|  |  |
| --- | --- |
| Chatbot Portál VŠ  15.7.2023 | Abstract  Řešení Chatbotu pro Portál VŠ Slovenské republiky  Vladimír Dědourek, Michal Stoklasa |

Obsah

[1 Portál VŠ 2](#_Toc141258336)

[1.1 Cíle řešení 2](#_Toc141258337)

[2 ChatBot 2](#_Toc141258338)

[2.1 Analýza struktury portálu 3](#_Toc141258339)

[2.1.1 Struktura souboru VŠ 3](#_Toc141258340)

[2.1.2 Orientační dotazy studentů 5](#_Toc141258341)

[2.2 Model embeddings 7](#_Toc141258342)

[2.2.1 Strukturální model textu 7](#_Toc141258343)

[2.2.2 Indexování strukturovaného textu 10](#_Toc141258344)

[2.2.3 Příklady operací v DB Qdrant 12](#_Toc141258345)

[2.3 Dotaz odpověď 14](#_Toc141258346)

[2.3.1 Struktura dat v DB 14](#_Toc141258347)

[2.3.2 Konstrukce promptu pro výběr textů na základě otázky 16](#_Toc141258348)

[2.3.2.1 Určení parametrů dotazu 17](#_Toc141258349)

[2.3.2.2 Určení přesné kategorie z uživatelem zadaných dat 18](#_Toc141258350)

[2.3.2.3 Sestavení podmínky pro výběr dat operací search z DB Qdrant 19](#_Toc141258351)

[2.3.2.4 Konstrukce promptu 19](#_Toc141258352)

# Portál VŠ

Webový portál VŠ je přístupný na drese

https://www.portalvs.sk/sk/

Základní motto portálu: 34 vysokých škol, 117 fakult, 4292 studijních programů na jednom místě.

Z toho plynou následující průměrné hodnoty:

3 – 4 fakulty na 1 universitu

37 programů na 1 fakultu

Cílem je vytvořit chatbota, který komunikuje se studenty s pomocí dat obsažených v portálu.

## Cíle řešení

Poptávaného AI chatbota (či Q&A bota) navrhujeme řešit metodou similarity search za pomoci vektorové databáze s použitím embeddings-ada-002 a ChatGPT modelů od OpenAI. Součástí řešení Chatbota by mělo být API rozhraní, umožňující jednoduchou implementaci do portálu. Chatbot se bude snažit vybrat z velkého množství textu relevantní odpověď k otázce uživatele.

Prvním krokem implementace tohoto řešení je provedení analýzy v rozsahu:

* Navržení architektury chatbota
* Vytvoření funkčního prototypu jako MVP („minimum viable product“)
  + Do prototypu by byl zahrnut ukázkový vzorek dat z webu <https://www.portalvs.sk/>
* Otestování prototypu za účelem splnění požadovaných funkcí
* Stanovení přesné pracnosti následného vývoje
* Prezentace výsledků

# ChatBot

Požadavky na funkcionalitu:

- rychlost

- maximální přesnost

- prioritně komunikace ve slovenském jazyce

- vícejazyčnost (možnost komunikace I v jiných jazycích)

- možnost jednoduchého komunikačního režimu (dotaz/výzva -> odpověď)

- možnost vytváření komunikačních řetězů

- založené na vlastních textových datech

- otevřené pro různé zdroje vstupních dat

## Analýza struktury portálu

Vlastní portál je značně rozsáhlý a obsahuje tisíce stránek. Proto bylo rozhodnuto použít soubory s agregovanými daty pro každou vysokou školu.

Tedy vstupem by mělo být 34 souborů ve shodné struktuře.

