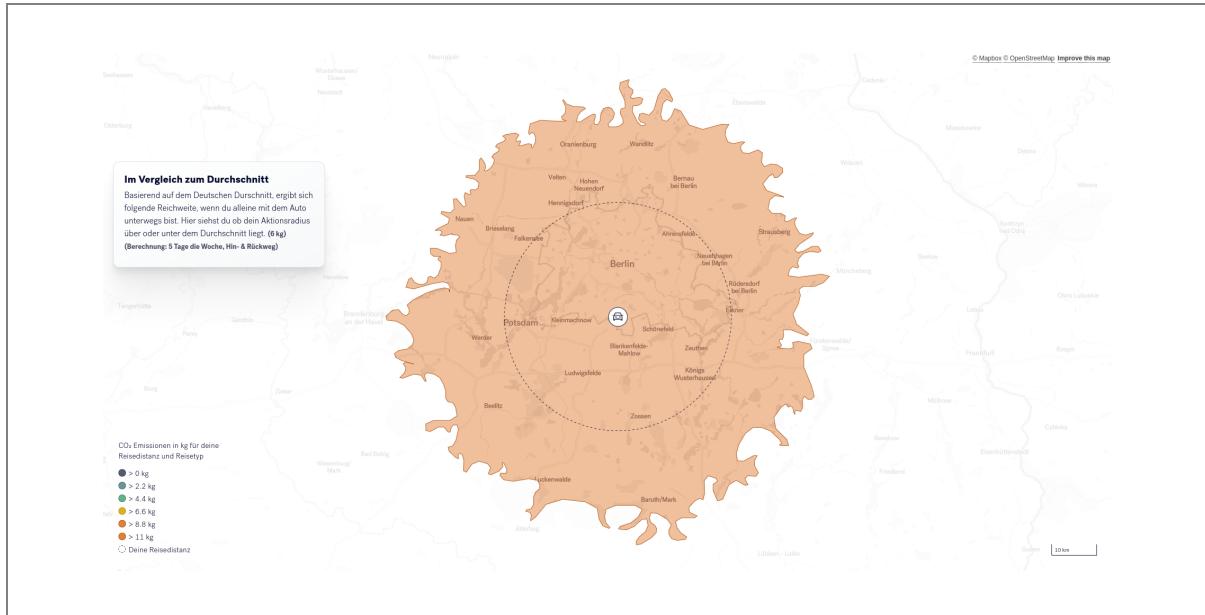
Die personalisierte Erzählung beginnt dann mit einem Wechsel zur von der Nutzer*in eingegebenen Postleitzahl. Dort wird durch einen Kreis die durchschnittliche tägliche Reisedistanz aufgezeigt, basierend auf dem gewählten Fortbewegungsmittel (basierend auf [Mobilität in Tabellen](#), DLR).

Fig. 4.4.4 Protoyp Mobilität: Darstellung der ausgewählten Region mit der durchschnittlichen Reiseentfernung anderer Bürger*innen



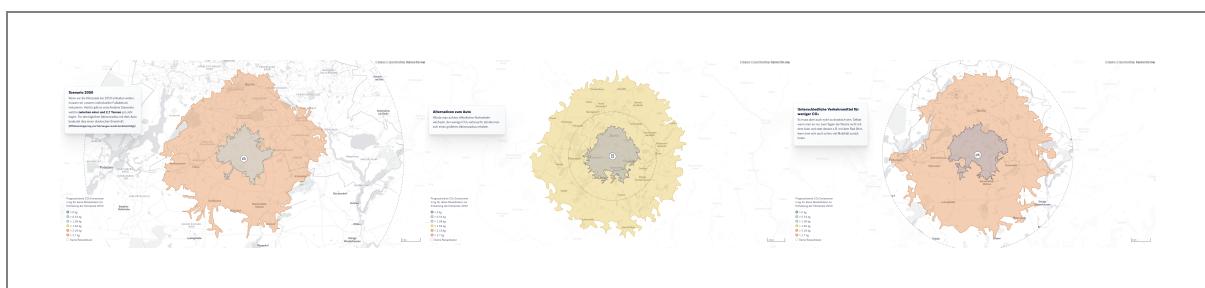
Im nächsten Schritt wird der Durchschnitt mit der Angabe der Leser*in kombiniert und erlaubt ein erstes Reflektieren der eigenen Mobilität mit dem Durchschnitt in der Region. Der Einbezug regionalisierter Durchschnitte ist durchaus relevant, da es regionale Unterschiede im Mobilitätsverhalten gibt, z.B. zwischen ländlichen und städtischen Regionen.

