

Zadání bakalářské práce



163475

Ústav: Ústav informačních systémů (UIFS)
Student: **Valík Adam**
Program: Informační technologie
Název: **Zpracování a vyhledávání dokumentů s využitím vektorových databází a jazykového modelu**
Kategorie: Data mining
Akademický rok: 2024/25

Zadání:

1. Seznamte se s problematikou zpracování přirozeného jazyka a vyhledávání textových dokumentů. Seznamte se s vektorovými databázemi.
2. Analyzujte požadavky na systém pro vyhledávání dokumentů, který rozdělí vstupní text na menší části vhodné velikosti, doplní k těmto částem dodatečné informace a kontext a upraví je do vektorové podoby. Dále umožní tvorbu dotazu uživatelem, provede vyhledání v databázi a zpracování odpovědi jazykovým modelem.
3. Navrhněte systém dle požadavků. Návrh konzultujte s vedoucím.
4. Implementujte navržený systém.
5. Otestujte vytvořený systém na vhodném vzorku dat a experimentálně ověřte úspěšnost vyhledávání.
6. Zhodnoťte dosažené výsledky a další možnosti pokračování tohoto projektu.

Literatura:

- Vaidyamath, R.C.: Vector Databases Unleashed: Navigating the Future of Data Analytics. Independent, 2024. ISBN 979-8321023488.
- Buttcher, S.: Information Retrieval: Implementing and Evaluating Search Engines. MIT Press Ltd., 2016. ISBN 978-0262528870.

Při obhajobě semestrální části projektu je požadováno:

Body 1-3.

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Bartík Vladimír, Ing., Ph.D.**
Vedoucí ústavu: Kolář Dušan, doc. Dr. Ing.
Datum zadání: 1.11.2024
Termín pro odevzdání: 14.5.2025
Datum schválení: 22.10.2024