

DEMONSTRATIONSBEISPIEL

KI-basiertes innerbetriebliches Wissensmanagement für KMU

Problemstellung

Aktives und intelligentes Wissensmanagement kann für KMU viele Vorteile bieten, ist jedoch mit zahlreichen Herausforderungen verbunden. Ähnlich wie eine adäquate Dokumentation als Teil der Geschäftsprozesse kann auch das Wissensmanagement zu konkreten **Wettbewerbsvorteilen** führen, u.a. durch:

- Sicherung von kritischem Expertenwissen über Mitarbeitergenerationen hinweg und geeignete Zugriffsmöglichkeiten darauf
- Kostenreduktion und Vermeidung von „Reibungsverlusten“ bei Routinetätigkeiten (u.a. Nachfragen, Zuständigkeiten, Fehlervermeidung)
- Unterstützung bei der Einarbeitung und Qualifizierung von Mitarbeitenden

Lösung

Mit der Verfügbarkeit neuer **KI-Technologien** zur Repräsentation von Wissen und einfachen Zugriffsmöglichkeiten darauf haben sich in den letzten Jahren auch für KMU neue und leistungsfähige Optionen für das innerbetriebliche Wissensmanagement ergeben. Technologische Grundlage sind sogenannte **große Sprachmodelle (Large Language Models – LLM)**.



Webbasierte Nutzerschnittstelle des Demonstrators

Bei der **Retrieval Augmented Generation (RAG)** Technologie liegt das Wissen in einer speziellen, sogenannten Vektordatenbank vor, welche eine **semantische Suche** ermöglicht, bei der die Bedeutung der Suchanfrage und nicht einzelne Suchbegriffe im Mittelpunkt stehen. Vektordatenbanken lassen sich automatisiert aus einer in Form digitaler Dokumente vorliegenden Wissensbasis, wie z.B. das Intranet einer Firma, aufbauen. Für den möglichst einfachen Zugriff auf das in den Vektordatenbanken enthaltene Wissen kommen bei RAG-Systemen LLM zum Einsatz: Fragen können über einen **Chatbot** gestellt werden, welcher die Antworten als natürlichsprachigen Text zurückliefert. Zusätzlich kann ein RAG-System auch **direkte Verweise** bzw. anklickbare Links auf die Passagen in Originaldokumenten liefern, welche die Antworten auf die Anfrage des Nutzers enthalten.

Gefördert durch:

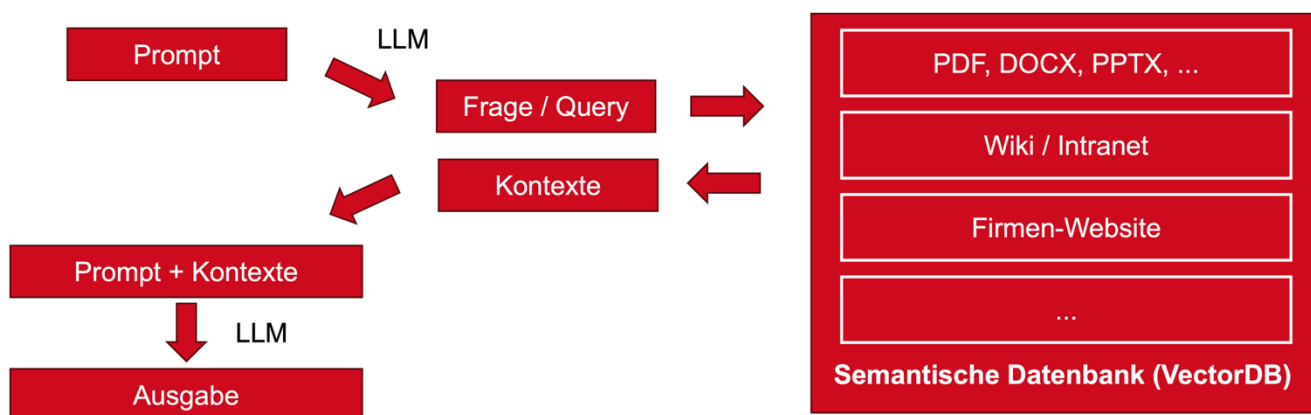


aufgrund eines Beschlusses
des Deutschen Bundestages

RAG-Systeme gibt es in zahlreichen Varianten, die sich u.a. bzgl. Kosten, Funktionalität, Verarbeitungsgeschwindigkeit, Schutz vertraulicher Daten und digitaler Souveränität unterscheiden. Die Systeme lassen sich sowohl cloudbasiert als auch auf lokalen Servern realisieren, was bei der **Verarbeitung von vertraulichen und personenbezogenen Daten** ein wichtiges Kriterium sein kann. Unternehmen, welche sich für RAG-Systeme und darauf basierendes Wissensmanagement interessieren, können diese Technologie mit dem hier vorgestellten Demonstrator interaktiv erleben. Die Unterschiede zwischen herkömmlichen, ausschließlich auf LLM basierenden Chatbots und den hier beschriebenen RAG-Systemen werden mit dem Demonstrator ebenfalls veranschaulicht.

Vorteile

- RAG-Systeme nutzen für die Generierung der Antworten ausschließlich die **firmenspezifische Wissensbasis** und können konkrete Verweise auf relevante interne Dokumente liefern.
- Die Wissensbasis in Form einer Vektordatenbank lässt sich mit vergleichsweise geringem Aufwand aufbauen und pflegen.
- Das zum Einsatz kommende LLM dient lediglich als Schnittstelle für den Datenabruf und die Antwortgenerierung in natürlicher Sprache, weshalb auch kleinere frei verfügbare LLM zum Einsatz kommen können. Dies ermöglicht auch KMU ein sogenanntes **lokales RAG-System**, welches vollständig im Firmenintranet laufen kann.



Funktionsweise der Retrieval Augmented Generation (RAG) Technologie

Impressum

Mittelstand-Digital Zentrum Ilmenau
Gustav-Kirchhoff-Platz 2
98693 Ilmenau

www.zentrum-ilmenau.digital

Ansprechpartner

Modellfabrik Smarte Sensorsysteme

Dr. Frank Spiller

03677 8749 361

spiller@kompetenzzentrum-ilmenau.de