Národní knihovna – koncept PSM, PSK

Zpracovatel:

Bootiq s.r.o.
Hybernská 1007/20, 110 00, Praha 1
IČO: 291 55 495

Obsah

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of biq.group

Prohlášení o pravdivosti informací

Já, níže podepsaný, [redacted], jednatel společnosti Bootiq s.r.o., IČO 29155495, tímto prohlašuji, že všechny informace uvedené v předloženém Konceptu jsou úplné, pravdivé a odpovídají skutečnosti.

V Praze dne 20.01.2025

[redacted]

BOOTIQ s.r.o.

Podepsáno v Signi | Signi.com

[redacted], jednatel

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of big.group

Představení společnosti

BOOTIQ je česko-slovenská IT společnost specializující se na vývoj softwaru, digitální transformaci a inovativní IT řešení. Naše týmy tvoří špičkoví odborníci v oblasti vývoje, analýzy a implementace IT projektů, kteří se vždy dokážou přizpůsobit specifickým potřebám klientů. Díky širokému portfoliu služeb, od vývoje na míru až po implementaci komplexních platforem, pomáháme našim klientům růst a posouvat jejich podnikání směrem k digitální budoucnosti.

Společnost BOOTIQ je součástí dynamicky rostoucí česko-slovenské ICT skupiny BIQ Group, která se opírá o více než 25 let zkušeností v oboru. Skupina sdružuje více než 450 odborníků v 11 pobočkách a dlouhodobě podporuje úspěšné online a digitální projekty. BIQ Group nabízí široké portfolio služeb a produktů, které pokrývají potřeby rostoucích firem, zavedených korporací i veřejného sektoru. Naším cílem je být spolehlivým a dlouhodobě důvěryhodným partnerem, který pomáhá klientům dosahovat jejich cílů.

Oblasti našeho zaměření v BIQ Group:
Consulting, front-end a back-end vývoj, grafika, design, UX/UI, CMS, zakázkový vývoj aplikací, systémů, portálů, řízení procesů, integrace, Data & AI, DevOps, SAP, Microsoft, Atlassian a mnoho dalších.

Kontaktní údaje

[redacted] – obchod, [redacted]@biq.group, [redacted]

[redacted] – architekt a vývoj, [redacted]@bootiq.io, [redacted]

BOOTIQ
BIQ Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of biq.group

Reakce na technickou specifikaci

V následující sekci přikládáme reakci na jednotlivé body z technické specifikace, případné dotazy bude nutné zodpovědět ze strany NK před finálním podáním nabídky.

1. Hlasové zadání dotazu:

   V dokumentu č. 4 technické specifikace, se nachází požadavek na real-time přepis dotazů z mluveného slova na text (tzv. Speech-to-Text), zde doporučujeme změnu na zpracování slova na text až po dokončení celého dotazu. Taková služba bude mít menší náklady, aktuálně uvažujeme o využití některé ze služeb 3. strany, finální výběr bude na základě kvality přepisu poskytnutého vzorku dat.
2. Výběr kategorie A (scénář č. 0):
   a. Zde bude nutné vybrat správný model, který budeme používat k vytváření embeddings pro jednotlivé dotazy, hlavní otázkou zde bude nastavení cenové politiky.
   b. Upřesnit intervaly confidence, aktuálně není definované chování pro hodnotu 80%.
   c. Bude potřeba upřesnit statistiky, jaké budeme sbírat.
   d. Dle našeho názoru bude nutné postupovat s pomocí PoC (Proof-of-Concept), toto menší řešení ukáže, jestli je daný postup dostatečný pro dané použití. Hlavní výhodou je možnost velice rychle upravit chování konceptu.
3. Navrhování odpovědí na nevyřízené dotazy (scénáře 2a, 2b a 10):
   a. Zde je hlavním bodem to, jak doplníme do AI data, je totiž možné vzít již existující model a doučit jej o nová data, ale lepším řešením, které i my navrhujeme, je využití RAG (Resource Acquisition Generation) pipeline. Je důležité taky zmínit, že v případě použití existujícího modelu, který budeme učit vlastními daty, bude výsledný model vždy s novou sadou dat potřeba znovu natrénovat a zkontrolovat jeho přesnost (odchylka od nějakého lokálního maxima).
   b. Použití RAG by pak mělo i výhodu pro jednoduchou integraci s více AI chat nástroji.
4. Generování štítků typu A v archivu dotazů:
   a. Proč chcete 2 různé metody ke generování štítků? Implementovat obě metody bude nákladné.

