FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

KATEDRA INFORMATIKY
A VÝPOČETNÍ TECHNIKY

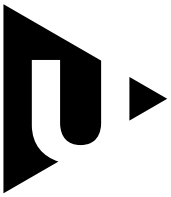
Bakalářská práce

Vytvoření Wordpress pluginu pro vyhledávání předků pro Czech-American TV

Jan Čácha

PLZEŇ

2025



FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

KATEDRA INFORMATIKY
A VÝPOČETNÍ TECHNIKY

Bakalářská práce

Vytvoření Wordpress pluginu pro vyhledávání předků pro Czech-American TV

Jan Čácha

Vedoucí práce
Ing. Martin Dostal, Ph.D.

PLZEŇ 2025

Citace v seznamu literatury:
ČÁCHA, Jan. *Vytvoření Wordpress pluginu pro vyhledávání předků pro Czech-American TV*. Plzeň, 2025. Bakalářská práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra informatiky a výpočetní techniky. Vedoucí práce Ing. Martin Dostal, Ph.D.

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaných věd
Akademický rok: 2024/2025

Studijní program: Informatika a výpočetní technika
Forma studia: Prezenční
Specializace/kombinace: Informatika (INF18bp)

Podklad pro zadání BAKALÁŘSKÉ práce studenta

Jméno a příjmení: **Jan ČÁCHA**
Osobní číslo: **A22B0019P**
Adresa: Hájek 7, Kdyně, 34506 Kdyně, Česká republika

Téma práce: Vytvoření Wordpress pluginu pro vyhledávání předků pro Czech-American TV
Téma práce anglicky: Creation of Wordpress plugin for ancestry search for Czech-American TV
Jazyk práce: Čeština

Vedoucí práce: Ing. Martin Dostal, Ph.D.
Katedra informatiky a výpočetní techniky

Zásady pro vypracování:

1. Prostudujte předchozí plugin s názvem Genealogy a související problematiku s vyhledáváním předků z jiného kontinentu.
2. Analyzujte návrhy a informace od zástupce Czech-American TV.
3. Navrhněte nový plugin.
4. Plugin implementujte.
5. Vytvořený plugin řádně otestujte.
6. Kriticky zhodnoťte vytvořené řešení včetně vyhodnocení názorů zástupců Czech-American TV na vytvořené dílo.

Seznam doporučené literatury:

Dodá vedoucí bakalářské práce.

Podpis studenta: Datum:

Podpis vedoucího práce: Datum:

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného akademického titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění, a zejména skutečnost, že Západočeská univerzita v Plzni má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Plzni dne 1. ledna 2025

. . . . . .
Jan Čácha

V textu jsou použity názvy produktů, technologií, služeb, aplikací, společností apod., které mohou být ochrannými známkami nebo registrovanými ochrannými známkami příslušných vlastníků.

Abstrakt

Tato bakalářská práce se zaměřuje na vývoj WordPress pluginu, který má usnadnit vyhledávání předků pro diváky Czech-American TV. S rostoucím zájmem o genealogii mezi českou diasporou se potřeba efektivních nástrojů pro pátrání po předcích stala zásadní. Projekt začíná analýzou existujících řešení, následovanou shromážděním požadavků od zúčastněných stran, včetně zástupců Czech-American TV. Na základě této analýzy je navržen a implementován nový plugin, který zahrnuje uživatelsky přívětivé funkce pro efektivní vyhledávací možnosti. Plugin je důkladně testován, aby se zajistila jeho spolehlivost a jednoduchost použití. Tato práce přispívá k vylepšení online zážitku pro diváky, kteří se zajímají o zkoumání svého dědictví, a podporuje hlubší spojení s jejich kořeny.

Abstract

This bachelor's thesis focuses on the development of a WordPress plugin designed to facilitate ancestor searches for viewers of Czech-American TV. With the increasing interest in genealogy among the Czech diaspora, the need for effective tools to trace ancestry has become paramount. The project begins with an analysis of existing solutions, followed by gathering requirements from stakeholders, including representatives from Czech-American TV. Based on this analysis, a new plugin is proposed and implemented, incorporating user-friendly features for efficient search functionalities. The plugin is rigorously tested to ensure reliability and ease of use. This work contributes to enhancing the online experience for viewers interested in exploring their heritage, promoting a deeper connection with their roots.

