

AIDA

AI ASISTENT
PRO
VYHLEDÁVÁNÍ
DOKUMENTŮ





Začátek – jazykové modely

Technologický boom. Strojové „porozumění“ významu přirozeného jazyka na konverzační úrovni a jeho využití při komunikaci v přirozeném jazyce.

Základní myšlenka využití

Dokument -> „něco“ ho přečte místo mě a pak...

...najdi, porovnej, ověř, sepiš.



AIDA – Asistent vyhledávání

Nástroj pro vyhledávání dokumentů uložených v digitální podobě ve firemním dokumentovém archivu.

Za pomoci technologie jazykových modelů vyhledá relevantní dokumenty na základě dotazu v přirozeném jazyce a opatří je mírou relevance jejich obsahu k položenému dotazu.

Vyhledává na základě významu zpracovávaného obsahu a může poskytnout i sumarizaci vyhledaných informací.

AI Document Advisor

OČEKÁVANÉ BENEFITY

Rozšiřte možnosti vyhledávání dokumentů z archivu o zadávání vyhledávacích kritérií v přirozeném jazyce a vyhledávání podle podobnosti významu a nikoliv podle podobnosti zápisu dané informace.

Vyhledaný obsah sumarizujte v souladu se zadaným dotazem rovněž v přirozeném jazyce.