Fig. 4.4.5 Protoyp Mobilität: Vergleich der durchschnittlichen Reiseentfernung mit der eigenen Angabe



In den darauf folgenden Schritten wird der CO2-Fußabdruck der eigenen Mobilität in den Fokus genommen. Basierend auf Prognosen des Umweltbundesamts werden verschiedene Szenarien eröffnet, um wieviel man die eigene Mobilität einschränken müsste. Um Alternativen aufzuzeigen, werden anschließend unterschiedliche Mobilitätskombinationen aufgezeigt (z.B. ÖPNV+PKW oder ÖPNV+Rad), um so Mobilitätspotentiale aufzuzeigen. Die Art der Kombinationen und Szenarien hängt von der oben getroffenen Auswahl ab.

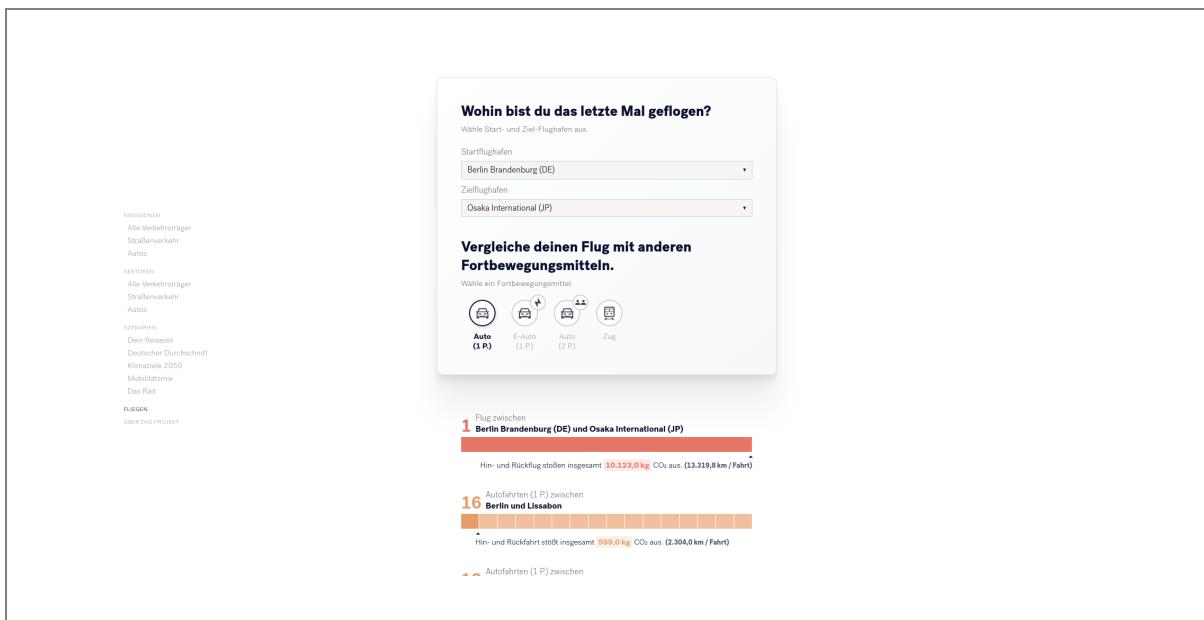
Fig. 4.4.6 Protoyp Mobilität: Darstellung von CO2-Reduktions Szenarien in Kombination mit verschiedenen Mobilitäts-Mix-Kombinationen



Da Flugreisen so viel CO2 produzieren, dass diese im ersten Teil nicht abgebildet werden konnten, haben wir ein abschließendes Modul entwickelt, welches den nächstgelegenen

Flughafen zur angegebenen Postleitzahl berechnet und dann CO2-Emissionen für Flugreisen mit anderen Mobilitätsformen für Reisen innerhalb Europa aufzeigt (z.B. für den Ort Weissach im Tal ist der nächstgelegene Flughafen Stuttgart, eine Reise von Stuttgart nach Melbourn entspricht 105 Zugreisen nach Madrid).