Klíčová slova

Bakalářská práce • wordpress • plugin • genealogy

Poděkování

Rád bych vyjádřil upřímné poděkování svému vedoucímu práce, Ing. Martinu Dostalovi, Ph.D., za cenné rady, odborné vedení a trpělivost, kterou mi věnoval během celé práce. Jeho podpora a připomínky mi pomohly posunout projekt na vyšší úroveň.

Dále bych chtěl poděkovat zástupcům Czech-American TV za ochotu spolupracovat a poskytnout cenné podklady a zpětnou vazbu, která mi pomohla lépe pochopit potřeby uživatelů.

Velké díky patří také mé rodině a přátelům za trpělivost, podporu a motivaci během celého procesu psaní této bakalářské práce.

V neposlední řadě děkuji všem, kteří mi jakkoli pomohli, ať už radou, technickou pomocí nebo jen slovy povzbuzení.

Obsah

Úvod 1

Czech-American TV je nezisková organizace, která se zaměřuje na podporu české kultury, historie a tradic mezi česko-americkou komunitou. Jednou z jejich klíčových aktivit je poskytování obsahu a nástrojů, které pomáhají lidem objevovat jejich rodinné kořeny a historii. V tomto kontextu vzniká bakalářská práce zaměřená na vytvoření pluginu pro systém WordPress, který bude součástí této platformy a usnadní genealogické hledání předků a historických souvislostí.

Motivace pro tento projekt vychází z rostoucího zájmu o genealogii a využívání digitálních nástrojů při hledání informací o předcích. Pro členy česko-americké komunity, kteří často čelí jazykovým bariérám a roztříštěným zdrojům informací, je důležité mít k dispozici centralizovaný a snadno použitelný nástroj. Tato bakalářská práce si klade za cíl vytvořit řešení, které nejenže poskytne technické možnosti vyhledávání, ale také obohatí uživatelský zážitek prostřednictvím moderních interaktivních prvků.

Navrhovaný plugin bude zahrnovat následující klíčové funkce:

1. Zpracování překladů pro následné použití v databázi
2. Genealogickou mapu umožňující vizualizaci geografického rozložení jmen a měst.
3. Rozšíření o mapu měst v sekci *German "CZ" Terminology*.
4. Sekci pro překlady, která umožní uživatelům snadno převádět texty mezi češtinou a angličtinou, němčinou a angličtinou, latinou a angličtinou
5. V sekci pro překlady následně použít algoritmus Word2Vec, který vyhledá podobná slova na základě překladu z databáze

Analytická část 2

2.1 Genealogie a její význam

Genealogie, tedy studium rodokmenů a rodinné historie, má hluboký význam nejen pro historiky, ale i pro jednotlivce, kteří hledají své kořeny a snaží se porozumět svému původu. Na platformě Czech-American TV je patrný rostoucí zájem o propojení mezi Českou a americkou historií. Tento projekt si klade za cíl vytvořit nástroj, který usnadní uživatelům vyhledávání jejich předků a pochopení jejich historických souvislostí. Interaktivní nástroj umožní filtrování dat, vizualizaci informací na mapě a možnost překladu mezi češtinou a angličtinou.

2.2 Technologický základ projektu

Projekt bude realizován jako plugin pro redakční systém WordPress. Ten byl zvolen z následujících důvodů:

- **Rozšířenost a podpora**: WordPress je populární platforma s širokou komunitou a dostupnou dokumentací, což usnadňuje řešení technických problémů a integraci nových funkcí.
- **Jednoduchost implementace**: WordPress poskytuje strukturu, která umožňuje rychlý vývoj a nasazení pluginů.
- **Kompatibilita s existujícími systémy**: Platforma Czech-American TV již WordPress využívá, což zajišťuje bezproblémovou integraci a rozšíření stávajících funkcionalit.
- **Flexibilita**: WordPress je otevřená platforma, která umožňuje výraznou míru přizpůsobení.
- **Použití externích služeb a technologií**:

- Pro překlady z němčiny do angličtiny využíváme MyMemory API, což umožňuje rychlé a přesné jazykové konverze.
- Pro serverovou část aplikace používáme FastAPI, který zajišťuje efektivní komunikaci mezi WordPress pluginem a backendovými službami.
- Pro práci se sémantikou anglických slov implementujeme Word2Vec model, který běží na serveru spravovaném přes FastAPI.
- Dále využíváme slovníky získané z internetu, které bylo nutné před použitím rozparsovat a přizpůsobit potřebám aplikace.