Příklady těchto souborů:

Slovenská poľnohospodárska univerzita v Nitre:

<https://portalvs.develop.maurit.sk/sk/ai/vysoka-skola/slovenska-polnohospodarska-univerzita-v-nitre>

Technická univerzita v Košiciach:

[https://portalvs.develop.maurit.sk/sk/ai/vysoka-skola/technicka-univerzita-v-kosiciach](https://portalvs.develop.maurit.sk/sk/ai/vysoka-skola/technicka-univerzita-v-kosiciach?fbclid=IwAR20gbNOjUWbfqeM2uinrkuV1cv457GCvvAoxOw_lyuWQ3IvdduayAJVPGU)

Žilinská univerzita v Žiline

[https://portalvs.develop.maurit.sk/sk/ai/vysoka-skola/zilinska-univerzita-v-ziline](https://portalvs.develop.maurit.sk/sk/ai/vysoka-skola/zilinska-univerzita-v-ziline?fbclid=IwAR08Hczx0dUIj41anikcHOASeGbmgrv7Je_UxlbkKlQndc7cXkXmTHHziqw)

### Struktura souboru VŠ

Orientační délka a počet tokenů je uveden pro „Slovenská poľnohospodárska univerzita v Nitre“.

|  |  |  |  |
| --- | --- | --- | --- |
| Hierarchie | Titulek | Popis | Délka / počet tokenů |
| 1 | VŠ | Např.  Slovenská poľnohospodárska univerzita v Nitre  Technická univerzita v Košiciach  Žilinská univerzita v Žiline | 1114719 / 496038 |
| 2 | Profil univerzity | Např.  Profil university  Obsahuje profil, adresu a seznam fakult | 4451/2007 |
| 1.1 | Fakulta | Např.  Fakulta agrobiológie a potravinových zdrojov  Fakulta biotechnológie a potravinárstva | 242157/111491 |
| 1.1.1 | Forma studia | Např.  Bakalárske štúdium  Inžinierske a Magisterské štúdium  Doktorandské štúdium | 87137/39874 |
| 1.1.2 | Profil fakulty |  | 1856/829 |
| 1.1.3 | Ubytování a stravování | Např.  Ubytovanie a stravovanie | 2763/1297 |
| 1.1.4 | Detail fakulty | Např.  Detail univerzity  Obsahuje kontakt, adresu, počet studentů, webovou adresu aj. | 524/234 |
| 1.1.1.1 | Studijní program | Např.  bezpečnostný manažment (denné)  bezpečnostný manažment (externé)  krízový manažment (denné)  krízový manažment (externé)  záchranné služby (denné)  záchranné služby (externé)  V textu je nedostatek v tom, že nadpis programu neobsahuje údaj v závorce, a to může působit zmatečně. | 6889/3129 |
| 1.1.1.1.1 | Titulek detailu | Např.  Bezpečnostný manažment  Uplatnenie absolventov  Podmienky prijatia  Doplňujúce informácie  Podmienky pre zahraničných študentov  Podmienky prijatia bez prijímacej skúšky  Koordinátor pre študentov so špecifickými potřebami  Všeobecné informácie k prijímacej skúške  Detail študijného programu  Předmět studia je popsaný v úvodním titulku studijního programu. Tedy v tomto případě je to pod titulkem detailu „Bezpečnostný manažment“ | 2789/1313  Pro Všeobecné informácie k prijímacej skúške |
|  |  |  |  |

Popis titulků detailu

|  |  |  |  |
| --- | --- | --- | --- |
| Pořadí | Titulek detailu | Popis | Délka / počet tokenů |
| 1 | Predme | Předmět studia. Např.  Bezpečnostný manažment |  |
| 2 | Uplatnenie absolventov |  |  |
| 3 | Podmienky prijatia |  |  |
| 4 | Doplňujúce informácie | Doplňující informace k přijímací zkoušce |  |
| 5 | Podmienky pre zahraničných študentov |  |  |
| 6 | Podmienky prijatia bez prijímacej skúšky |  |  |
| 7 | Koordinátor pre študentov so špecifickými potrebami |  |  |
| 8 | Všeobecné informácie k prijímacej skúške | Obsahuje i platby za přijímací zkoušku a roční školné a též fakturační údaje |  |
| 9 | Detail študijného programu | Obsahuje krom jiného kód programu, počet přijatých, délku studia, vyučovací jazyk |  |
|  |  |  |  |

Vysoký počet tokenů pro texty ve slovenském jazyce je dáno patrně tím, že modely nejsou příliš trénovány v těchto jazycích. Podrobněji se problémy s embeddings věnuje výzkum v <https://aclanthology.org/W19-6107.pdf>.