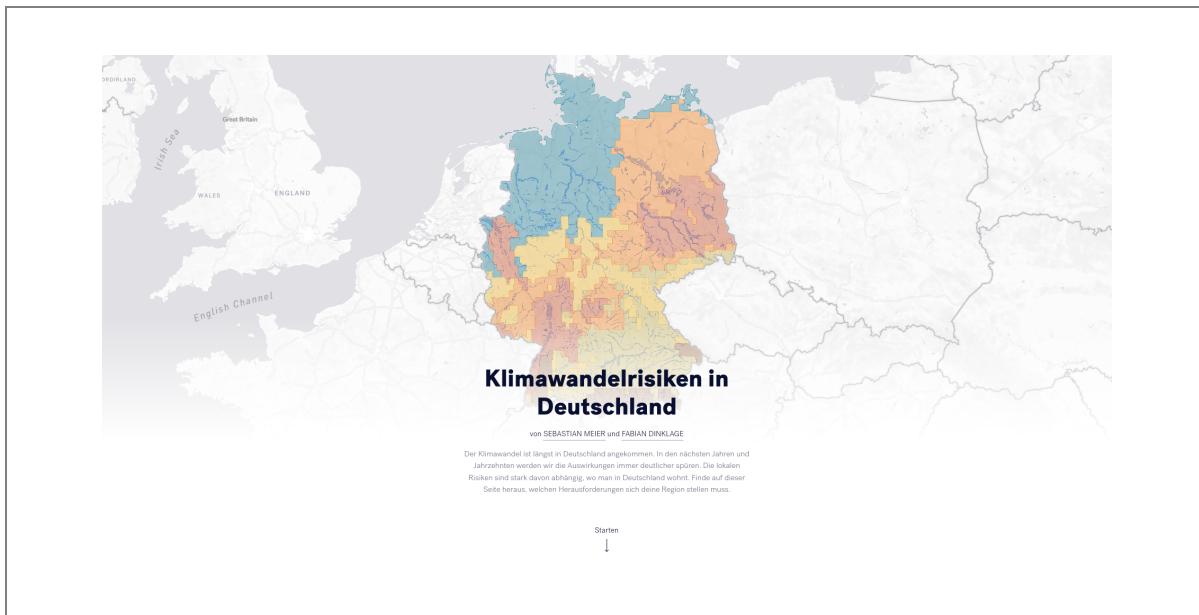
Fig. 4.4.7 Prototyp Mobilität: Visualisierung des Fußabdrucks von Flugreisen



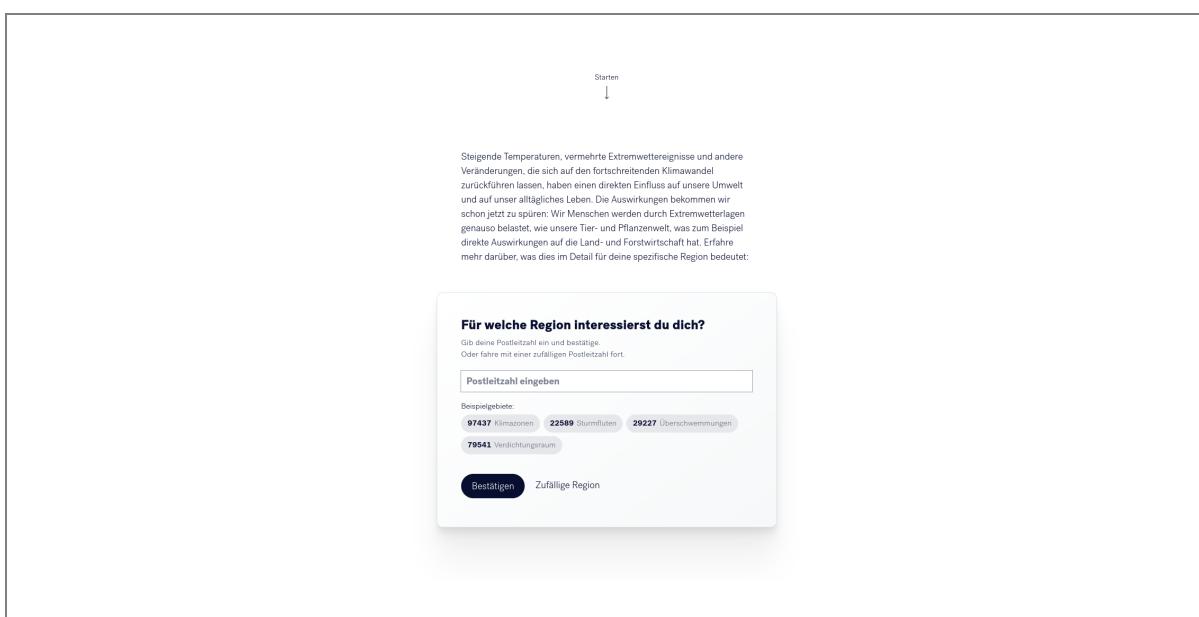
4.5 Prototyp II - Klimawandelrisiken in Deutschland

