

Challenge 3: Empowering OGD-users with an AI- powered Chatbot

Goal: Optimize search journey with an AI-powered **Search-Bot**

GovTech Hackathon 2024

Portal opendata.swiss

Geschäftsstelle Open Government Data, Bundesamt für Statistik BFS

Ein grosses Dankeschön an die Hackers



Hinten v. l. n. r.: Dominik Bruggisser, Patrick Arnecke, Benoît Regamey, Philip Gretner, Patrick Kull, Mathis Kretz, Floriane Pochon Levit, Maik Roth
Vorne v. l. n. r.: Michèle Spichtig, Kim Lan Vu
Im Bild fehlt: Lena Fuhrmann

Pitch and results

Challenge Empowering OGD-users with an AI-powered Chatbot - GovTech Hackathon 2024:

<https://hack.opendata.ch/project/1094>

Github repo of the challenge:

<https://github.com/MaikRothOGD/govtech24/tree/main>

Librarian analogy

opendata.swiss is like a library for searching open data from Switzerland. The intelligent search function provides me with a "librarian". It evaluates what I'm looking for and gives me hints if I get stuck.



<https://www.flaticon.com/>

PoC: From lexical to intelligent search

<https://vimeo.com/923743750> (Video Presentation GovTech Hackathon 2024)

<https://github.com/MaikRothOGD/govtech24/tree/main/opendatasearchapp>

The screenshot displays the user interface of the 'verbesserte Suche opendata.swiss' application. At the top, a navigation bar features a circular profile icon with the letter 'D', the name 'Dominik', and the word 'Daten'. A search bar is positioned below the navigation bar, containing the text 'Datensätze suchen...'. To the right of the search bar, there are icons for a heart, a clock, and a share symbol. Below the search bar, the search results are displayed. The main heading reads '10'401 Datensätze gefunden'. A dropdown menu for 'Sortieren nach' is set to 'Relevanz'. The first search result is titled 'Wohnungen nach Kanton, Anzahl Zimmer, Bau- und Renovationsperiode, 1990 und 2000' and is attributed to 'Bundesamt für Statistik BFS'. The second result is titled 'Wohnbevölkerung am wirtschaftlichen Wohnsitz nach institutionellen Gliederungen, Geschlecht, beruflicher Tätigkeit und Staatsangehörigkeit [Kategorie], 2000' and is also attributed to 'Bundesamt für Statistik BFS'. On the left side of the search results, there are sections for 'Linked Data' and 'Kategorien'. The 'Ges' logo is visible in the bottom left corner of the screenshot. A video player interface is overlaid at the bottom, showing a play button, a progress bar at 02:39, and the Vimeo logo.

On the top! Semantic Search (tryout)