Pro similarity search je zde uváděno, že lepších výsledků je například dosaženo knihovnou FastText https://fasttext.cc/

### Orientační dotazy studentů

Studentům byla položena následující otázka:

Představte si, že končíte střední školu a rozhodujete se kam dál půjdete na vysokou. S výběrem by vám měl pomoci webový portál vysokých škol, na kterém najdete informace o každé univerzitě, oboru, přijímací zkoušce atd... Webový portál má AI chatbota, který je schopen studentům zodpovědět veškeré otázky související s vysokými školami. Na jaké otázky byste se chatbota ptali (alespoň 4) nebo co by vás na takovém portálu zajímalo nejvíce?

Otázky:

- zeptal bych se, aby mi našel školy podle mých zájmů, požadavků, atd.

- jaké školy jsou v určeném okolí nebo městech

- požadavky na přijetí, nutnost dělání přijímacích zkoušek, případně co musí být splněno, aby mi byly odpuštěny

- něco o studiu a oborech, které ty školy nabízejí

- kdy se konají zkoušky a všechny termíny; vyučování předměty; požadavky ke zkouškám; info o daných oborech na VŠ;

- co a kde najdu jako jsou jídelny, podniky poblíž, studijní oddělení, kabinety, obory jaké jsou; určitě recenze studentů a jejich hodnocení daně VŠ; jaké programy nabízejí (bc./Mg./Ph.D) a doba trvání + možnost prodloužení či ne; počet přijatých studentů v daném oboru a daně VŠ; kredity a pokažení zkoušky a jak postupovat k napravení; co očekávat od VŠ; nějaký vzorový harmonogram.

- No to je těžká otázka, věci, na které by se měl člověk ptát, mu většinou dotečou až v průběhu studia a ne při výběru školy. Ale určitě nějaký rozbor oborů a zaměření, uplatnění absolventů, koleje, stravování, pro sociálně slabší třeba i stipendia a hlavně důležité termíny či potřebné URL.

- jaká je úspěšnost studia

- jaké je uplatnění absolventů ( % zaměstnanost nebo konkrétní profese)

- v jakém prostředí absolventi pracuji (národní/mezinárodní)

- jaký plat může úspěšný absolvent očekávat (nástup nebo po 5 letech)

- jaká je budoucnost daného oboru/profese v rámci AI a změn na trhu práce

- Kde najdu potřebné informace, jak probíhá přijímací řízení, jak zařídit ubytování, doporučení kolejí, jak zařídit stravování, studijní obory, uplatnitelnost v praxi, všechny potřebné termíny, možnost doporučení škol podle oborů nebo zaměstnání.

- Obor, místo, prezenční/dálkové/kombinované studium

- Využitelnost v praxi. Zaměstnanost absolventů.

Z výše uvedených otázek vyplývá následující:

- řada informací je obsažena v kontextu informací o VŠ (požadavky na přijetí, nutnost přijímací zkoušky, výběr podle zájmů …)

- některé informace nejsou v kontextu dostupné, ale daly by se zjistit v obecném GPT modelu (jaké školy jsou dostupné v aktuálním okolí, v okolí města)

- některé informace nejsou dostupné nikde (jaká je úspěšnost studia)

- určitě by mělo být v systému Chatbota prováděno logování dotazů a odpovědí. Z nich by se dala vytvořit zpětná vazba na tvorbu doporučení pro zkvalitnění obsahu webu, který by více odpovídal potřebám studentů

## Model embeddings



### Strukturální model textu

Pro jednu univerzitu je k dispozici jeden textový soubor, který lze extrahovat z URL pro příslušnou univerzitu.

