

RAG-Chatbot für Schweizer Geschichte

Illustrierte Schweizer Geschichte (Troxler, 1925)

The screenshot shows a web-based application titled "Swiss History RAG". On the left, there is a sidebar with "Information" about the Retrieval-Augmented Generation (RAG) system. It includes instructions for how it works (entering a question in German, searching historical texts, generating an answer), features (semantic search, AI-powered answers, source citations, German language support), and a "Deploy" button. The main area has a dark background with white text. At the top right is a Swiss flag icon followed by the title "Swiss History RAG". Below that is a sub-header "Ask questions about Swiss history in German". A "Ask your question" input field contains the text "Sag mir etwas interessant über Zürich?". To its right is a pink "Search" button. Underneath the input field, the text "Collection(name=swiss_history)" is visible. In the bottom right corner of the main area, there is a placeholder text "Question: Sag mir etwas interessant über Zürich?".

Alice Gregus, Amin Karbassi

ZHAW - Zürcher Hochschule für Angewandte Wissenschaften

30. Januar 2026

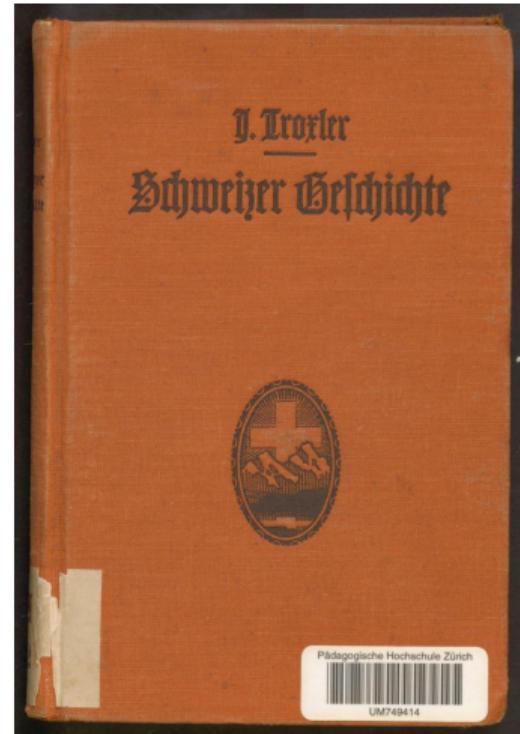
Projektübersicht

Quelle:

- **Titel:** Illustrierte Schweizer Geschichte für Sekundar- und Mittelschulen
- **Autor:** Joseph Troxler
- **Jahr:** 1925
- **Umfang:** 274 Seiten
- **Sprache:** Deutsch

Projektziel:

- Konversationelles QA-System
- Quellenangaben
- RAG-Architektur



Phase 1: Textextraktion aus PDF

Herausforderungen:

- Historisches Dokument (1925)
- Frakturschrift-ähnliche Formatierung
- Komplexes Layout mit Bildern
- Mehrspaltige Seiten
- Scan-Artefakte

Lösung:

- **PyPDF** für Textextraktion
- **Doclign** als Alternative getestet
- Seiten 12-274 verarbeitet
- Spezielle Textbereinigung:
 - Entfernung von Zeilenumbrüchen
 - Korrektur deutscher Sonderzeichen
 - Bereinigung von Scan-Artefakten

Ergebnis: Extrahierter Text in extracted_text.json

Herausforderungen bei der Textextraktion

Original aus dem PDF:

des Landes Gut und Blut einzusezen und die schuldigen Abgaben auch fernerym zu entrichten, aber die Bögte (wenn möglich, ohne Blutvergießen) zu vertreiben. Am Neujahrsmorgen 1308 sollte dieser Beschlüß ausgeführt werden.

5. Der Apfelschuh. Geßler ließ zu Altdorf auf dem öffentlichen Platze einen österreichischen Herzogshut auf einer Stange aufhängen und befahl, jeder Vorübergehende solle vor dem Hute sein Haupt entblößen. Daran wollte er prüfen, wer österreichisch gesinnt sei, und wer nicht. Wilhelm Tell, ein trefflicher Bogenschütze aus Bürgeln, ging mit seinem Knaben an dem Hute vorüber, ohne ihm die befohlene Ehrte zu erweisen. Die aufgestellte Wache führte ihn deshalb vor den Vogt, der ihm befahl, zur Strafe seinem eigenen Kinde einen Apfel vom Kopfe zu schießen. Tell bat, man möge ihm die Strafe erlassen. Geßler aber drohte, er werde ihn mit dem Kinde töten lassen. Da betete Tell zu Gott, stießt noch einen zweiten Pfeil in den Köcher, zielte und schoß den Apfel mitten durch. Das Volk jubelte; der Vogt aber fragte Tell, was jener zweite Pfeil zu bedeuten habe, und

Extrahierter Text:

Chunk ID: 105

Page: 57

Content:

. Wilhelm Tell, ein trefflicher Bogenschütze aus Bürgeln, ging mit seinem Knaben an dem Hute vorüber, oh aufgestellte Wache führte ihn deshalb vor den Vogt, der ihm befahl, zur Strafe seinem eigenen Kinde einer möge ihm die Strafe erlassen. Geßler aber drohte, erwerbe ihn mit dem Kinde töten lassen. Da betete Tell den Köcher, zielte und schoß den Apfel mitten durch. D...