<https://github.com/MaikRothOGD/govtech24/tree/main/opendatasearchapp/tryouts>

Deploy

10397 datasets found

Show descriptions

Limit results

50

0 200

Lexical versus semantic search on opendata.swiss

Suchbegriffe oder Frage

Bevölkerung

11/2000

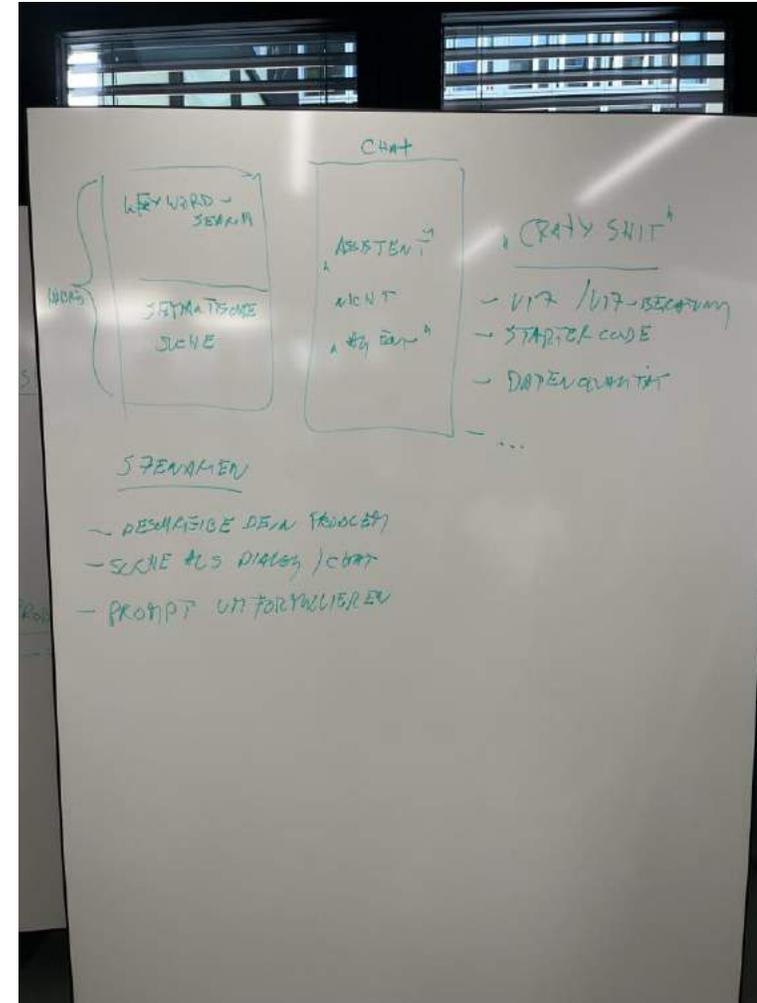
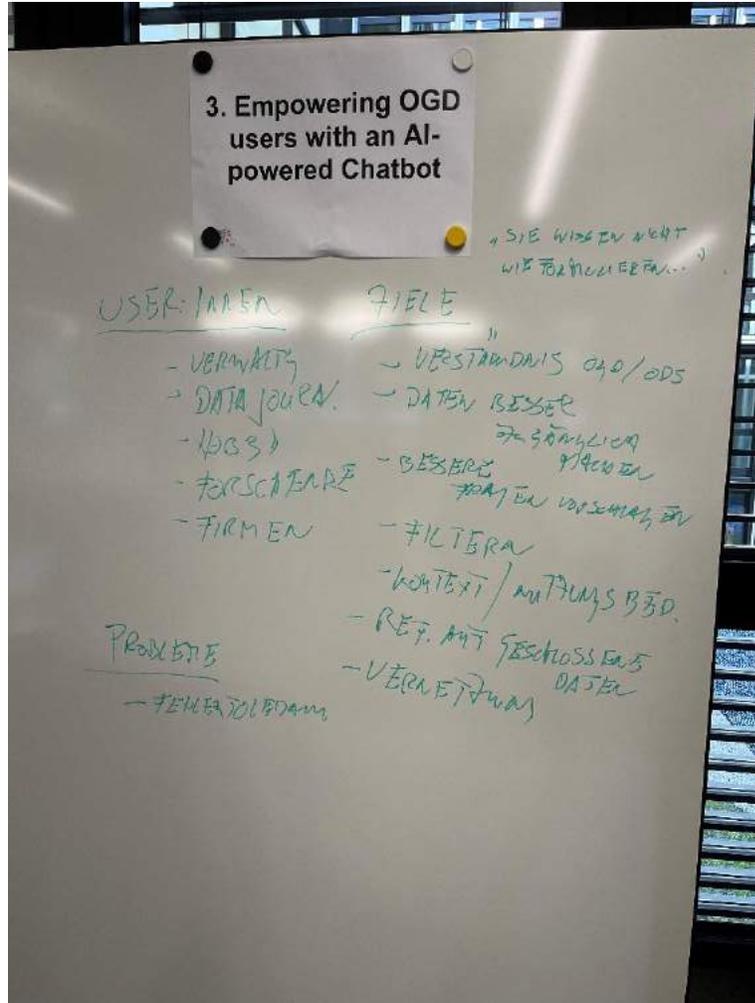
CKAN Search Results