Členění v hierarchii je následující:

|  |  |  |
| --- | --- | --- |
| Úroveň hierarchie | Název úrovně | Zkratka |
| 1 | Univerzita | U |
| 2 | Fakulta | F |
| 3 | Forma studia | FS |
| 4 | Studijní program | P |
| 5 | Titulek detailu | T |

Text v hierarchii pro jeden soubor bude strukturalizován následujícím způsobem:

Univerzita:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Univerzita | Fakulta | Forma studia | Studijní program | Titulek detailu | Popis |
| Název univerzity |  |  |  | Profil univerzity | Profil univerzity |
| dtto |  |  |  | Univerzita | Název univerzity |
|  |  |  |  |  | Zbývající text je dělen pro jednotlivé fakulty |

Fakulta

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Univerzita | Fakulta | Forma studia | Studijní program | Titulek detailu | Popis |
| Název univerzity | Název fakulty |  |  | Profil fakulty | Profil fakulty |
| dtto | dtto |  |  | Fakulta | Název fakulty |
| dtto | dtto |  |  | Ubytovanie a stravovanie | Ubytování a stravování |
| dtto | dtto |  |  | Detail fakulty | Detailní informace o fakultě |
|  |  |  |  |  | Zbývající text je dělen pro jednotlivé formy studia |

Forma studia

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Univerzita | Fakulta | Forma studia | Studijní program | Titulek detailu | Popis |
| Název univerzity | Název fakulty | Forma studia |  | Zoznam študijných programov | Seznam studijních programů |
|  |  |  |  |  | Zbývající text je dělen pro jednotlivé studijní programy |

Studijní program

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Univerzita | Fakulta | Forma studia | Studijní program | Číslo studijního programu | Titulek detailu | Popis |
| Název univerzity | Název fakulty | Forma studia | Název studijního programu | dtto | Predmet | Předmět studia  Předmět nemusí být uveden, potom je třeba použít Uplatnění absolventů |
| dtto | dtto | dtto | dtto | dtto | Program | Kód a název studijního programu |
| dtto | dtto | dtto | dtto | dtto | Uplatnenie absolventov |  |
| dtto | dtto | dtto | dtto | dtto | Podmienky prijatia |  |
| dtto | dtto | dtto | dtto |  | Podmienky pre zahraničných študentov |  |
|  |  |  |  | dtto | Podmienky prijatia bez prijímacej skúšky |  |
| dtto | dtto | dtto | dtto | dtto | Koordinátor pre študentov so špecifickými potrebami |  |
| dtto | dtto | dtto | dtto |  | Všeobecné informácie k prijímacej skúške |  |
|  |  |  |  |  | Školné, poplatky a fakturačné údaje | Školné, poplatky a fakturační údaje  Ročné školné |
| dtto | dtto | dtto | dtto |  | Detail študijného programu |  |

Finální struktura textu není normalizovaná a je složena z řádků se stejnou strukturou. Některá pole nemusí být naplněna viz tabulky výše

Tabulka metadat:

|  |  |
| --- | --- |
| Název pole | Popis |
| university | Název univerzity |
| faculty | Název fakulty |
| form | Forma studia (bakalářské, magisterské, doktorské) |
| program | Název studijního programu |
| code | Kód studijního programu (175409, …) |
| heading | Název paragrafu |
| text | Text paragrafu |

Pro každý textový soubor FILE\*.txt je vytvořen soubor strukturovaného textu FILE\*.xlsx.

A computer screen shot of a program

Description automatically generated

A screenshot of a computer

Description automatically generated

### Indexování strukturovaného textu

Pro indexování se berou texty z tabulky metadat. Tyto texty jsou rozděleny na menší segmenty (chunks) tak, aby maximální velikost nepřekročila maximální počet tokenu (max\_tokens).

Protože pouhým indexováním těchto segmentů (chunks) bez příslušnosti k názvu paragrafu (heading), by mohlo dojít ke ztrátě významové souvislosti, Jsou názvy paragrafu připojeny vždy na začátek jednotlivých segmentů.

Při dělení na segment jsou nedělitelné jednotky celé věty. V případě, že by se věta nevešla do segment, je provedeno dělení na úrovni slov. K rozdělení slov mezi segment nedochází.

Pro každý segment (chunk) jsou ve vektorové databázi přiřazena následující metadata.