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of biq.group

5. Generování obrázků k dotazům:
    a. Jaká licence má být u generovaných obrázků? (commercial, noncommercial, attribution...)
6. Obecně k AI:
    a. Hlavním faktorem, který bude ovlivňovat výběr technologií je velikost dat a počet dotazů.
7. SMTP server:
    a. V různých krocích je potřeba odesílat emailové notifikace, kdo bude provozovat SMTP server?
8. Uživatelské rozhraní:
    a. Při vytváření rozhraní a jeho testování je potřeba splnit požadavky EAA (European Accessability Act).

Technická specifikace slouží jako dobrý základ pro budoucí analýzu celého projektu. Nicméně v průběhu jejího vypracování mohou vzniknout další dotazy, které bude nutné zodpovědět ze strany NK.

Reakce na předpokládaný harmonogram

Děkujeme za podrobnou zadávací dokumentaci a definovaný harmonogram vývoje systémů PSK a PSM. Po důkladném prostudování bychom rádi poskytli naše stanovisko k projektové metodice, preferovanému přístupu a návrhy na možné úpravy harmonogramu s cílem dosáhnout vyšší míry agility, kterou považujeme za optimální pro podobné projekty.

1. Dotazy k metodice waterfall vs. agile

Zadávací dokument zahrnuje pevně stanovené fáze a milníky, což se přibližuje iterativnímu waterfall modelu. Nicméně některé rysy agilního přístupu jsou patrné, například důraz na průběžnou zpětnou vazbu (např. testování beta verzí). Rádi bychom položili několik otázek pro upřesnění:

- Bude během jednotlivých fází (např. návrh systému nebo testování beta verze) zadavatel zapojen do pravidelných kontrolních setkání (např. formou sprint review)?
- Má zadavatel preferenci ohledně periodicity doručování výsledků (např. průběžné inkrementy místo velkých výstupů po ukončení celé fáze)?
- Je prostor pro implementaci produktového backlogu a pravidelnou prioritizaci úkolů ve spolupráci se zadavatelem?

BOOTIQ
BIQ Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of biq.group

2. Co upravit, aby se přístup přiblížil agilní metodice?

Abychom harmonogram více přizpůsobili agilnímu přístupu, navrhujeme tyto úpravy:

- **Rozdělení dlouhých fází na menší iterace:** Například fázi návrhu systému (5 měsíců) rozdělit na dvou až třítýdenní iterace, během nichž by byly pravidelně prezentovány dílčí výsledky, např. konkrétní moduly UML modelu, návrhy architektury nebo ukázky prvků UX.
- **Průběžná akceptace:** Zavést průběžné akceptace menších částí místo akceptace celých fází. To by umožnilo flexibilnější reakce na změny požadavků.
- **Zapojení zadavatele do iterativního procesu:** Organizovat pravidelná setkání (např. každé dva týdny) za účelem kontroly průběhu vývoje, prioritizace požadavků a hodnocení pokroku.
- **Pilotní nasazení:** U beta verzí doporučujeme postupné nasazení s možností průběžného zapracování zpětné vazby od uživatelů, což umožní efektivnější odstranění nedostatků.