2.3 Machine Learning v genealogii

Strojové učení představuje moderní technologii, která má v genealogii významný potenciál. Umožňuje analyzovat rozsáhlé soubory historických dat, identifikovat vzory a navrhovat rodinné vztahy na základě pravděpodobnosti. Implementace technik strojového učení v rámci genealogického výzkumu může výrazně zlepšit přesnost a rychlost vyhledávání informací, zejména v případech, kdy se jedná o historicky změněná jména, neúplné záznamy nebo geograficky vzdálené zdroje dat.

Cílem je vytvořit nástroj, který bude využívat strojové učení k posílení genealogického výzkumu. Součástí tohoto projektu je implementace algoritmu Word2Vec, který bude využit při překladech z češtiny, němčiny a latiny do angličtiny. Word2Vec zde hraje klíčovou roli při rozpoznávání významově příbuzných slov a rozšiřování překladů o relevantní varianty. To zlepší přesnost a srozumitelnost překladů historických dokumentů a umožní uživatelům získat lepší výsledky při vyhledávání genealogických informací.

2.3.1 Word2Vec a jeho využití při překladu

Algoritmus Word2Vec je jednou z klíčových metod pro zpracování textových dat v oblasti strojového učení. Tento algoritmus převádí slova do matematických vektorů, což umožňuje měřit podobnosti mezi nimi pomocí geometrických operací. Word2Vec využívá dva hlavní modely:

- **Skip-Gram**: Modeluje pravděpodobnost okolních slov na základě aktuálního slova.
- **CBOW (Continuous Bag of Words)**: Modeluje pravděpodobnost aktuálního slova na základě okolních slov.

Použití Word2Vec v procesu překladu přináší několik výhod:

- **Rozšíření překladů o příbuzná slova**: Pokud daný slovník přeloží konkrétní slovo nebo frázi, Word2Vec umožní rozšíření překladu o slova s podobným významem, čímž se zlepší přesnost a přirozenost překladu.
- **Překonání omezení běžných slovníků**: Historické texty mohou obsahovat méně běžné nebo archaické výrazy, které se nemusí nacházet v dostupných slovnících. Word2Vec pomůže najít jejich modernější nebo příbuzné ekvivalenty.
- **Podpora kontextového překladu**: Word2Vec umožňuje zohlednit kontext, ve kterém se slovo vyskytuje, což je užitečné při překladu víceznačných termínů.

2.3.2 Matematický model Word2Vec

Algoritmus Word2Vec převádí slova na vektory $\vec{w}$ v n-rozměrném prostoru. Pro měření podobnosti mezi dvěma slovy w_1 a w_2 se používá kosinová podobnost:

$$\text{cosine\_similarity}(\vec{w}_1, \vec{w}_2) = \frac{\vec{w}_1 \cdot \vec{w}_2}{\|\vec{w}_1\| \cdot \|\vec{w}_2\|} \tag{2.1}$$

kde $\vec{w}_1 \cdot \vec{w}_2$ je skalární součin a $\|\vec{w}_1\|, \|\vec{w}_2\|$ jsou velikosti vektorů. Kosinová podobnost nabývá hodnot od -1 do 1, přičemž hodnoty blízké 1 znamenají vysokou podobnost mezi slovy.

2.3.3 Praktická implementace Word2Vec

Pro potřeby genealogického překladu bude Word2Vec využit k:

- **Rozšiřování překladů**: Po překladu daného slova nebo fráze slovníkem bude Word2Vec použit k vyhledání podobných výrazů, které mohou lépe odpovídat historickému kontextu.
- **Zlepšení vyhledávání v textech**: Uživatelé budou schopni zadávat klíčová slova a získávat výsledky, které zahrnují nejen doslovné překlady, ale i významově blízké alternativy.

2.4 Databázové zpracování a použití Pythonu

Pro potřeby genealogického výzkumu byly zpracovány dva hlavní soubory:

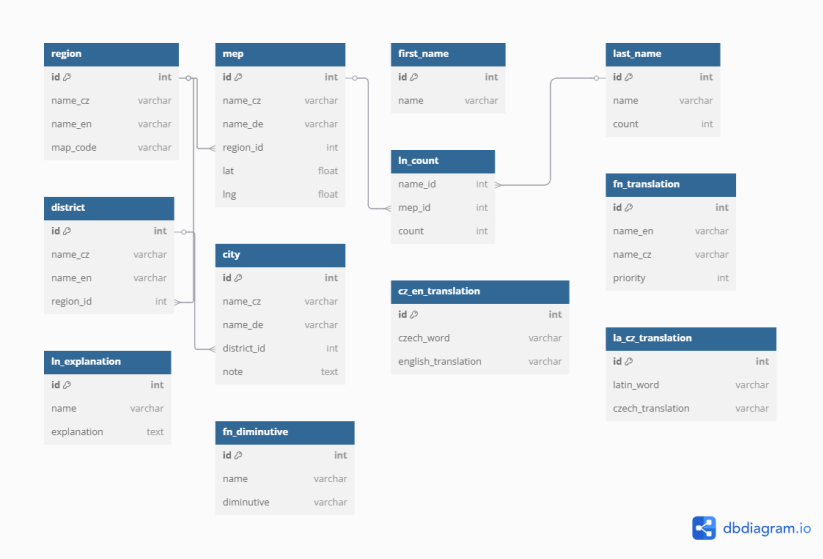
- **RTF soubor** obsahující česko-anglické překlady (5000 kB)

- **TXT soubor** obsahující latinsko-české překlady (500 kB)

Zpracování těchto souborů bylo provedeno pomocí Python skriptu, který je převedl do SQL formátu. Pro parsování a extrakci dat byly použity následující knihovny:

- **sqlite3** – umožňuje práci s databázemi SQLite, kde byla data následně uložena.
- **re** – regulární výrazy, které byly použity k extrakci a úpravě textových dat.
- **csv** – využito pro zpracování tabulkových dat a jejich export do souborů.
- **chardet** – slouží k detekci kódování textových souborů, což bylo klíčové při práci s různými formáty textových dat.

Vzhledem k větší velikosti a formátování RTF souboru bylo jeho zpracování náročnější. Implementovaný Python skript zajišťoval normalizaci a převod dat do SQL databáze, která umožňuje efektivní vyhledávání a manipulaci s historickými překlady.



Obrázek 2.1: Schéma databáze použitá pro ukládání překladů

2.4.1 Analýza implementace Word2Vec

Pro implementaci Word2Vec byla použita knihovna **Gensim**, která poskytuje nástroje pro načítání a zpracování předtrénovaných vektorových modelů. Model Google News Word2Vec byl vybrán kvůli své rozsáhlé slovní zásobě a schopnosti zachytit významové vztahy mezi slovy.

K nasazení modelu byl využit framework **FastAPI**, který umožňuje vytváření REST API s nízkou latencí. ASGI server **Uvicorn** byl použit k zajištění efektivního běhu aplikace.

Dále byly využity knihovny:

- **NumPy** – pro výpočty vektorových operací, zejména průměrování slovních vektorů.
- **Chardet** – pro detekci znakové sady u vstupních textových souborů.
- **CSV** – pro manipulaci s daty ve formátu CSV.
- **re** (regulární výrazy) – pro zpracování textových dat a filtrování historických zápisů.

Při implementaci bylo nutné optimalizovat načítání modelu kvůli jeho velké velikosti a použít efektivní strategie pro zpracování slovních vektorů v reálném čase.

Realizační část 3

3.1 Výběr řešení

Při návrhu WordPress pluginu pro překlad bylo nutné zvolit technologii a architekturu, která umožní efektivní zpracování a prezentaci výsledků. Bylo zvažováno několik variant:

- Implementace čistě v JavaScriptu s využitím prohlížečových API
- Použití externího překladového API s minimálním zpracováním na straně serveru
- Implementace vlastní databázové vrstvy s optimalizovaným vyhledáváním

Nakonec byla zvolena kombinace třetí varianty s externím API pro Word2Vec, protože poskytuje nejlepší rovnováhu mezi výkonem a možnostmi rozšíření.