Tabulka metadat ve vektorové databázi:

|  |  |
| --- | --- |
| Název pole | Popis |
| university | Název univerzity |
| faculty | Název fakulty |
| form | Forma studia (bakalářské, magisterské, doktorské) |
| program | Název studijního programu |
| code | Kód studijního programu (175409, …) |
| heading | Název paragrafu |
| text | Segment textu, který odpovídá vektoru embeddings |
| n\_tokens | Počet tokenů segmentu dat (chunk) |

Průběh indexování 3 univerzit trvalo 796 s.

A screenshot of a computer

Description automatically generated

Indexovaná data jsou uložena v:

|  |  |
| --- | --- |
| Název | Hodnota |
| Databáze | Qdrant |
| Collection | www.portalvs.sk |
| Qdrant url | https://f86a5598-6877-4dca-a5f1-e8555124d321.us-east-1-0.aws.cloud.qdrant.io:6333 |
| Qdrant API key | VgIFxRf3pEdRUJAPnVWrg0ixn5UAkp5KPLkDZtuBBZPY3pbhDnugQA |
| Vector size | 1536 |
| Vector distance | models.Distance.COSINE |
| Payload | "university", "faculty", "form", "program", "code", "heading", "text", "n\_tokens" |
|  |  |

Dokumentace Qdrant databáze: https://qdrant.tech/documentaion/

### Příklady operací v DB Qdrant

Připojení k DB Qdrant

from qdrant\_client import QdrantClient

# open Qdrant database

qdrant\_client = QdrantClient(

url=os.getenv("QDRANT\_URL"),

api\_key=os.getenv("QDRANT\_API\_KEY"),

)

Vytvoření vektoru embeddings ze zadaného dotazu:

# get embeddings vector

question\_embeddings = openai.Embedding.create(input=emb\_question, engine=engine)['data'][0]['embedding']

Načtení dat z DB prostřednictvím similarity search:

# 1 - search vectors in Qdrant with the best distance

result = qdrant\_client.search(

collection\_name = "www.portalvs.sk",

search\_params=models.SearchParams(

hnsw\_ef=128,

exact=True

),

query\_vector = question\_embeddings,

limit=5 # Return 5 closest points

)

Načtení dat z DB prostřednictvím similarity search s dodatečnou podmínkou:

# 2 - search vectors in Qdrant with the best distance and filter condition

result = qdrant\_client.search(

collection\_name="www.portalvs.sk",

query\_filter=models.Filter(

must=[

models.FieldCondition(

key="heading",

match=models.MatchValue(

value="Predmet",

),

)

]

),

search\_params=models.SearchParams(

hnsw\_ef=128,

exact=True

),

query\_vector=question\_embeddings,

limit=5 # Return 5 closest points with condition

)

print("\n2 The best distance and filter condition")

for record in result:

university = record.payload["university"]

code = record.payload["code"]

program = record.payload["program"]

print(f"Univerzita: {university} Program: {code} {program}")

Zjištění počtu záznamů splňujících podmínku:

# 3 - search count records in Qdrant with filter condition

record\_count = qdrant\_client.count(

collection\_name="www.portalvs.sk",

count\_filter=models.Filter(

must=[

models.FieldCondition(

key="heading",

match=models.MatchValue(value="Predmet"),

),

]

),

exact=True,

)

print(f"\n3 Number of records: {record\_count}")

Výběr dat z DB pouze dle zadané podmínky:

# 4 - scroll records in Qdrant with condition (programs)

print("\n4 Retrieve records with filter condition")

program\_tokens = 0

counter = 0

next\_id = 0

data = []

while next\_id != None:

result = qdrant\_client.scroll(

collection\_name="www.portalvs.sk",

offset = next\_id,

scroll\_filter=models.Filter(

must=[

models.FieldCondition(

key="heading",

match=models.MatchValue(value="Predmet"),

),

]

),

consistency = 'all',

with\_payload=True,

with\_vectors=False,

limit=10000,

)

next\_id = result[1]

for record in result[0]:

program\_tokens += record.payload["n\_tokens"]

university = record.payload["university"]

faculty = record.payload["faculty"]

form = record.payload["form"]

code = record.payload["code"]

program = record.payload["program"]

heading = record.payload["heading"]

text = record.payload["text"]

counter += 1

print(f"{counter:0=4} {university} | {faculty} | {code} {program} | {heading}")

#data.append((university, code, program, heading, text))

print(f"\nPočet tokenů: {program\_tokens} počet záznamů {counter}")

## Dotaz odpověď

### Struktura dat v DB

Data v DB mají následující logickou strukturu.



