

Komunikační agent pro informace o Brně

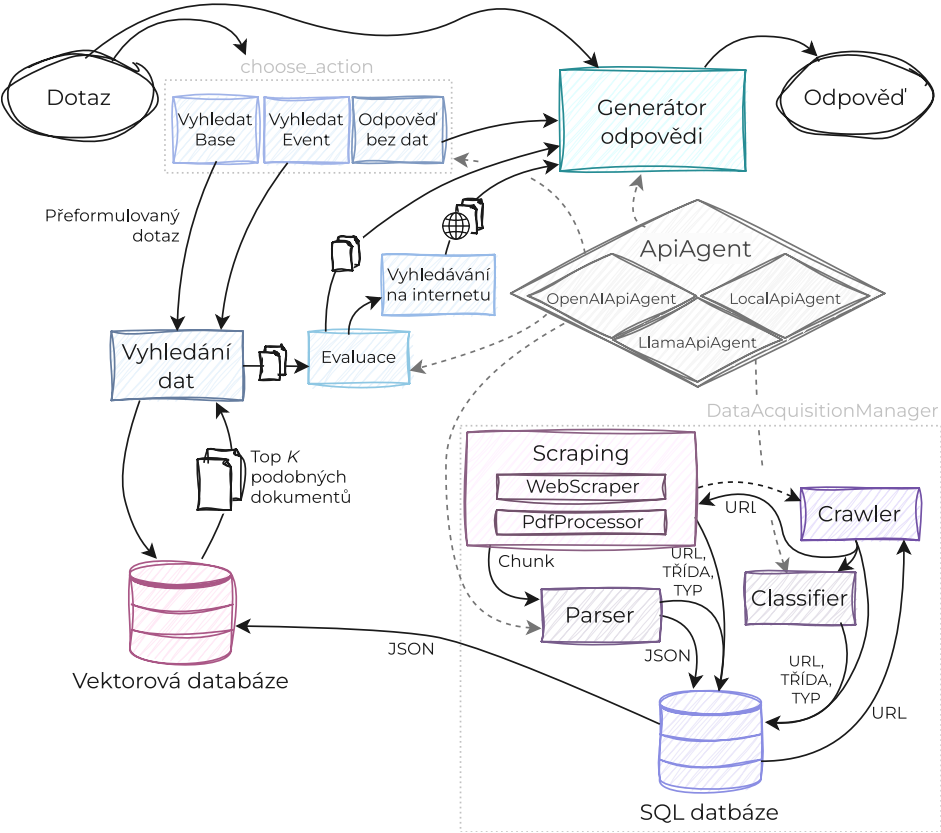
Autor: Bc. Kateřina Neprašová | Vedoucí práce: Smrž Pavel, doc. RNDr., Ph.D.

Cíl práce:

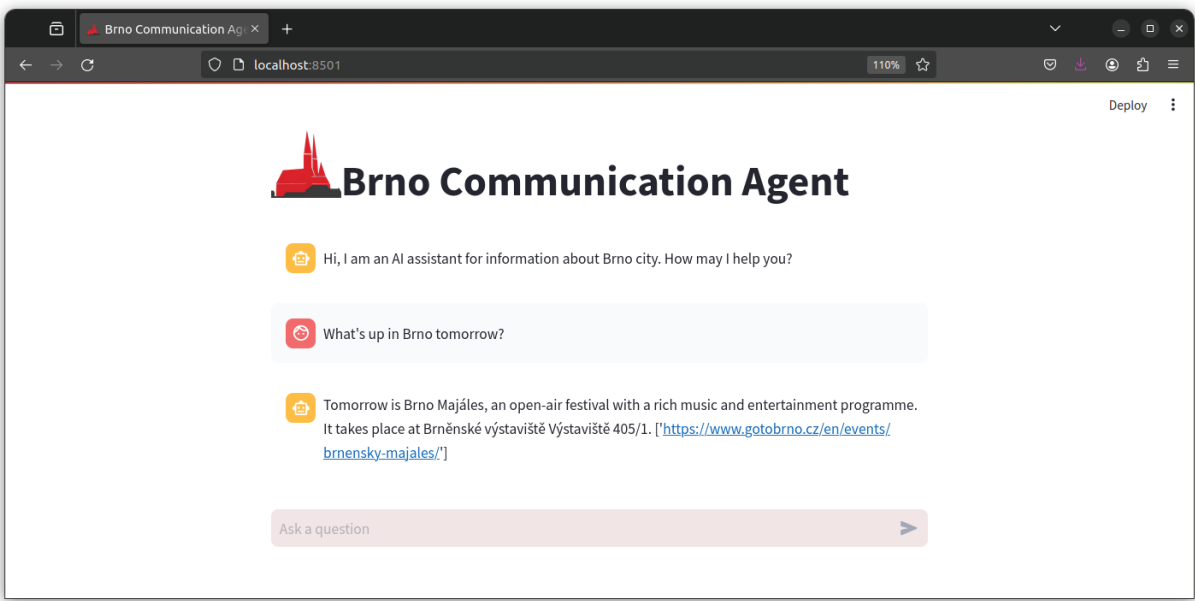
Tato diplomová práce se zaměřuje na doménově specifické komunikační agenty, cílem je vytvořit funkčního komunikačního agenta pro turisty i místní obyvatele Brna, poskytujícího relevantní a aktuální informace podle individuálních potřeb uživatelů.

Návrh systému:

Aby bylo zajištěno, že má komunikační agent k dispozici potřebné informace, je postaven na generování s rozšířeným vyhledáváním (RAG, Retrieval-Augmented Generation), tedy spojení jazykového modelu a vektorové databáze. Systém navíc obsahuje komponenty pro získávání dat, která se spolu s dalšími parametry ukládají do příslušných znalostních databází. Jádrem implementace je agent pro komunikaci s velkými jazykovými modely, který podporuje jak zpoplatněná API, tak open-source modely spouštěné lokálně. Tento klíčový prvek je kromě komunikace s uživatelem používán i při zpracování dat a volbě postupu pro generování odpovědi, evaluaci získaných dat pro odpověď a rozhoduje o potřebě vyhledávání dalších informací na internetu.



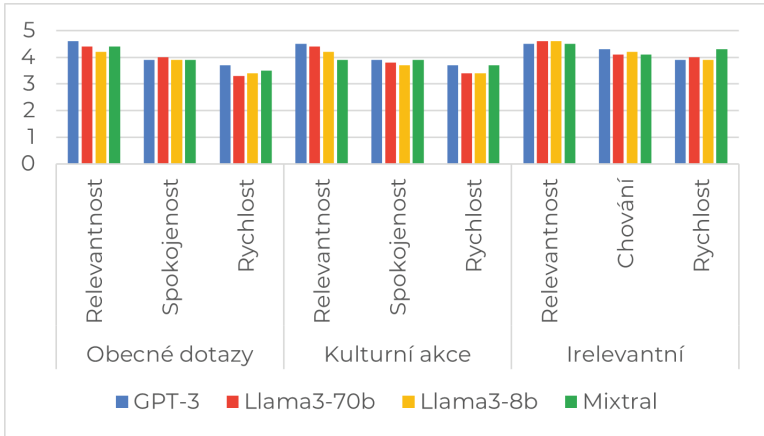
Výsledná aplikace:



Výsledky evaluace:

Model	Přesnost kontextu	Relevance odpovědi	Věrohodnost	Vyvolání kontextu
llama3-70b	0.803	0.899	0.868	0.689
llama3-8b	0.708	0.818	0.820	0.561
mixtral	0.835	0.899	0.924	0.648
gpt-3.5	0.886	0.925	0.864	0.699

Srovnání výkonnosti modelů vybranými RAGAs metrikami:



Průměrné hodnocení odpovědí systému respondenty dle modelu