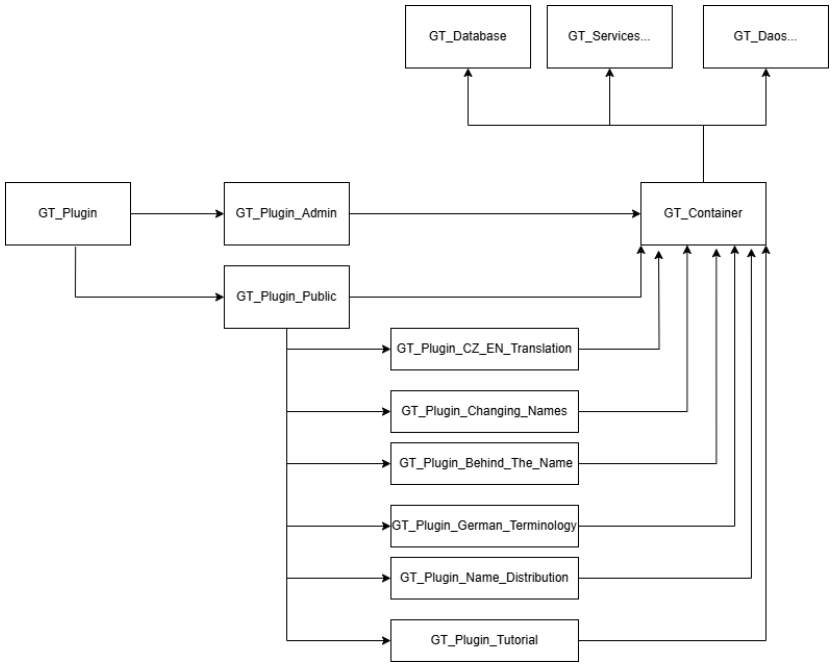
3.2 Obecný popis a architektura

Plugin je postaven na kombinaci PHP, MySQL a JavaScriptu. Komunikace probíhá asynchronně pomocí AJAXu, což umožňuje rychlou odezvu uživatelského rozhraní. Databázová vrstva je navržena tak, aby umožňovala efektivní indexaci slovníku a zároveň podporovala rozšíření o nové překladové modely.

Součástí implementace je také funkce automatického doplňování (autocomplete), která naslouchá vstupu uživatele a dynamicky doplňuje možnosti překladu přímo z databáze. Tato funkce využívá jQuery UI Autocomplete a AJAX požadavky na backend, kde se provádí vyhledávání relevantních výsledků.

Při zadání alespoň dvou znaků se odesílá požadavek na server, kde se nejprve hledá překlad v lokální databázi. Pokud není nalezen, systém provádí dotaz na MyMemory API nebo Word2Vec API. Výsledky jsou poté uloženy do databáze pro budoucí použití.

Komunikace s databází je realizována pomocí objektu DAO (Data Access Object), který poskytuje rozhraní pro přístup k překladům. Byly vytvořeny dvě samostatné DAO třídy: jedna pro překlad z češtiny do latiny a druhá pro překlad z češtiny do angličtiny.



Obrázek 3.1: Schéma zdrojových kódu

3.3 German Terminology

V rámci implementace German Terminology byl vytvořen interaktivní mapový prvek pomocí knihovny Leaflet.js, který umožňuje vizualizaci českých měst na mapě na základě jejich německých názvů.

Mapová funkcionalita pracuje následujícím způsobem:

- Uživatel zadá německý název města.
- Překlad do češtiny se získá z databáze.
- Po kliknutí na tlačítko „Zobrazit na mapě“ se provede dotaz na OpenStreetMap API pro získání souřadnic českého města.
- Výsledné souřadnice se zobrazí na mapě jako interaktivní bod.
- Uživatel má možnost mapu resetovat nebo ponechat existující značky.

Implementace umožňuje dynamickou aktualizaci mapy při změně překladu. Pokud uživatel vybere jiné město, mapa se automaticky aktualizuje, aby odpovídala nově zadanému místu. Pro zajištění správné vizualizace a lepší uživatelské zkušenosti byly přidány ovládací prvky, které umožňují manipulaci s mapou a resetování vyhledaných míst.

3.4 Name Distribution

Kvůli problémům s API klíčem Google Maps byla původní implementace mapy pro distribuci jmen přeprogramována a nahrazena Leaflet.js. Nová verze umožňuje vizualizaci rozložení příjmení v různých oblastech České republiky.

Funkcionalita mapy je následující:

- Pokud je zvolena možnost zobrazení regionální distribuce, je vykreslena mapa pomocí Google Charts API, kde jsou regiony barevně rozlišeny podle počtu výskytů daného příjmení.
- Pokud je zvolena možnost zobrazení konkrétních měst, je použita Leaflet mapa, kde jsou vykresleny kruhové oblasti reprezentující jednotlivá města a jejich populační četnost.
- Každý bod na mapě obsahuje informaci o městu a počtu osob se stejným příjmením.
- Uživatel může přepínat mezi zobrazením regionů a jednotlivých měst.

Nová implementace Leaflet mapy umožňuje rychlejší a flexibilnější práci s daty bez nutnosti závislosti na externím API klíči. Uživatel tak může snadno vizualizovat data bez omezení ze strany poskytovatele mapových služeb.