Každý vrchol stromu může obsahovat různé kategorie textů.





Jednotlivé záznamy jsou uloženy v jednotné struktuře viz následující tabulka

|  |  |
| --- | --- |
| Název pole | Popis |
| university | Název univerzity |
| faculty | Název fakulty |
| form | Forma studia (bakalářské, magisterské, doktorské) |
| program | Název studijního programu. Údaj přísluší ke kódu studijního programu. |
| code | Kód studijního programu (175409, …). Kód studijního programu je unikátní přes všechny univerzity. Proto při vyhledání programu je tím jednoznačně určena forma studia, fakulta i univerzita. |
| heading | Název paragrafu. Každý text má jednoznačný název paragrafu. Proto lze v databázi vyhledávat texty podle názvu paragrafu. |
| text | Segment textu, který odpovídá vektoru embeddings |
| n\_tokens | Počet tokenů segmentu dat (chunk) |

Seznam všech názvů paragrafů

|  |  |  |
| --- | --- | --- |
| heading | Příslušnost k | Popis |
| Univerzita | Univerzita | Název univerzity |
| Profil univerzity | Univerzita | Stručná charakteristika univerzity |
| Fakulta | Fakulta | Název fakulty |
| Profil fakulty | Fakulta | Stručná charakteristika předmětů studia a uplatnění absolventů fakulty. Ovšem v některých případech jsou zde pouze údaje o založení a odkazy. |
| Ubytovanie a stravovanie | Fakulta | Informace o ubytování a stravování na fakultě |
| Detail fakulty | Fakulta | Kontaktní informace fakulty. Počty studentů (denních a externích). Odkaz na elektronickou přihlášku a uplatnění absolventů. |
| Zoznam študijných programov | Forma studia | Seznam názvů studijních programů pro danou formu studia |
| Program | Studijní program | Kód a název studijního programu |
| Predmet | Studijní program | Předmět studijního programu |
| Uplatnenie absolventov | Studijní program | Uplatnění absolventů |
| Podmienky prijatia | Studijní program | Podmínky pro přijetí studenta na daný studijní program |
| Podmienky pre zahraničných študentov | Studijní program | Podmínky pro přijetí zahraničních studentů |
| Podmienky prijatia bez prijímacej skúšky | Studijní program | Podmínky pro přijetí studenta bez přijímací zkoušky na daný studijní program |
| Koordinátor pre študentov so špecifickými potrebami | Studijní program | Kontakt na koordinátora pro studenty se specifickými potřebami |
| Všeobecné informácie k prijímacej skúške | Studijní program | Detailní informace o přijímací zkoušce. Termín podání přihlášky, termín zkoušky, termín zasedání přijímací komice. Předpokládaný počet přijatých studentů, počet přihlášených studentů. |
| Školné, poplatky a fakturačné údaje | Studijní program | Poplatek za přijímací zkoušku, poplatky za studium. Fakturační údaje. |
| Detail študijného programu | Studijní program | Předpokládaný počet přijatých studentů. Kód studijního programu. Stupeň studia, délka studia, metoda studia, titul absolventů, vyučovací jazyky |

Na data uložená v databázi lze pohlížet následujícím způsobem



Každý záznam je složen z dimenzí a faktů.

Dimenze jsou tvořeny vlastnostmi s konečnými množinami hodnot. Lze na ně uplatnit výběrovou podmínku.

Fakta jsou tvořena jednotlivými typy textů. Lze na uplatnit výběr (pouze některé sloupce).

V databázi jsou jednotlivá fakta dána dvojicí (heading, text).