Der zweite Prototyp setzt sich mit der Thematik der Klimawandelrisiken in Deutschland auseinander. Wenn man den aktuellen Nachrichten folgt, bekommt man häufig das Gefühl vermittelt, dass der Klimawandel zwar akut ist, aber bisher primär andere Länder betrifft, wie exotische Inseln die vom Meeresanstieg bedroht sind oder Waldbrände in Kalifornien und Buschfeuer in Australien. Studien des Umweltbundesamts zeigen aber auf, dass die Folgen des Klimawandels längst in Deutschland angekommen sind und in den nächsten Jahren weiter zunehmen werden.

In dieser Storytelling-Anwendungen werden Erkenntnisse des Umweltbundesamts zu Klimawandelrisiken auf den Wohnort der Leser*in heruntergebrochen und lokal visualisiert.

Fig. 4.5.1 Prototyp Kliamwandel: Intro

Hierzu beginnt der personalisierte Teil der Anwendung, wie im ersten Prototypen, mit dem eingeben der eigenen Postleitzahl. Daraufhin werden die für diese Postleitzahl notwendigen Daten geladen und dargestellt.

Fig. 4.5.2 Prototyp Kliamwandel: Auswahl der Region

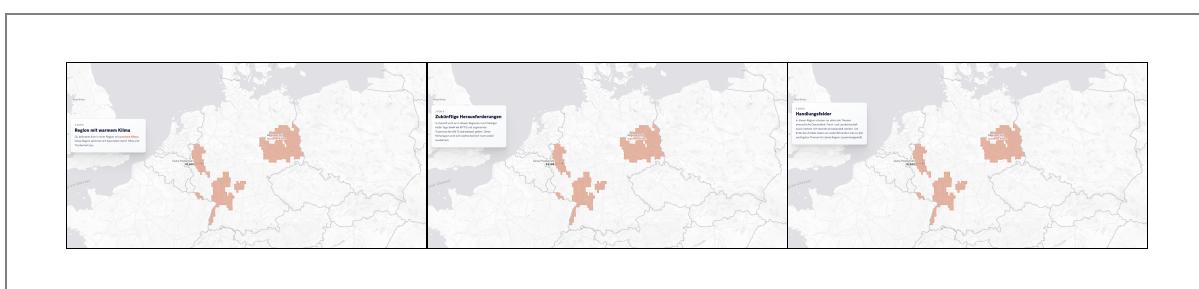
Der eigentliche Storytelling-Abschnitt beginnt mit einer Verortung der Erzählung in der angegebenen Postleitzahl auf einer Karte.

Fig. 4.5.3 Prototyp Kliamwandel: Darstellung der ausgewählten Region



Danach wird erklärt in welcher Klimazone sich dieses Gebiet befindet. In den nächsten Schritten wird die Bedeutung der Klimazone erklärt, zukünftige Entwicklungen und Bereich aus Umwelt, Gesellschaft und Wirtschaft die besonders von diesen Auswirkungen betroffen sind.