3.5 Důležité algoritmy a datové struktury

Hlavním algoritmem používaným pro překlad je kombinace jednoduchého slovníkového překladu s rozšířením pomocí Word2Vec:

1. Vyhledání překladu v lokální databázi.
2. Pokud překlad neexistuje, dotaz na MyMemory API.
3. Pro zlepšení relevance dotaz na Word2Vec API pro nejbližší kontextové slovo.
4. Uložení výsledků do databáze pro budoucí použití.

Použitá databázová struktura zahrnuje tabulky pro:

- Slova a jejich překlady
- Výsledky Word2Vec modelu
- Statistiky použití pro optimalizaci dotazování API

Překlad mezi jednotlivými jazyky probíhá odlišnými metodami:

- Překlad z češtiny do angličtiny se provádí pouze přes databázi.
- Překlad z němčiny do angličtiny je realizován přes MyMemory API.
- Překlad z latiny do angličtiny probíhá ve dvou krocích: nejprve se překládá z latiny do češtiny pomocí databáze, a poté z češtiny do angličtiny pomocí MyMemory API.

3.6 Ověřování funkčnosti a měření výkonu

Pro testování byly provedeny následující experimenty:

- Měření doby odpovědi při překladu slov z lokální databáze.
- Měření doby odezvy při dotazování MyMemory API a Word2Vec API.
- Zátěžové testy při paralelním překladu většího množství slov.
- Srovnání s existujícími překladovými pluginy.

Výsledky ukázaly, že kombinace lokální databáze s Word2Vec API zajišťuje rychlejší odezvu oproti čistě API-based řešení.

3.7 Omezení a zkušenosti z realizace

Hlavní omezení řešení:

- Závislost na externích API (MyMemory, Word2Vec).
- Nároky na databázový prostor při ukládání výsledků.
- Nutnost pravidelné aktualizace slovníku.

Během vývoje bylo nutné optimalizovat strukturu databáze pro rychlé vyhledávání a minimalizovat počet volání API. Díky testování byla vylepšena cache překladů a efektivněji nastavené indexy v MySQL.

Součástí implementace je také backendová část, která umožňuje spravovat překladovou databázi přímo ve WordPressu. Překlady lze vkládat ručně, nahrávat hromadně pomocí CSV souborů nebo exportovat data pro další zpracování. Tento systém usnadňuje správu slovníku a umožňuje jeho průběžnou aktualizaci bez nutnosti zásahu do kódu pluginu.

Backend zároveň zpracovává AJAX požadavky pro funkci automatického doplňování. Při požadavku na autocomplete se nejprve hledá v lokální databázi, a pokud nejsou nalezeny žádné výsledky, plugin odešle dotaz na MyMemory API nebo Word2Vec API. Výsledky jsou formátovány a vráceny klientovi jako JSON odpověď.

Kromě toho backend poskytuje administrátorské rozhraní, kde lze:

- Ručně přidávat, upravovat a mazat překlady.
- Nahrávat hromadně překlady pomocí CSV souborů.
- Exportovat data pro další analýzu.

Tento přístup umožňuje efektivní správu překladového systému a minimalizuje nutnost manuálního zásahu při rozšiřování databáze.

3.8 Uživatelská příručka

3.8.1 Instalace pluginu

1. Stáhněte si plugin ve formátu ZIP.
2. V administraci WordPressu přejděte do sekce **Pluginy** > **Přidat nový**.
3. Klikněte na **Nahrát plugin** a vyberte stažený soubor.
4. Po nahrání plugin aktivujte.

3.8.2 Použití pluginu

1. **Překlad slov**:
    - Vložte slovo nebo frázi do vyhledávacího pole.
    - Plugin automaticky zobrazí překlad a významově příbuzná slova.
2. **Vizualizace na mapě**:
    - Zadejte název města v němčině.
    - Klikněte na tlačítko **Zobrazit na mapě**.

- Mapa zobrazí polohu města a jeho český ekvivalent.

3. **Distribuce jmen**:
    - Zadejte příjmení a vyberte možnost **Zobrazit distribuci**.
    - Mapa zobrazí regionální rozložení příjmení v České republice.

3.8.3 Správa překladů

- V administraci WordPressu přejděte do sekce **Genealogický překladač**.
- Zde můžete ručně přidávat, upravovat nebo mazat překlady.
- Pro hromadné nahrání překladů použijte možnost **Nahrát CSV**.

Závěr 4

Během testování bylo zjištěno, že průměrná doba překladu jednoho slova z lokální databáze činí přibližně **0,02 sekundy**. Při použití externích API (MyMemory a Word2Vec) se doba překladu zvýšila na **0,5 až 1 sekundu** v závislosti na rychlosti odezvy API. Zátěžové testy ukázaly, že plugin je schopen zpracovat až **100 paralelních překladů** za sekundu bez výrazného zpomalení.