- [Bevölkerung](#)
- [Bevölkerung](#)
- [Bevölkerung \[Pers.\]](#)
- [Bevölkerung: Männeranteil \[%\]](#)
- [Bevölkerung: Frauenanteil \[%\]](#)
- [Vornamen der baselstädtischen Bevölkerung](#)
- [Nachnamen der baselstädtischen Bevölkerung](#)
- [Bevölkerung: katholisch \[Pers.\]](#)
- [Bevölkerung: Durchschnittsalter \[Jahre\]](#)
- [Bevölkerung: Frauen \[Pers.\]](#)
- [Bevölkerung: Männer \[Pers.\]](#)

Semantic Search Results

- [Bevölkerung](#)
- [Szenarien zur Bevölkerungsentwicklung, Ergebnisse des Referenzszenarios](#)
- [Szenarien zur Bevölkerungsentwicklung nach Altersgruppen, Ergebnisse des Referenzszenarios](#)
- [Szenarien zur Bevölkerungsentwicklung, Ergebnisse des hohen Szenarios](#)
- [Szenarien zur Bevölkerungsentwicklung, Ergebnisse des tiefen Szenarios](#)
- [Bevölkerung](#)
- [Bevölkerung: Anteil 0-14-Jährige \[%\]](#)
- [Bevölkerung: Anteil 20-39-Jährige \[%\]](#)
- [Ständige Wohnbevölkerung Kanton Thurgau 2007-2014 nach Gemeinden, Geschlecht und Fünfjahresaltersklassen](#)
- [Szenarien zur Bevölkerungsentwicklung, Hypothesen zu den drei Szenarien](#)

1. Problem definition

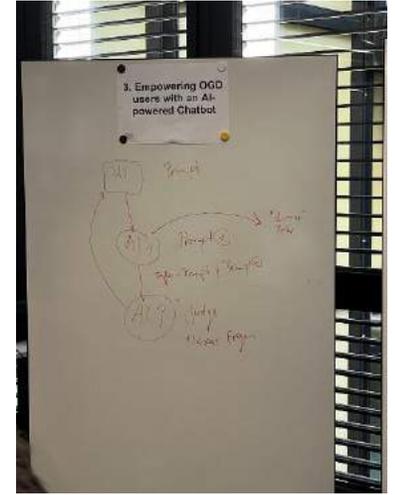
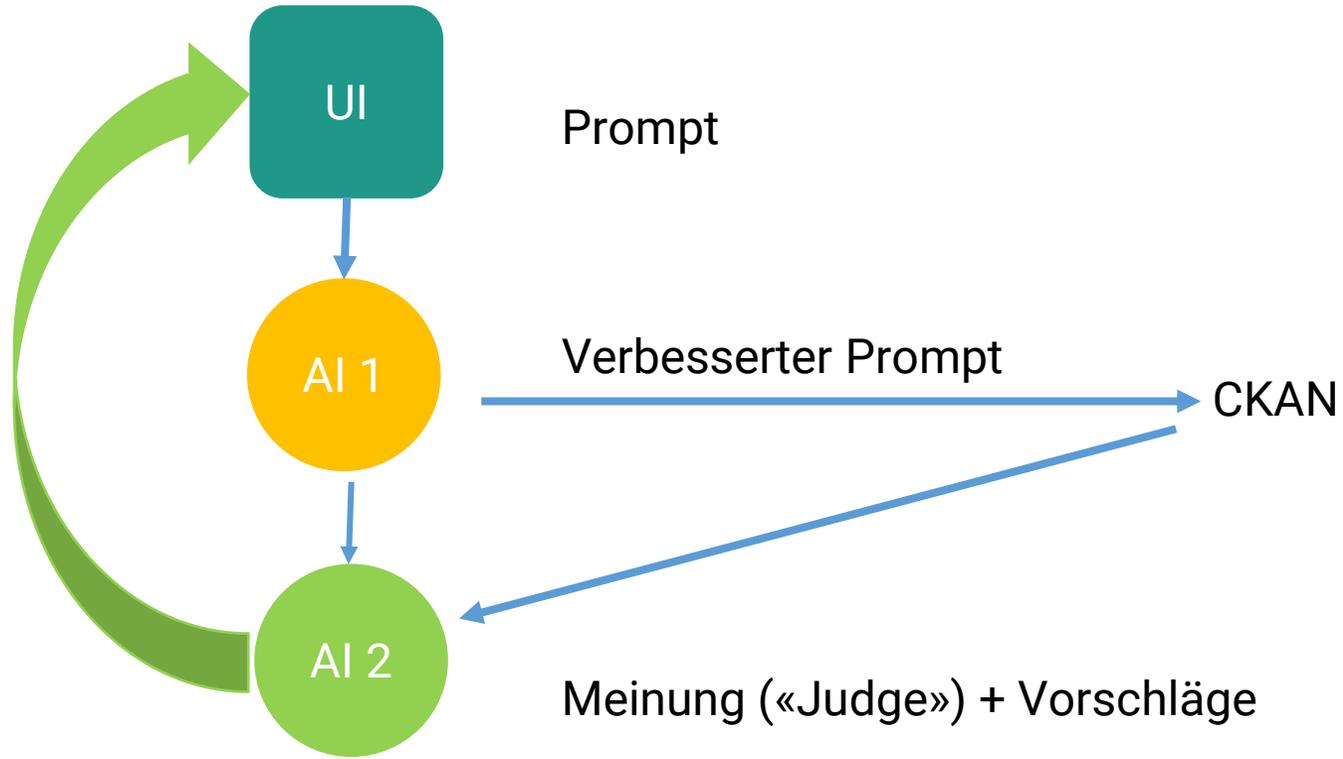