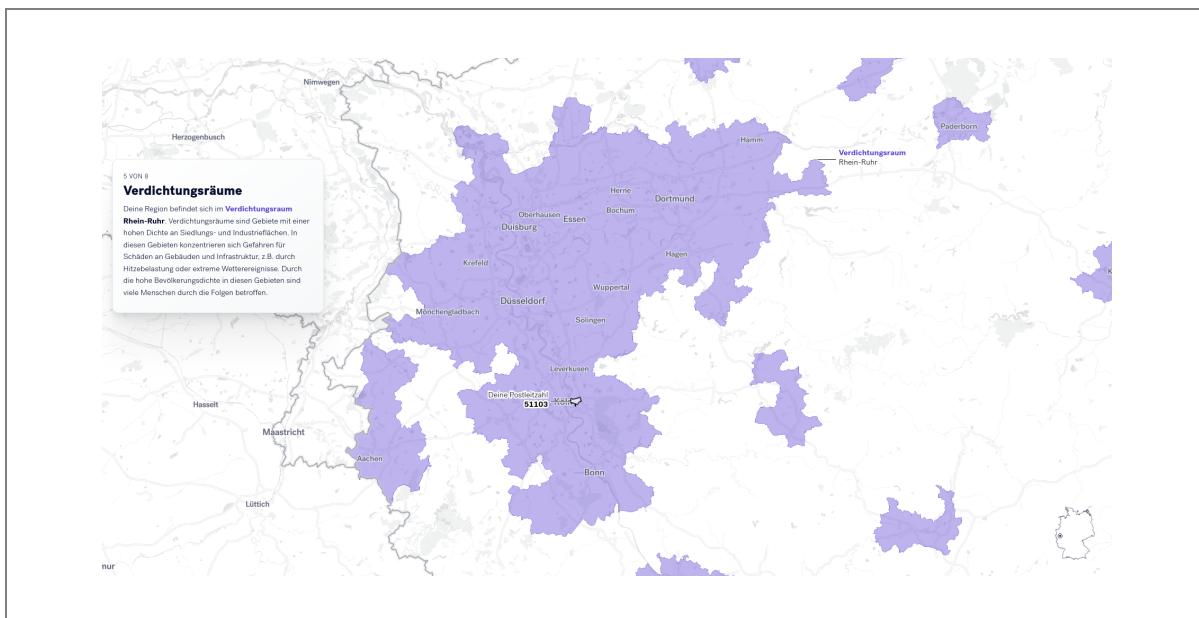
Fig. 4.5.4 Prototyp Kliamwandel: Details zur Klimaregion und zukünftigen Herausforderungen



Abhängig von der eingegebenen Postleitzahl werden weitere Risiken mit lokalem Bezug visualisiert. Diese speziellen Risiken erstrecken sich über Verdichtungsräume (siehe

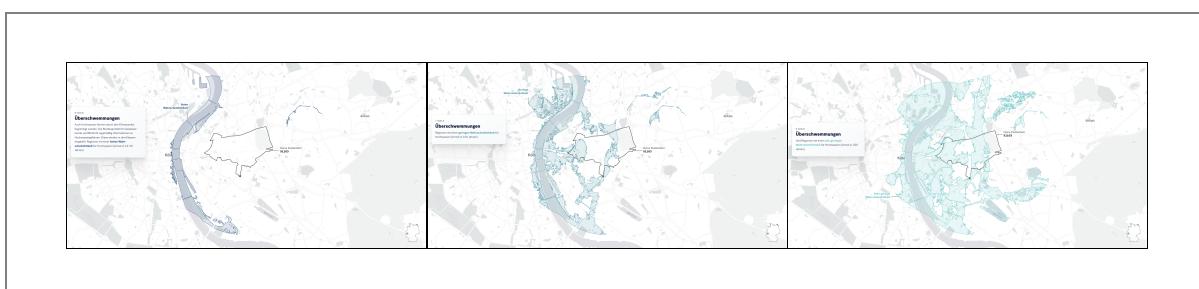
Abbildung), Gebiete an der Küste mit Sturmflutrisiken, bis hin zu Hochwassern in Flussgebieten (siehe nächster Abschnitt).

Fig. 4.5.5 Protoyp Klimawandel: Verdichtungsräume



Für die Visualisierung der lokalen Hochwassergefahren, wurden Daten der [Bundesanstalt für Gewässerkunde](#) genutzt. Hierzu werden den Leser*innen verschiedene Szenarien (Wahrscheinlichkeiten) für Hochwasser aufgezeigt.

