

Frankfurter Allgemeine

Das KI-MVP des F.A.Z.-Archivs

Olivera Kipic – Leiterin Frankfurter Allgemeine Archiv & Rights Management

Frankfurt/Main, 14.03.2024 – Vortrag im Rahmen der 2. MVFP Future Media Now 2024



Eines der umfangreichsten und bestsortierten Pressearchive der Welt

Frankfurter Allgemeine Archiv und Rights Management – verantwortet die Informationsversorgung der Redaktionen, betreibt das digitale Archiv der Zeitung und vermarktet den Content sämtlicher Publikationen der F.A.Z.-Gruppe und anderer Mandanten.

Das Archiv als Hüter des F.A.Z.-Datenschatzes



Umfangreiches Archiv

Das Archiv der F.A.Z. ist mit ca. 8 Millionen Artikeln seit 1949 eines der umfangreichsten Pressearchive der Welt



Strukturierter Bestand

Die Artikel im Archiv sind strukturiert und mit Metadaten versehen



Vollständige Metadaten

Für jeden Artikel liegen vollständige Metadaten wie Erscheinungsdatum, Ressort, Autor etc. vor

Das gut strukturierte und mit Metadaten angereicherte Archiv der F.A.Z. ist ein einzigartiger Datenschatz für Recherchen.

Innovation & Bewahren

Das Frankfurter Allgemeine Archiv ist im Rahmen seiner Innovationsstrategie permanent auf der Suche nach neuen technischen Möglichkeiten zur Gewinnung von Beschreibungsdaten und zur Verbesserung der Rechercheperformance.

Dabei legt es großen Wert auf Datenschutz und Daten-Souveränität.

Kooperation speziell bezüglich generativer KI möglichst nur mit Partnern, die folgendes bieten:

- ✓ „On-Premise-Lösungen“ (in geschützter Umgebung)
- ✓ Unternehmenssitz in Deutschland bzw. EU
- ✓ Eigener Bestand wird nicht ungefragt als Trainingsdatenkorpus verwandt.
- ✓ F.A.Z. bleibt stets „Herrin“ über ihre eigenen Daten.
- ✓ Vermeidung von Imageschäden und Vertrauensverlust durch „unexplainable AI“ und „Halluzinationen“.

Projektziel und Projektnutzen



MVP mit 3 Kernfunktionen

Keyword-Extraction, Summarizer und Semantic Search werden in einer Demoanwendung integriert.



Interne Demoanwendung

Die Demoanwendung dient der internen Präsentation und Evaluation.

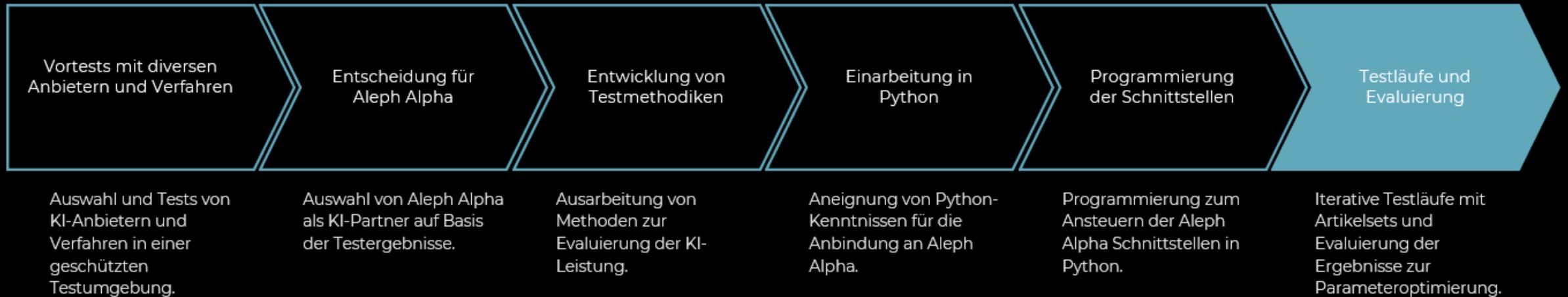


Weitere Anwendungsfälle

Die Technologie kann für Recherche, Dossiergenerierung und Suchmaschinen genutzt werden.

Das Minimum Viable Product (MVP) umfasst die wichtigsten KI-Funktionen in einer Demoanwendung für interne Präsentation und Evaluation mit Potenzial für weitere Anwendungsfälle.