2. Focussing, identifying possible solutions

Empowering OGD-users with an AI-powered Chatbot

Nutzen	Rahmenbedingungen	Instrument	Design
Warum?	Für wen?	Was?	Wie?
<p>Was ist das Problem? Was wollen wir damit erreichen? Wer verlangt dies?</p> <p>Problem:</p> <ol style="list-style-type: none"> Gewisse User finden nicht die gesuchten Datensätze, weil sie nicht wissen, wie mit der Suche umgehen. Die User stellen mehrheitlich keine Fragen und wissen nicht wie formulieren Keine Kontextsuche <u>bsp.</u> Wetter / Meteo Es gibt keine Treffer, wenn die Wörter falsch eingegeben werden. (Fehlertoleranz, Tippfehler) Trefferliste ist zu lang (nicht klar, ob der relevanteste Eintrag ganz oben ist) <p>Ziel:</p> <ul style="list-style-type: none"> Hauptziel: User journey zu optimieren, um den gesuchten Datensatz aufzufinden (Fokus Problem 1-5) Suchanfrage (Prompt) mit LLM aufwerten, um semantisch zu suchen (Problem c.) Prompt und Suchergebnisse analysieren lassen und Empfehlungen generieren (Problem a. & b.) Spezifische Lösung vorschlagen für zu kurze und zu lange Trefferliste (Problem d. & e.) <p>Lösungsansätze:</p> <ul style="list-style-type: none"> Bessere Fragen vorschlagen Ursprüngliche Quelle soll angezeigt werden Hilfe, wie die Fragen verfeinert werden können. Ähnlich wie bei Google. Technisch: App nativ auf Webseite anbinden <p>Problemstellungen (kein Fokus)</p> <ol style="list-style-type: none"> Die Anzahl der Schritte bis zum Datensatz ist relativ lang 	<p>Wer sind die Kunden? Arbeitsweise der Kunden? Woher kommen die Daten? In welchem Format haben wir die Daten?</p> <p>User:</p> <ul style="list-style-type: none"> Verwaltung Data Journalisten Forschende Firmen Hobby <p>Typische Suchanfrage:</p> <ul style="list-style-type: none"> Stichwörter Bsp. Ich suche nach relativen Schattierungen nach EMS Was sind die typischen Suchanfragen auf Opendata.swiss <p>Datenquelle:</p> <ul style="list-style-type: none"> Opendata.swiss CKAN API API guide — CKAN 2.9.10 documentation Team AI 2: opendata.swiss API nutzen — Handbuch Opendata.swiss 1.0 Dokumentation 	<p>Welche Systeme können wir für die Problemlösung verwenden?</p> <ul style="list-style-type: none"> Vorhandene Suche (Index basierte) verwenden <p>Team AI 1: Suche verbessern (synonyme Suche erlauben)</p> <ol style="list-style-type: none"> Erhalten Suchanfrage von Webseite Werten Prompt mit <u>OpenAI</u> API auf Müssen auf CKAN API zugreifen Liefern Prompts und Trefferliste an Team AI 2 Muss sich Testfälle überlegen, anhand derer geprüft werden kann, ob die Suche besser (semantischer) ist <p>Team AI 2: Suchtreffer analysieren und Empfehlungen generieren</p> <ul style="list-style-type: none"> Prompts und Trefferliste von Team AI 1 analysieren Empfehlungen für User generieren. Mindestens für zu kurze und zu lange Trefferlist <p>Systeme</p> <ul style="list-style-type: none"> <u>OpenAI</u> <u>Streamlit</u> <u>Llangchain</u> <p>Team UX</p> <ul style="list-style-type: none"> Empfehlung für UX/UI und User Journey. Integration: Wie taucht es auf, wie Tonalität/Persona der Empfehlung Was passiert, wenn der User der Empfehlung folgt Wie messen wir (Metriken) das, dass sich die Customer Journey verbessert hat Wie können wir die Nutzung des AI-Assistenten promoten Wie könnten wir User-Feedback integrieren Können wir aus der User-Historie lernen und diese bei einem weiteren Website Besuch berücksichtigen Datenschutz (Website Cookies) 	<p>Wie soll das Design aussehen? Was soll das System können?</p> <p>User Journey</p> <ol style="list-style-type: none"> User gibt einen Suchbegriff oder Suchanfrage (Satz) ein technisch: LLM wertet Prompt semantisch auf und wir lösen eine Suche auf der API aus. technisch: Original-Prompt + verbesserter Prompt + Suchergebnisse gehen ein zweites Mal an ein LLM. Das LLM analysiert und generiert Vorschläge für die Nutzer, um seine Suche zu verbessern User bekommt Trefferliste und Empfehlung <p>Crazy shit</p> <ul style="list-style-type: none"> Code Generierung Starter code Datenqualitäts-Checker Vorschlag, mit was der Datensatz verbunden werden könnte Visualisierungsvorschlag der Daten

3. Architektur



4. Proof of Concept - Typo

<https://vimeo.com/923743750>

Current search

The screenshot shows the opendata.swiss website with a search bar containing the word "farrad". Below the search bar, it displays "0 Datensätze gefunden" (0 data sets found). The page includes navigation links for "Daten", "Organisationen", "Showcases", "Kontakt", and "Portal". There are also sections for "Linked Data" and "Kategorien" (Categories), both indicating that no results were found for those categories. A "Suche zurücksetzen" (Reset search) button is visible.

Improved search with AI assistant

The screenshot shows the same search results as the previous image, but with an AI assistant overlay. The search bar still contains "farrad". The AI assistant provides a message: "Alle genannten Begriffe stehen im Zusammenhang mit Fahrrädern und deren Nutzung. Um die Suche zu optimieren und die Ergebnisse auf eine passende Anzahl zu reduzieren, sollten Sie Ihre Suchkriterien spezifizieren oder bestimmte Aspekte wie den Fahrradtyp oder den Nutzungszweck hervorheben." Below this, it offers suggestions to expand the search: "Hier sind ein paar Vorschläge, um die Suche auszuweiten:" followed by suggestions like "Fahrrad Fahrer Rad Fahrradfahren Radfahren Mountainbike Drahtesel Velofahren Fahrradnutzung OR Fahrradkorb". It also offers suggestions to narrow the search: "Hier sind ein paar Vorschläge, um die Suche weiter einzuzengen:" followed by suggestions like "Fahrrad Fahrer Rad Fahrradfahren Radfahren Mountainbike Drahtesel Velofahren Fahrradnutzung AND Fahrradrepauratur". At the bottom, it shows "20 Datensätze gefunden" (20 data sets found) with a list of organizations including "Bundesamt für Statistik BFS 1007" and "SITG - Genfer Geoinformationssystem".