3. V čem nevnímáme úplnou agilitu?

- **Dlouhé fáze s pevnými milníky:** Harmonogram aktuálně obsahuje rozsáhlé časové bloky (např. 5 měsíců na návrh systému), což není v souladu s kratšími iteracemi typickými pro agilní vývoj.
- **Chronologická závislost fází:** Každá fáze začíná až po dokončení a akceptaci předchozí, což snižuje flexibilitu vývoje.
- **Fixní požadavky:** Pokud zadavatel očekává možnost průběžných změn a doplňování funkcionalit, současný model nemusí plně podporovat takovou dynamiku. Např. AI se rychle vyvíjí a půl roku může např. znamenat zastarání používaného modelu.

4. Dotazy na upřesnění

- Jaká jsou očekávání od dodavatele během jednotlivých fází, pokud jde o průběžnou komunikaci a kontrolu výstupů?
- Je prostor pro úpravu harmonogramu tak, aby reflektoval kratší iterace?
- Jak zadavatel vnímá riziko změn během vývoje a jejich možný dopad na harmonogram?
- Bude preferováno využívání agilních nástrojů (např. nástroje na správu backlogu, pravidelné sprint review)?

Závěr

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of big.group

Na základě výše uvedených návrhů jsme připraveni přizpůsobit náš přístup tak, abychom maximalizovali přidanou hodnotu pro váš projekt. Naše preferovaná metodika je agilní vývoj, který umožňuje flexibilitu, průběžnou kontrolu a inkrementální přístup k tvorbě systému, čímž snižujeme riziko nedostatků v pozdějších fázích.

Rádi s vámi detailně prodiskutujeme možnosti úprav harmonogramu a nastavení spolupráce tak, aby odpovídaly vašim potřebám a očekáváním.

Klíčové oblasti

- Využití AI při přiřazování kategorií.
    - Cena za provoz, ta se odvíjí od počtu volání, při využití komerčních API.
    - Trénování vlastního modelu je dražší varianta ovšem s větší flexibilitou.
- UI/UX
    - Návrh bude další klíčová oblast, aplikace by měla být uživatelsky přívětivá a v moderním provedení, které bude myslet na jednoduchost použití i pro starší uživatele.
    - Výsledný web musí splňovat náležitosti, jako například EAA.
- Provoz
    - Náklady na provozování cloudových služeb.
    - Stabilita a bezpečnost řešení, splňování NIS2.
    - Podpora pro koncové uživatele a správce systému.

BOOTIQ
BIQ Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of big.group

Příloha č. 1 – Návrh architektury

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

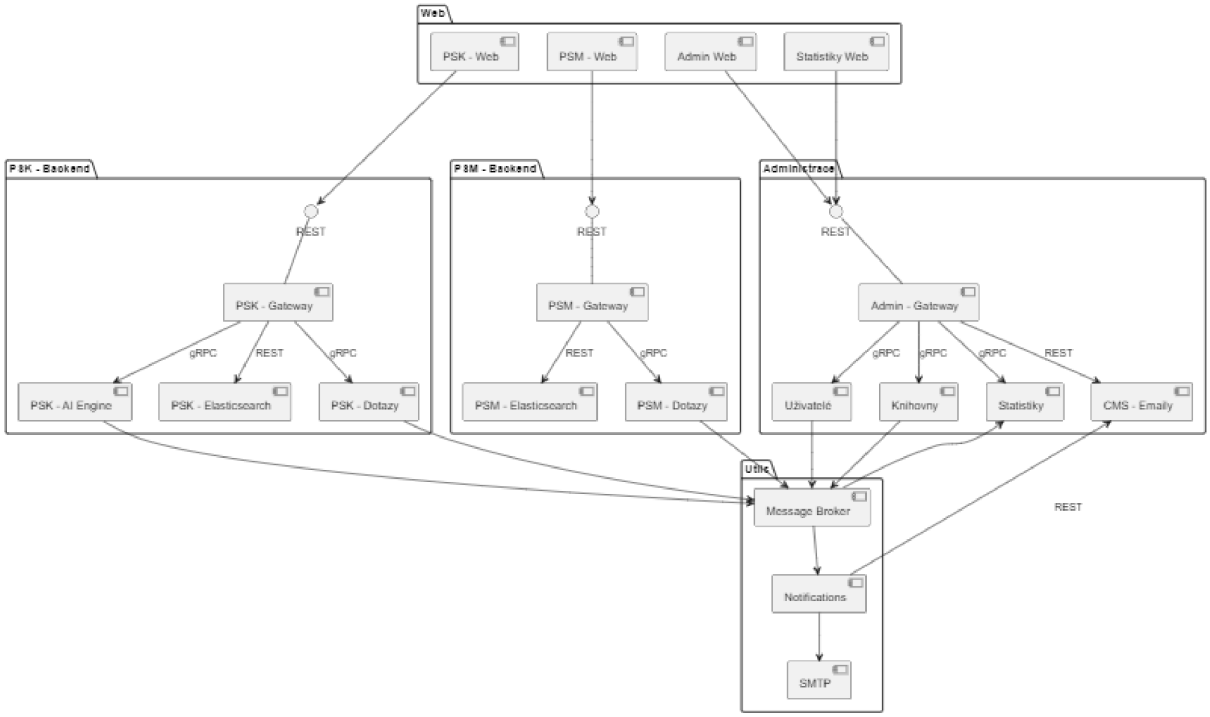
www.bootiq.io
info@bootiq.io
part of big.group