Významné zrychlení přístupu k relevantním informacím

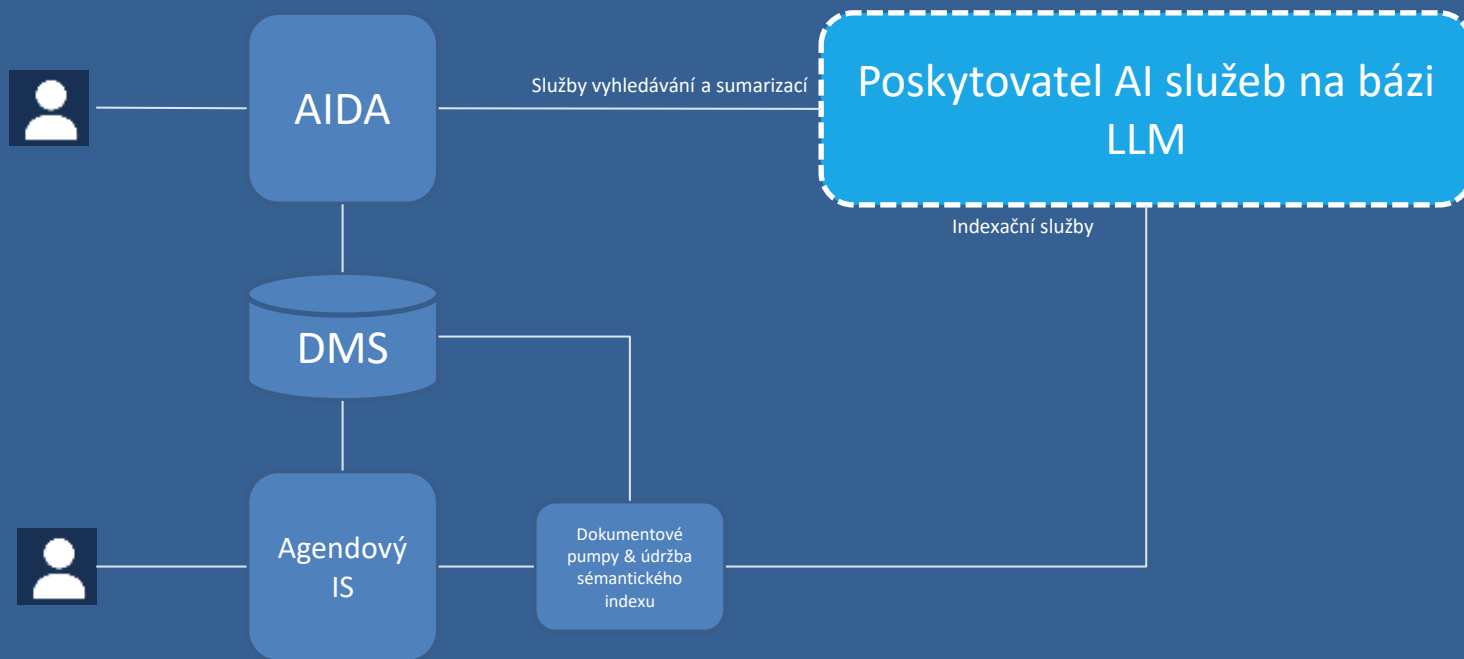


Zjednodušení vytěžování informací z dokumentů s neznámým obsahem

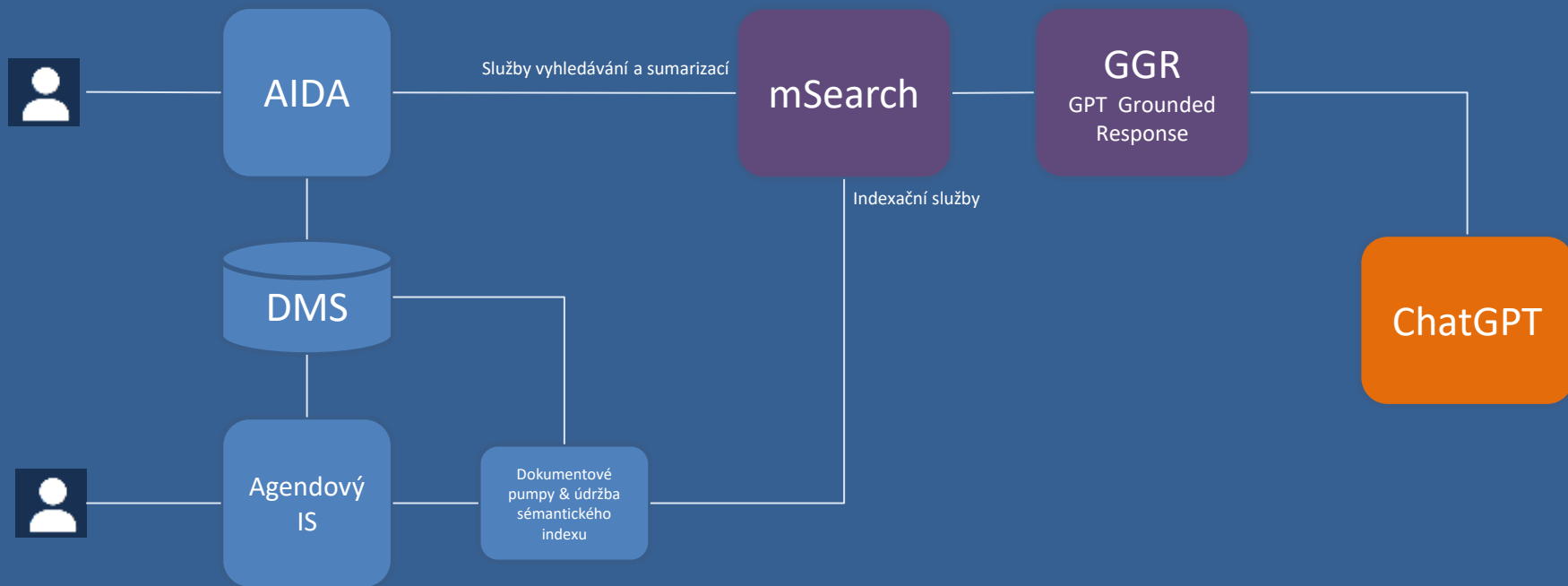


Uživatelé věnují více času zpracování informací než jejich vyhledávání

AIDA - Architektura



AIDA - Architektura



AIDA - Asistent vyhledávání



1.

OCR
příprava
dokumentů
určených k
vyhledávání

2.

Analýza obsahu
dokumentů
pomocí
jazykového
modelu

3.

Klientská
aplikace pro
sémantické
vyhledávání

AIDA – Klientská aplikace



3.

Zadávání a
správa
dotazů,
zobrazení
výsledků
vyhledávání

Zajišťuje zadávání a zpracování dotazů v přirozeném jazyce jejich výsledkem je seznam relevantních dokumentů s možností textové sumarizace nalezených výsledků.

Při vyhledávání je možné kombinovat přirozený jazyk s metadatovými filtry.

Součástí řešení je vedení historie dotazů pro jednotlivé uživatele s možností vytváření katalogu oblíbených dotazů nebo jejich sdílení s dalšími uživateli.

AIDA – Klientská aplikace

The screenshot displays the AIDA client application interface. At the top, a sidebar on the left contains navigation links. The main header area features a 'Welcome Martin!' message and a search bar with the text 'Existují dokumenty týkající se problematiky sémantického vyhledávání?'. Below the search bar, a paragraph of text provides context about the search results. The bottom section of the interface shows a table with search results, including columns for document name, category, comment, agenda, and system. A sidebar on the right lists various filters and categories.