4. Proof of Concept - Multilingualism

<https://vimeo.com/923743750>

Current search

The screenshot shows the opendata.swiss search interface. The search bar contains the Arabic text 'طاقة الرياح'. Below the search bar, it indicates '0 Datensätze gefunden' (0 datasets found). A DeepL translator is integrated into the interface, showing the translation of 'Windkraft' from German to Arabic as 'طاقة الرياح'.

Improved search with AI assistant

The screenshot shows the improved search interface. The search bar contains 'طاقة الرياح'. The AI assistant provides suggestions and context, resulting in '7 Datensätze gefunden' (7 datasets found). The AI assistant also provides a summary of the search results and suggests related terms.

4. Proof of Concept – Question in natural language

<https://vimeo.com/923743750>

Current search

opendata.swiss

Daten Organisationen Showcases Kontakt Portal - Q

Startseite Daten

Daten

Wievil Chüeh gits i de Schwiz?

Suche zurücksetzen

0 Datensätze gefunden

Sortieren nach Relevanz

Linked Data
Zu dieser Suche gibt es keine Linked Data

Kategorien
Zu dieser Suche gibt es keine Kategorien

Schlagwörter

Improved search with AI assistant

Daten

Wievil Chüeh gits i de Schwiz?

Suche zurücksetzen

Linked Data
RDF XML 757
JSON-LD 743
R3 743
RDF Turtle 743
SPARQL 56

Kategorien
Regionen und Städte 3538
Umwelt 3022
Bevölkerung und Gesellschaft 2224
Regierung und öffentlicher Sektor 1648
Bildung, Kultur und Sport 1408
Wirtschaft und Finanzen 1022
Verkehr 854
Landwirtschaft, Fischerei, Forstwirtschaft und Nahrungsmittel 740
Energie 411
Gesundheit 345
Zeige mehr Kategorien

Schlagwörter
geschlecht 488
abstimmung 488
frau 428
mann 428
bgl-bundesgründungsinfrastruktur 401
bevölkerung 401
national 401
gemeinden 374
grodalen 373
kantonsweisch 347
Zeige mehr Schlagwörter

Organisationen

Diese Begriffe sind mit der Agrarwirtschaft und der statistischen Erfassung von Viehbeständen in der Schweiz verbunden. Für präzisere Ergebnisse sollten Sie Ihre Suche durch zusätzliche spezifische Begriffe einschränken.

Hier sind ein paar Vorschläge, um die Suche auszuweiten:

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung OR Ziegenzucht

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung OR Schaffhaltung

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung OR Zählung

Hier sind ein paar Vorschläge, um die Suche weiter einzuzugrenzen:

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung AND Schweinezucht

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung AND Tierhaltung

Kuh Landwirtschaft Viehbestand Schweiz Zählung Statistik Tierhaltung AND Landwirtschaft

Ich habe auch spannende alternative Suchergebnisse gefunden:

Landwirtschaftszählung

Statistik

Tierhaltung

Sortieren nach Relevanz

1881 Datensätze gefunden

opendata.swiss



Kontakt opendata@bfs.admin.ch

Michèle Spichtig (Challenge-Ownerin)
michele.spichtig@bfs.admin.ch

Kim Lan Vu (Challenge-Ownerin)
kimlan.vu@bfs.admin.ch

Portal opendata.swiss
X (Twitter) [@opendataswiss](https://twitter.com/opendataswiss)
NewsMail www.news-stat.admin.ch
LinkedIn <https://ch.linkedin.com/showcase/opendata-swiss/>
Weitere Informationen www.bfs.admin.ch/ogd