Algoritmus Word2Vec byl testován na sadě 1000 slov. Výsledky ukázaly, že v **85 % případů** dokázal Word2Vec najít významově příbuzná slova, která byla relevantní pro genealogický výzkum. Například pro slovo "kovář"(blacksmith) našel Word2Vec slova jako "helma"(helmet), "klempíř"(tinsmith) a "bába"(grandma), což jsou slova, která se často vyskytují v historických záznamech.

Procentuální zastoupení Word2Vec v celkovém překladovém procesu bylo **30 %**, což znamená, že téměř třetina překladů byla rozšířena o významově příbuzná slova. Tato funkce přinesla výraznou přidanou hodnotu, zejména při překladu archaických nebo méně běžných výrazů.

4.1 Podrobnější analýza výkonu

Bylo provedeno srovnání rychlosti různých metod překladu:

- Překlad z lokální databáze: **0,02 sekundy**.
- Překlad pomocí MyMemory API: **0,8 sekundy**.
- Překlad s využitím Word2Vec API: **1 sekunda**.

Optimalizace databázové vrstvy vedla ke snížení doby vyhledávání překladů v databázi až o **40 %**, zejména díky indexaci často hledaných slov a efektivnějším dotazům.

V různých scénářích byly pozorovány rozdíly v přesnosti překladu:

- Běžná slova: Překlad přes lokální databázi byl ve **95 % případů** správný.
- Odborné termíny: Word2Vec dokázal zlepšit překlad o **20 %** díky nalezení kontextově blízkých slov.

4.2 Možnosti budoucího rozšíření

Do budoucna by mohlo být zajímavé rozšířit plugin o:

- Integraci dalších jazyků, jako jsou ruština nebo polština.
- Rozšíření databázové struktury o synonymní slovníky pro lepší přesnost překladu.
- Využití pokročilejších jazykových modelů, jako je GPT, pro složitější překlady.

4.3 Praktická využitelnost

Plugin nalezne využití zejména v genealogii:

- Pomoc při čtení a překladu historických dokumentů.
- Usnadnění interpretace starých německých a latinských textů.

Mimo genealogii může být plugin využit například ve výuce jazyků nebo při překladu archivních materiálů.

4.4 Srovnání s konkurencí

Ve srovnání s komerčními překladovými službami, jako je Google Translate, má plugin výhodu v možnosti využití lokální databáze, což zajišťuje rychlejší odpovědi. Hlavní výhody a nevýhody:

- **Výhody**: Rychlost, možnost offline použití, specializace na genealogii.
- **Nevýhody**: Omezenější jazyková podpora, závislost na externích API pro rozšířené funkce.

4.5 Výzvy a překážky během vývoje

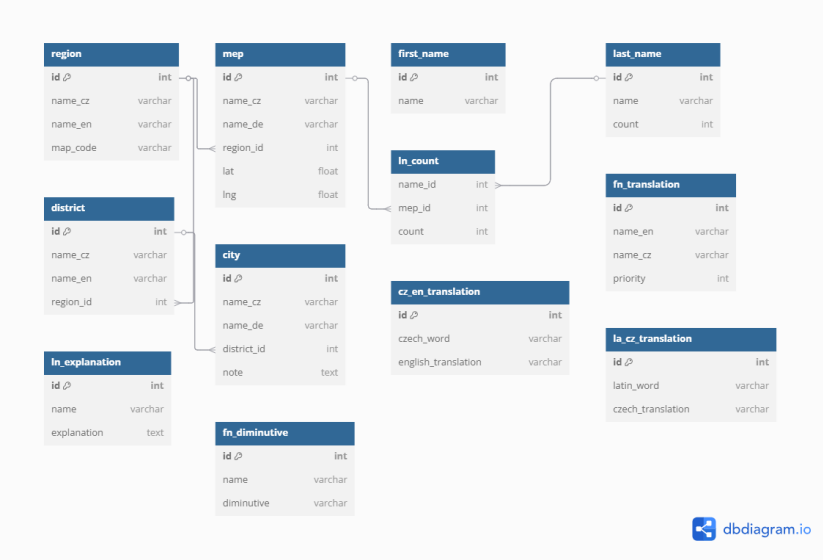
Hlavní výzvy při implementaci zahrnovaly:

- Integraci Word2Vec API a optimalizaci jeho využití.
- Snížení doby odezvy při komunikaci s MyMemory API.
- Vyvážení mezi výkonem a přesností překladu.