### Konstrukce promptu pro výběr textů na základě otázky

Implementace je provedena v jazyce Python v rámci třídy **QnA.**

Metody třídy **QnA**

|  |  |
| --- | --- |
| Název metody | Popis |
| get\_question\_parameters | Určení parametrů dotazu, které jsou použity ve filtru při sémantickém vyhledání v DB Qdrant |
| get\_un\_dimension | Určení přesného názvu kategorie dat zadaných uživatelem  Přibližný název univerzity -> přesný název univerzity  Přibližný název fakulty -> přesný název fakulty  Přibližný název studijního programu -> přesný název studijního programu |
| get\_question\_headings | Získání názvů polí heading, které budou získány z DB Qdrant |
| get\_qdrant\_condition | Sestavení podmínky pro výběr dat operací search z DB Qdrant |
| answer\_question | Získání odpovědi na zadanou otázku |
|  |  |

#### Určení parametrů dotazu

Po zadání otázky studenta je třeba určit následující informace

|  |  |  |
| --- | --- | --- |
| Informace | Zadaný dotaz směřuje k | Další informace |
| Výběr univerzity | výběru univerzity |  |
| Výběr fakulty | výběru fakulty |  |
| Výběr studijního programu | výběru studijního programu |  |
| Detailní info o univerzitě | získání informací o univerzitě | Název univerzity |
| Detailní info o fakultě | získání informací o fakultě | Název univerzity, název fakulty |
| Detailní info o studijním programu | získání informací o studijním programu | Kód studijního programu |

Způsob vyhledání požadovaných informací je na následujícím obrázku



Parametry:

question – uživatelem zadaná otázka

Výstup:

JSON struktura ve formátu

question\_type - Detail, Výber

question\_subject - Univerzita, Fakulta, Program

university - název univerzity

faculty - název fakulty

form - forma studia

program - název studijního programu

code - kód studijního programu

content - v případě, že se otázka týkala univerzit obsahuje None. Jestliže je úplně mimo toto téma, je zde naplněna odpověď z ChatGPT.

Výběr faktů pro jednotlivé druhy výběru

|  |  |
| --- | --- |
| Informace | headings |
| Výběr univerzity | Profil univerzity |
| Výběr fakulty | Profil univerzity, Profil fakulty, Detail fakulty |
| Výběr studijního programu | Predmet, Uplatnenie absolventov, Zoznam študijných programov |
| Detailní info o univerzitě | Profil univerzity |
| Detailní info o fakultě | Profil univerzity, Profil fakulty, Detail fakulty, Ubytovanie a stravovanie |
| Detailní info o studijním programu | Predmet, Uplatnenie absolventov, Podmienky prijatia, Podmienky pro zahraničných študentov, Podmienky prijatia bez prijímacej skúšky, Všeobecné informácie k prijímacej skúške, Školné, poplatky a fakturačné údaje, Detail študijného programu |



**get\_question\_heading**

Parametry:

question\_type - Detail, Výber

question\_subject - Univerzita, Fakulta, Program

Výstup:

Seznam názvů heading dle tabulky výše

#### Určení přesné kategorie z uživatelem zadaných dat

Při použití vyhledávací kategorie pro nastavení filtru na výběr dat z vektorové DB Qdrant je třeba použít exaktní název, který se v DB vyskytuje. Při zadání dotazu klientem nelze exaktní název zaručit, a proto je nutné provést jeho transformaci prohledáním v Qdrant databázi prostřednictvím sémantického vyhledání.

Způsob získání přesného názvu je na následujícím obrázku



Parametry:

dim\_type – typ kategorie (Univerzita, Fakulta, Program)

text – orientační název kategorie

Výstup:

Exaktní název kategorie

#### Sestavení podmínky pro výběr dat operací search z DB Qdrant

Podmínka pro sestavení filtru závisí:

- parametrech dotazu určených v get\_question\_parameters

- hlavičkách textů určených v get\_question\_heading

Nastavení filtrů dle parametrů dotazu

|  |  |  |
| --- | --- | --- |
| question\_type | question\_subject | Filtrované položky |
| Výber | Univerzita | - |
| Výber | Fakulta | university |
| Výber | Študijný program | university, faculty, form |
| Detail | Univerzita | university |
| Detail | Fakulta | university, faculty |
| Detail | Študijný program | university, faculty, form, code, program |
|  |  |  |

Parametry question\_type a question\_type jsou povinné a v parametrech jsou vždy uvedeny. Parametry pro filtrované položky nemusí být uvedeny a potom se použijí pouze ty, které jsou v parametrech uvedeny.