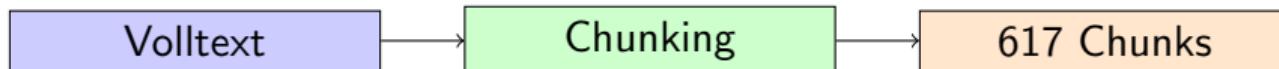
Phase 2: Text-Chunking

Chunking-Strategie:

- **Chunk-Größe:** 1000 Zeichen
- **Overlap:** 200 Zeichen
- **Methode:** RecursiveCharacterTextSplitter
- **Metadaten:** Seitennummern, Chunk-IDs

Ergebnisse:

- **Anzahl Chunks:** 617
- **Durchschn. Chunk-Länge:** 1000 Zeichen
- Erhaltung von Kontext durch Overlap
- Metadaten für Quellenangaben



Phase 3: Vektorisierung & Embeddings

Embedding-Modelle im Vergleich:

Modell	Dimensionen	Status
paraphrase-multilingual-mpnet-base-v2	768	Initial getestet
multilingual-e5-small	384	Final verwendet

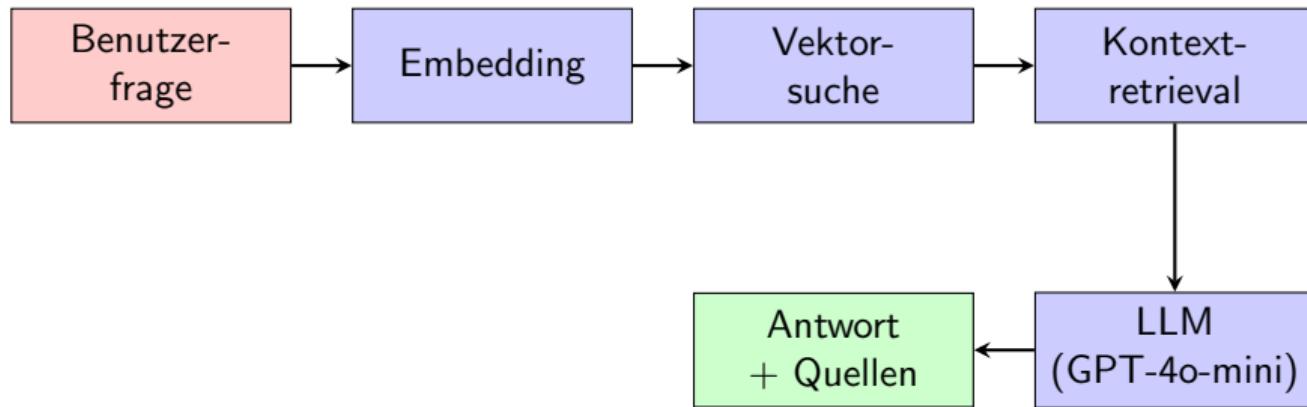
Vektordatenbank:

- **ChromaDB**
- 617 Dokumente
- Collection: `swiss_history`
- Persistierung auf Disk

Retrieval-Konfiguration:

- Top-K: 5 Chunks
- Semantic Search
- Cosinus-Ähnlichkeit

Phase 4: RAG-Pipeline



Technologie-Stack:

- **Framework:** LangChain
- **LLM:** OpenAI GPT-4o-mini (Temperature: 0.3)
- **Prompt:** Deutschsprachig, faktenbasiert
- **Interface:** Streamlit Web-App

Phase 5: Evaluation mit RAGAS

RAGAS-Metriken (LLM-as-a-Judge):

Metrik	Score	Bewertung
Faithfulness	0.8250	Excellent
Answer Relevancy	0.7505	Good
Context Precision	0.8506	Excellent
Context Recall	0.6167	Good

Testdaten:

- 10 Testfragen
- Ground-Truth Antworten
- Verschiedene Themen
- Automatisierte Bewertung

Erkenntnisse:

- Hohe Faktentreue
- Gute Kontext-Präzision
- Verbesserungspotenzial bei Recall
- System funktioniert zuverlässig

Demo:

The screenshot shows a Streamlit application interface. On the left, a sidebar contains sections for 'Information' (describing RAG), 'How it works' (a 3-step process), and 'Features' (a bulleted list). The main area has a title 'Swiss History RAG' with a Swiss flag icon, a subtitle 'Ask questions about Swiss history in German', and a 'Ask your question' input field with a placeholder 'Sag mir etwas interessant über Zürich?' and a 'Search' button. Below this is a code snippet 'Collection(name=swiss_history)'. The right side of the main area is a large empty white box labeled 'Answer'. At the bottom, there's a footer bar with navigation icons.

Information

This application uses Retrieval-Augmented Generation (RAG) to answer questions about Swiss history.

How it works:

1. Enter your question in German
2. The system searches the historical text
3. AI generates an answer based on the context

Features:

- Semantic search through Swiss history
- AI-powered answers with GPT-4o-mini
- Source citations with page numbers
- German language support

Deploy

Swiss History RAG

Ask questions about Swiss history in German

Ask your question

Sag mir etwas interessant über Zürich?

Collection(name=swiss_history)

Answer

Question: Sag mir etwas interessant über Zürich?

Live-Demo der Streamlit-Anwendung



Vielen Dank!

Fragen?

Projekt-Repository:

<https://github.com/lizinsane/llm-project-gregus-karbassi/>

Kontakt:

greguali@students.zhaw.ch
karbaami@students.zhaw.ch