Fig. 4.5.6 Protoyp Klimawandel: Prognosen für Hochwasser



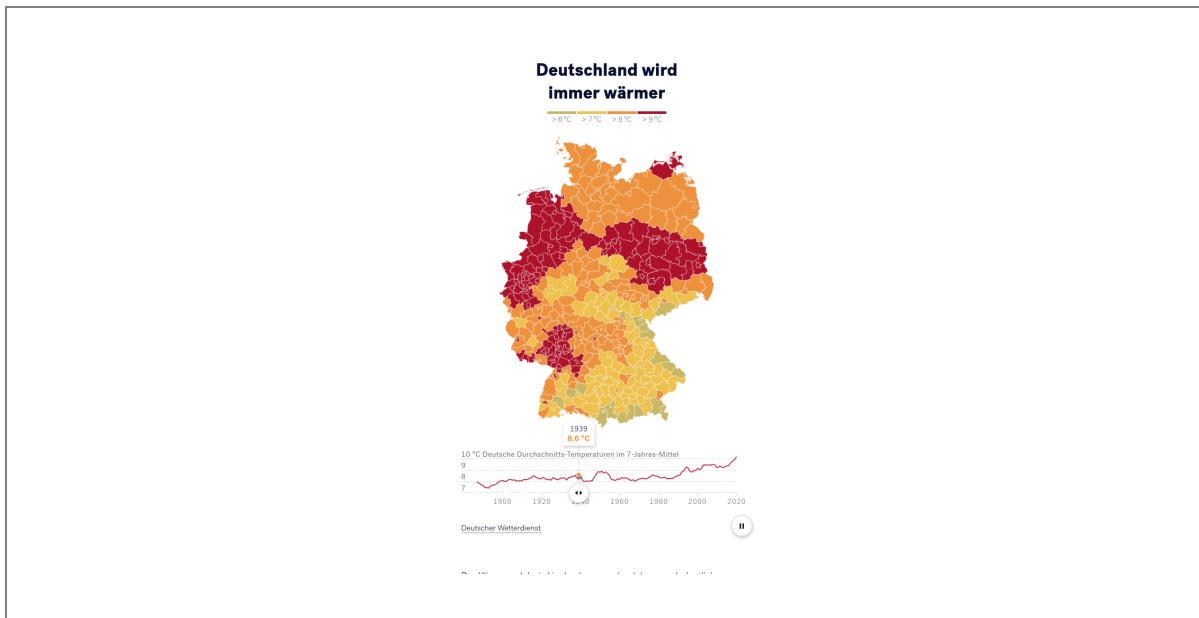
Um die zeitliche Entwicklung deutlich zu machen, werden zum Abschluss des personalisierten Bereichs die Entwicklungen der lokalen Durchschnittstemperaturen mit den Entwicklungen der Temperaturen in ganz Deutschland verglichen. Hierbei werden der deutliche Anstieg der Temperaturen unabhängig von vllt. lokal abweichenden Phänomenen hervorgehoben.

Fig. 4.5.7 Prototyp Klimawandel: Vergleich Temperaturen in Deutschland und der ausgewählten Postleitzahl



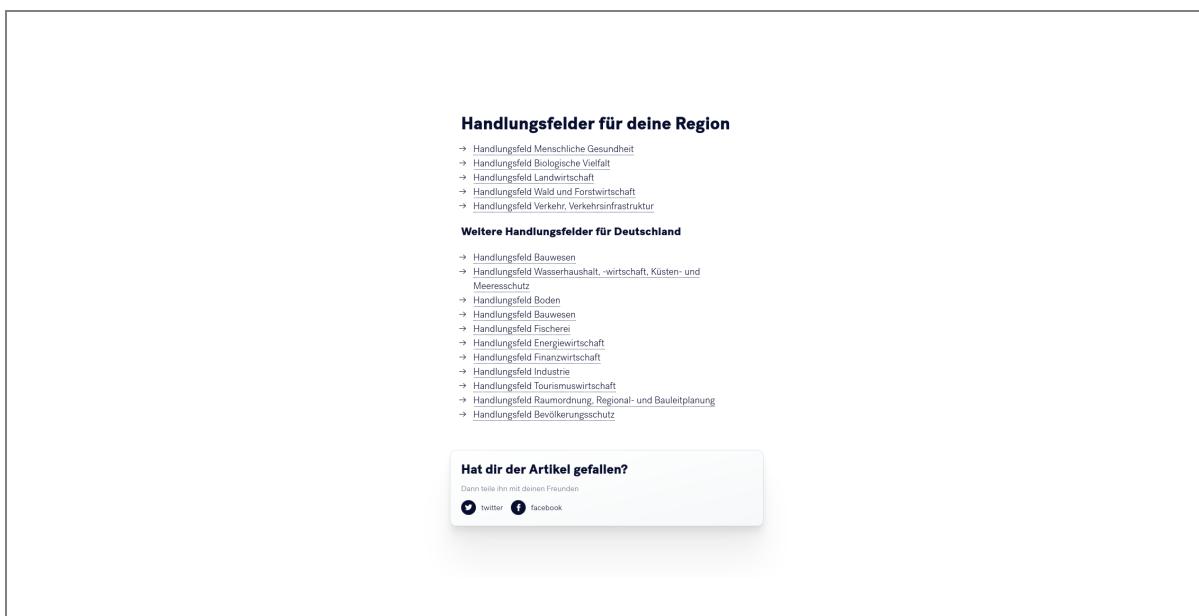
Der Vergleich zwischen den lokalen und nationalen Temperaturtrends schafft den Bezug zwischer der lokalen und nationalen Ebene. Dies schließt der letzte Abschnitt der Anwendung ab, indem auf einer Deutschlandkarte die Entwicklung der Temperaturen der letzten 100 Jahre abgebildet werden.