V případě parametrů (university, faculty, program) nemusí být jejich formulace přesná, a proto je provedeno zpřesnění nazvu metodou get\_un\_dimension.

Další filtry na položku heading jsou nastaveny z hlaviček textů určených v get\_question\_heading.



**get\_qdrant\_condition**

Parametry:

question\_parameters – seznam parametrů dotazu z get\_question\_parameters

question\_headigs – seznam hlaviček dotazu zjištěný v get\_question\_headings

Výstup:

Seznam podmínek FieldCondition, které jsou použity v search operaci v DB Qdrant.

#### Konstrukce promptu

Nejprve je analyzována otázka metodou get\_question\_parameters a jsou zjištěny parametry dotazu. Zde je použita funkce s modelem GPT-4 (model GPT-3.5 turbo tuto funkcionality příliš dobře nezvládá). Jestliže dotaz směřuje mimo oblast univerzit, potom není naplněn parametr question\_type a v parametru content je již přímo odpověď na dotaz.

Např. otázka:

"Jakou technologií je vyrobena atomová bomba?"

V content je odpověď.

„Prepáčte, ale ako univerzitný poradca nemám informácie o technológii výroby atómovej bomby. Môžem vám však pomôcť s otázkami týkajúcimi sa univerzít, fakúlt a študijných programov.“

Na základě parametrů question\_type a question\_subject jsou určeny typy vybíraných faktů metodou get\_question\_headings.

Nakonec je konstruována výběrová podmínka do DB Qdrant metodou get\_qdrant\_condition.



Prompt je poté sestaven z načtených segmentů z DB Qdrant prostřednictví operace search.

### Proces dotaz odpověď

Zadaná otázka je zpracována metodou get\_question\_parameters, ve které jsou určeny podmínky pro filtr dat v DB Qdrant. Poté je proveden similarity search pro danou otázku s aplikovaným filtrem a jsou získány segmenty zdrojových dat (chunks) pro konstrukci odpovědi. Jeliko6 segmenty jsou seřazeny dle ranking čísla, je provedeno seřazení jednotlivých segmentů do logické sekvence, která odpovídá řazení ve zdrojovém dokumentu.

Nakonec je proveden dotaz se zadaným kontextem funkcí ChatCompletion v OpenAI v modelu GPT-4.



Orientační časy pro vybrané otázky ukazuje následující tabulka

|  |  |
| --- | --- |
| Funkce | Doba zpracování [s] |
| Jaké jsou přijímací požadavky pro bakalářské studijní programy na Technické univerzitě v Košiciach? | |
| Get\_question\_parameters | 8.075 |
| Similarity search | 0.579 |
| Chat completion | 7.366 |
|  |  |
| Kdy se konají přijímací zkoušky na Žilinské univerzitě v Žilině pro akademický rok 2023/2024? | |
| Get\_question\_parameters | 5.402 |
| Similarity search | 0.738 |
| Chat completion | 5.179 |
|  |  |
| Mohl byste mi poskytnout informace o oborech, které nabízí Žilinská univerzita v Žilině? | |
| Get\_question\_parameters | 6.311 |
| Similarity search | 0.819 |
| Chat completion | 8.193 |

Zpracování je řešeno prostřednictvím třídy v Pythonu s následujícími parametry:

A screen shot of a computer program

Description automatically generated

## Souhrn

Výpočetní čas se převážně spotřebovává při zjištění parametrů dotazu a potom i při generování odpovědi. Vyhledání ve vektorové DB je zpravidla menší než 1 s.

Při zpracování otázky, kde data jsou v mnoha segmentech, nejsou odpovědi kompletní nebo správné. Důvodem je to, že data pro vytvoření vhodného kontextu jsou příliš rozsáhlá a přesahují povolený limit tokenů (viz <https://platform.openai.com/docs/models/overview>).

Ve zdrojových datech chybí informace o studovaných předmětech jednotlivých studijních programů. V případě seznamu studijních předmětů není spolu s názvem uveden i kód předmětu.