Úvod

Následující dokument reprezentuje návrh architektury, který vznikl na základě dodané technické specifikace a zároveň i z prvního dialogu, který proběhl 27. listopadu 2024. Tento koncept se bude měnit s upřesňovat na základě analýzy, která by vznikla před samotným započetím vývojových prací.

Architektura služeb

Aktuálně plánujeme provoz aplikace v cloudovém prostředí Azure od společnosti Microsoft. S tímto poskytovatelem máme již několikaleté zkušenosti. Mimo vlastních aplikací, které budou předmětem vývoje máme v plánu využívat i služby, které poskytuje tento cloudový provider. Jde například o databázový server. Tyto služby jde provozovat i přímo, ale při využití služby dojde k úspoře nákladů.



Obrázek 1 Diagram komponent

Naším cílem je rozdělit aplikaci do jednotlivých mikroslužeb, které budou mezi sebou komunikovat. Jednotlivé služby budou potom dostupné zvenčí skrze gateways.

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of big.group

Message broker

Message broker slouží pro předávání si informací mezi jednotlivými službami, jedná se například o propagaci využívání AI pro statistický modul nebo odesílání notifikací. Message broker zajistí odeslání informace tak, aby neblokovala odeslání informace uživateli. Díky tomu bude aplikace celkově rychlejší z uživatelského hlediska.

Elasticsearch

Aktuálně uvažuje o použití elasticsearch jako vyhledávácí engine, který umožňuje sémantické vyhledávání, ale je možnost, že řešení nakonec postavíme na úrovni relační DB např. Postgres + pgvector.

Vlastní mikroslužby

V aktuálním konceptu počítáme, jak již bylo řečeno, s rozdělením do mikroslužeb, jednotlivé mikroslužby budou mezi sebou komunikovat pomocí gRPC, pokud to bude technologicky možné.

AI

Aktuálně počítáme, že jednotlivé integrace AI vložíme do jedné mikroslužby, nicméně je možné, že dojde k dalšímu rozdělení na jednotlivé menší celky.

Speech-To-Text

Přepis hlasového vstupu do textové podoby. Obecně doporučujeme využít techniky přepisu po dokončení mluvy kvůli větší přesnosti a menších nákladech na běh a implementaci. Dokončení zpracování takovéto zprávy je zpravidla do pár sekund od konce mluvy.
Využijeme zde některé z řešení 3. strany dle analýzy kvality přepisu na poskytnutém vzorku dat:

- OpenAI Whisper
- Azure AI Speech

Bohužel pro češtinu nejsou žádné self-hosted modely produkční kvality.