Fig. 4.5.8 Prototyp Klimawandel: Temperaturentwicklung für ganz Deutschland



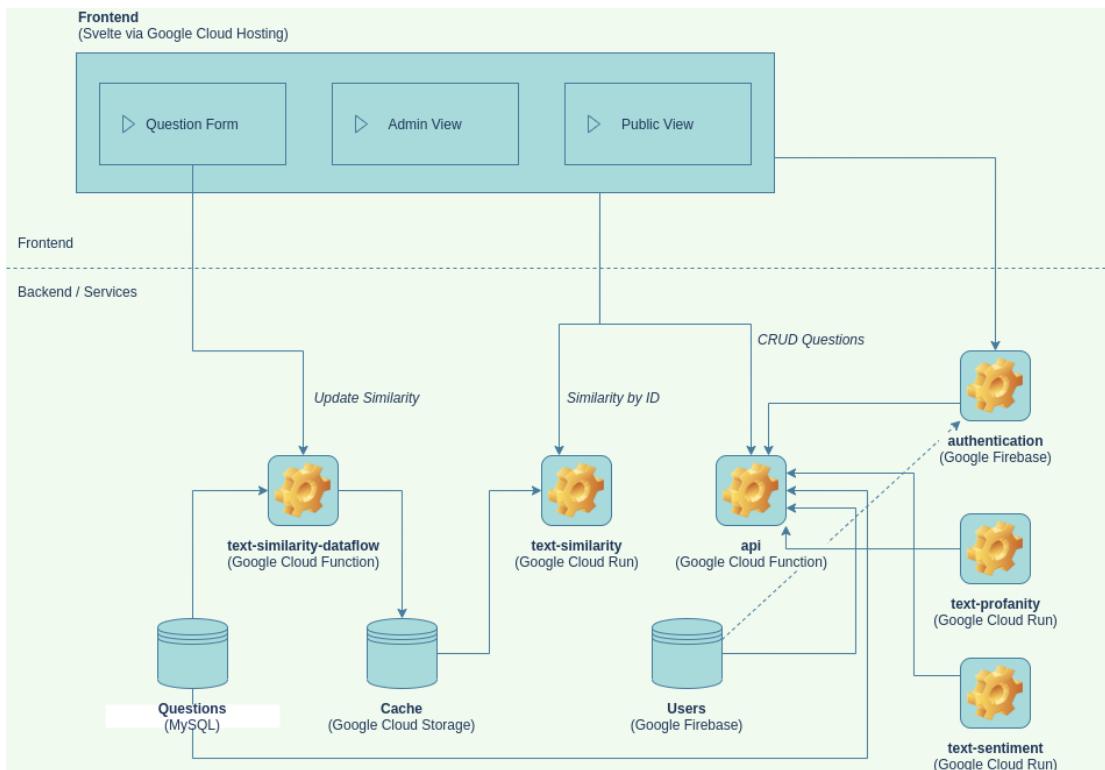
Für Leser*innen die sich weiter mit der Materie auseinandersetzen wollen, werden zum Abschluss der Anwendungen noch Leseempfehlungen gegeben. Diese Empfehlungen sind auch abhängig von den Angaben der Leser*innen personalisiert, sodass Themen für die Region der Leser*in besonders hervorgehoben werden.

Fig. 4.5.9 Prototyp Klimawandel: Personalisierte Literaturempfehlungen



5. Prototyp: Technische Dokumentation

Overview



Service Repositories

- [LoCobSS-platform](#)
- [LoCobSS-text-similarity](#)
- [LoCobSS-text-similarity-dataflow](#)
- [LoCobSS-text-profanity](#)
- [LoCobSS-text-sentiment](#)
- [LoCobSS-authentication-UI](#)
- [LoCobSS-dwd-transform](#)
- [LoCobSS-co2-data](#)
- [LoCobSS-stories...](#)