SEARCH WITH ANSWER

Existují dokumenty týkající se problematiky sémantického vyhledávání?

SEARCH

Ano, existují dokumenty týkající se problematiky mSearch. Jedná se o dokument s názvem "Specifikace AIDA.docx" z jednání, kde byla prezentována problematika indexace dokumentů pro vyhledávání v rámci agendy DocumentSearch. Tento dokument obsahuje informace o metadatech, které jsou indexovány pro účely sémantického vyhledávání, a popisuje postup indexace obsahu pro agendu DocumentSearch.

Records Count: 2	Search	Semantic Score	Document Name	Document Category	Comment	Agenda	System
☐	0.887	Specifikace AIDA.docx	Technická specifikace	ADA vyhledávání	IT	DMS	
☐	0.652	mSearch.pdf	Integrovaní API	IT	DMS		

Filters: Semantic Score, Document Name, Document Category, Comment, Agenda, System

AIDA – Analýza obsahu



2.

Analýza
obsahu
dokumentů
pomocí
jazykového
modelu

Dokumenty jsou analyzovány pomocí jazykového modelu. Výsledkem analýzy je vyhledávací index vhodný pro tzv. sémantické vyhledávání.

AIDA řešení zajišťuje nejen vytvoření vyhledávacího indexu pro již existující archiv dokumentů, ale i údržbu indexu s ohledem na nově vytvářené dokumenty nebo jejich změny.

AIDA využívá pro indexaci a následné vyhledávání technologii mSearch společnosti THE MAMA AI.

AIDA – Příprava dokumentů



1.

OCR

příprava
dokumentů
určených k
vyhledávání

Dokumenty určené pro AI vyhledávání mohou být v různých formátech, ale musí obsahovat textovou informaci, která je předmětem zpracování jazykovým modelem.

Pokud archiv dokumentů obsahuje skenované dokumenty pouze s obrazovou informací, je potřeby tyto dokumenty opatřit textovou vrstvou. Například OCR technologií.



AIDA – Postřehy a zkušenosti

- Cca 400 000 dokumentů
 - LLM zpracování 2 000 za hodinu v dávkovém režimu
 - Celková doba cca 3 týdny
- PDF, MS Word, TXT
- Ladění parametrů AI služeb podle obsahu, struktury a formátu dokumentů
- Záludnosti testování
- Záludnosti verzování
- Udržení integrity mezi DMS a AI indexem



Děkuji za pozornost

Martin.Seget@i.cz

www.iczgroup.com



mSearch

Klíčové vlastnosti

- Umožňuje hybridní vyhledávání (sémantika v kombinaci s fulltextem)
- Metadatové filtrování v kombinaci se semantickým vyhledáváním
- Čeština (např. rekonstrukce diakritiky)
- Vysvětlitelnost (označení relevantních pasáží, confidence score)
- Absolutní kontrola nad granularitou indexace přes REST API
- User access, role management

Provozní modely

- SaaS, on-premise deployment (český tým zajišťující podporu)
- Možnost offline řešení v kombinaci s opensource LLM



mSearch HW on-premise

cca 15 000 dokumentů a 10 kolekcí (rozsah do 50000 segmentů)

On Kubernetes (preferred)

- 2 separate namespaces in k8s cluster (for separate development and production environments)
- production env.: 16 CPU-cores with 32GB RAM, 1TB disk space
 - 3 pods with up to 16GB RAM allocation (for mSearch)
 - 2 pods with up to 8GB RAM allocation (for GGR)
 - 3 pods with up to 8GB RAM allocation (for Elastic)
- development env.: 8 CPU-cores with 16GB RAM, 1TB disk space
 - 3 pods with up to 8GB RAM allocation (for mSearch)
 - 2 pods with up to 4GB RAM allocation (for GGR)
 - 3 pods with up to 4GB RAM allocation (for Elastic)

On VMs

- production env.: 16 CPU-cores with 32GB RAM, 1TB disk space
- development env.: 8 CPU-cores with 16GB RAM, 1TB disk space
- 2x index a uživatel v vaší Elastic Search instalaci, pokud máte, popř. navíc 2xVM s 8 CPU-cores with 16GB RAM, 1TB disk space for Elastic