Umsetzung des Projekts



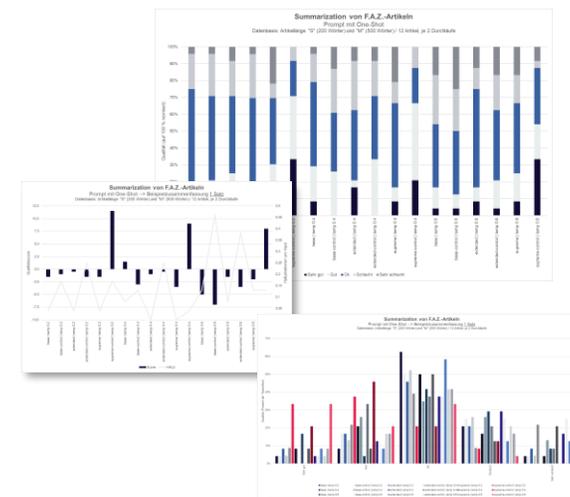
Timeline

Meilensteine im Projekt



Auswertungsbeispiel „Summarizer“

Testoutputs – intellektuelle Sichtungen ergänzt durch statistische Auswertungen

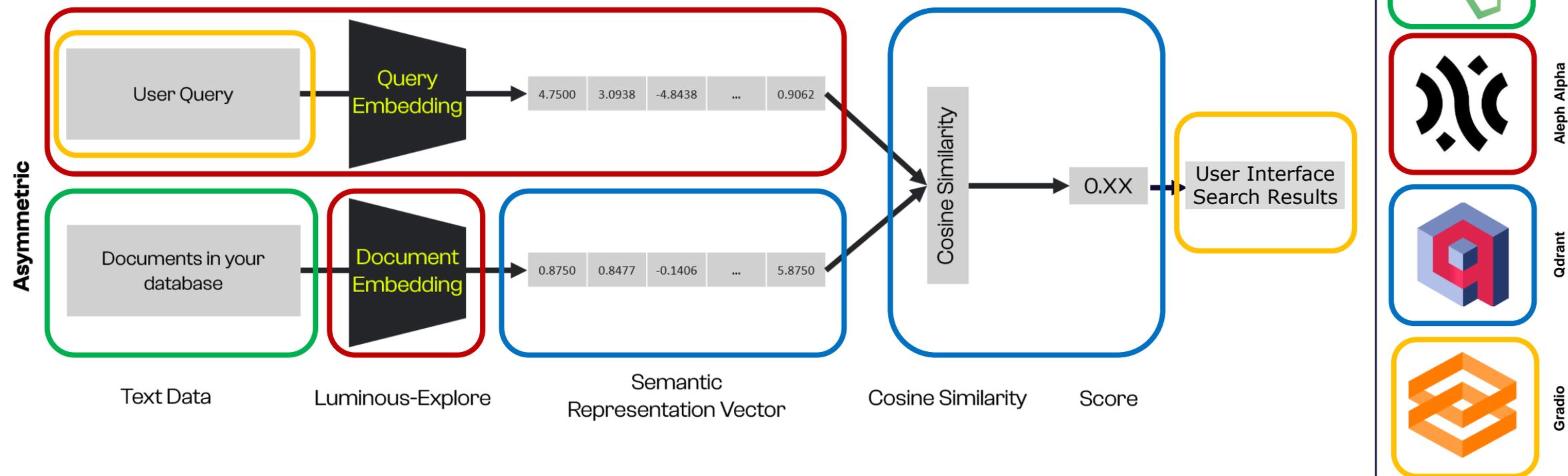


Qualitative Auswertung	Quantitative Auswertung
Die generierten Zusammenfassungen wurden von menschlichen Gutachtern auf Informationsgehalt, Kohärenz und Relevanz überprüft.	Es <u>wurden</u> <u>ausgewählte</u> , <u>repräsentative</u> Artikel in <u>über 1.000 Testläufen</u> mit dem Summarizer verarbeitet. Die durchschnittliche Komprimierungsrate betrug 23%.
Insgesamt wurden die Ergebnisse als gut bewertet, allerdings traten gelegentlich Falschzuordnungen und unsinnige Aussagen auf.	Bei 32% der Summaries mussten kleinere Korrekturen vorgenommen werden, um die Qualität zu verbessern.

Semantische Suche

Technisches Setup MVP

- Vektorisierung: Aleph Alpha Sprachmodell luminous-base
- Vektordatenbank/-suchmaschine: Qdrant GmbH, Berlin
- User Interface: Python-Framework Gradio



Ergebnisse



Demo-Anwendung fertiggestellt

Eine voll funktionsfähige Demo-Anwendung mit Keyword Extraction, Summarizer und Semantic Search wurde entwickelt



Implementierung in Produktivumgebungen möglich

Die Demo-Anwendung kann in Produktivumgebungen des Verlags implementiert werden



Weitere Entwicklung geplant

Weitere Entwicklung für KI-Assistenten, Wissensdatenbanken etc. ist geplant

Die Demo-Anwendung ist ein wichtiger Meilenstein für die weitere KI-Entwicklung beim F.A.Z.-Archiv

Ausblick

Archiv-Chatbot



KI-Assistenten für Redaktion

Wissensdatenbank

Optimierung bestehender
Anwendungen

Freiheit beginnt im Kopf.

— Vielen Dank für Ihre Aufmerksamkeit!



Olivera Kipic

Leitung Archiv und Rights Management

o.kipic@faz.de

+49 69 7591-2986

Live-Demo:

Keywords

Zusammenfassungen

Semantische Suche