Generátor obrázků

Jedná se o dvoukrokrovou AI pipeline – nejdříve pomocí jazykového modelu přetvoříme text na vhodný prompt, který pošleme do generativního obrazového modelu.
Jazykový model lze použít jakýkoliv, dobře ale ze zkušenosti funguje ChatGPT od verze 4.
Jelikož se jedná o obrázky pro nekomerční užití, máme k dispozici většinu existujících modelů, např.:

BOOTIQ
B!Q Group
BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle
www.bootiq.io
info@bootiq.io
part of big.group

- Flux.1
- Dall-E
- Midjourney
- Popř. libovolný model z CivitAI splňující licenční požadavky (vpravo dole na detailu)

Služba sémantické podobnosti

Vrátí podobnost dvou textů z hlediska jejich témat. Využijeme např. na přiřazení štítků a nabídku podobných dotazů. Na pozadí se jedná o jazykový model, který generuje embeddingy (n-dimenzionální vektory popisující daný prompt), a porovnává jednotkovou vzdálenost mezi vrácenými vektory.

Z existujících embeddingových modelů lze použít např.:

- OpenAI text-embedding-3-large
- Některý z modelů od Seznamu
- Jina.AI jina-embeddings-v3

RAG (Retrieval-augmented Generation)

Velmi velkou otázkou této komponenty je vyhledávání v online zdrojích. Budeme předpokládat, že se jedná o předem definované stránky s materiály, a ne fulltextové vyhledávání na celém webu, jelikož bychom tímto přišli o hodně možností, jak danou komponentu implementovat. Základem tohoto řešení je kombinace vektorové databáze, modelu na tvorbu embeddingů a generického velkého jazykového modelu (LLM).

Embeddingový model zvolíme dle výsledků služby sémantické podobnosti.

Vektorovou DB zvolíme dle formátu vstupních dat z:

- Azure Cosmos DB
- MongoDB
- PostgreSQL

Jazykový model lze použít prakticky jakýkoliv podporující češtinu v produkční kvalitě:

- OpenAI ChatGPT
- Anthropic Claude
- Google Gemini / Gemma2
- Qwen2.5

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of big.group

Existující i out-of-the-box RAG služby, zde bohužel ale nelze měnit AI model na pozadí, což je jeden z požadavků. Pokud by se požadavek změnil na přepínání celých RAG služeb, mohli bychom je použít a tím ušetřit čas na implementaci:

- Azure AI Search
- OpenAI SearchGPT (uzavřená beta, předpokládáme k dispozici v čase implementace)

Pozn.: při použití vyhledávání pomocí AI není potřeba uživateli našeptávat lepší formu psaní dotazu, jelikož vyhledávání funguje na sémantické bázi, ne pomocí klíčových slov v textu.
Pozn. 2: Jelikož primárním jazykem pro interakci se systémem bude čeština, jsme celkem omezení možnostmi výběru jazykových modelů. Aktuálně není lokálně nasaditelný model, který dokáže zpracovat český texty.

Dotazy

Jedná se o mikroslužbu, která má na správu pouze dotazy v rámci PSK nebo PSM. Uchovává jejich datovou strukturu a umožnuje vytváření a úpravu. Obsahuje také hodnocení a komentáře (toto se možná změní).

Administrace

Administrativní část celého řešení je rozdělena do jednotlivých služeb. Každá služba zajišťuje konkrétní funkcionalitu systému a dohromady tvoří celé řešení.

Statistiky

Tato část bude udržovat všechny statistické informace o aplikaci. Informace bude přijímat z jednotlivých systémů pomocí message brokera.

Uživatelé

Zahrnuje správu uživatelů a uchovávání jejich identifikačních údajů.

Knihovny

Obsahuje správu a databázi knihoven, které jsou zapojeny do PSK či PSM.

CMS – Emaily

Pro potřeby aplikace bude potřeba umožnit jednoduchou možnost úpravy emailů, k tomuto účelu bude připravené kompletní CMS pro emaily.

BOOTIQ
B!Q Group

BOOTIQ s. r. o.
Polygon House
Doudlebská 5
140 00 Praha 4 – Nusle

www.bootiq.io
info@bootiq.io
part of biq.group