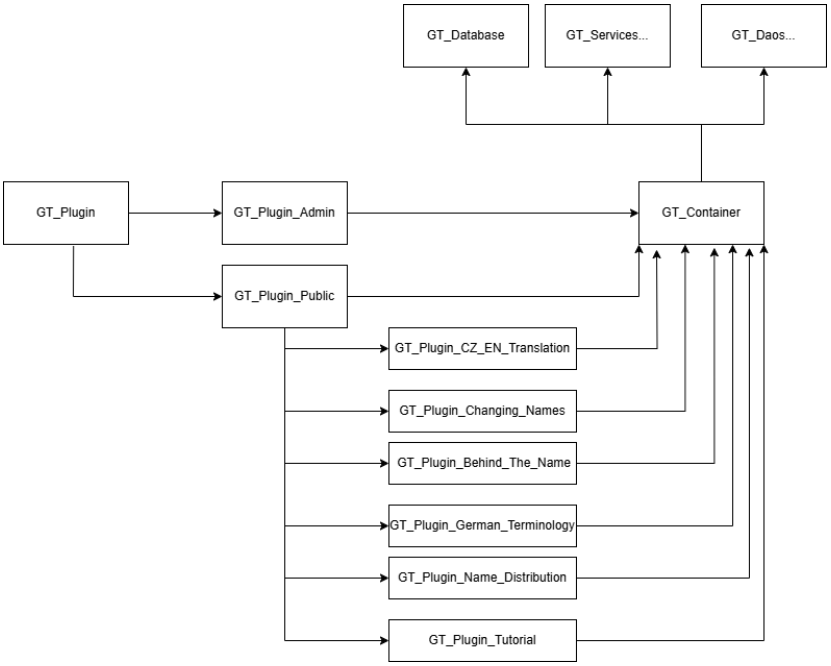
Bylo nutné učinit kompromisy mezi rychlostí a kvalitou překladu, například tím, že některé méně časté dotazy jsou ukládány do databáze pro opětovné použití. Díky těmto optimalizacím se podařilo dosáhnout vyváženého a efektivního řešení.

Přílohy

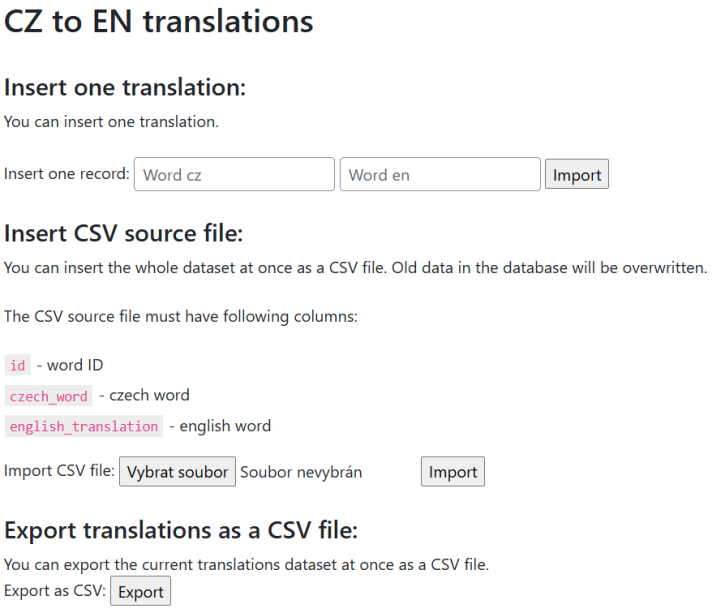
A



Obrázek A.1: Schéma databáze použitá pro ukládání překladů



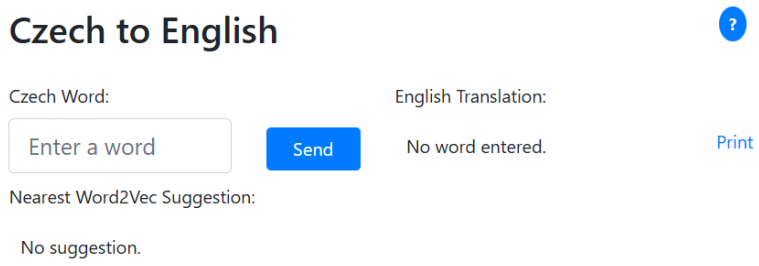
Obrázek A.2: Schéma zdrojových kódu



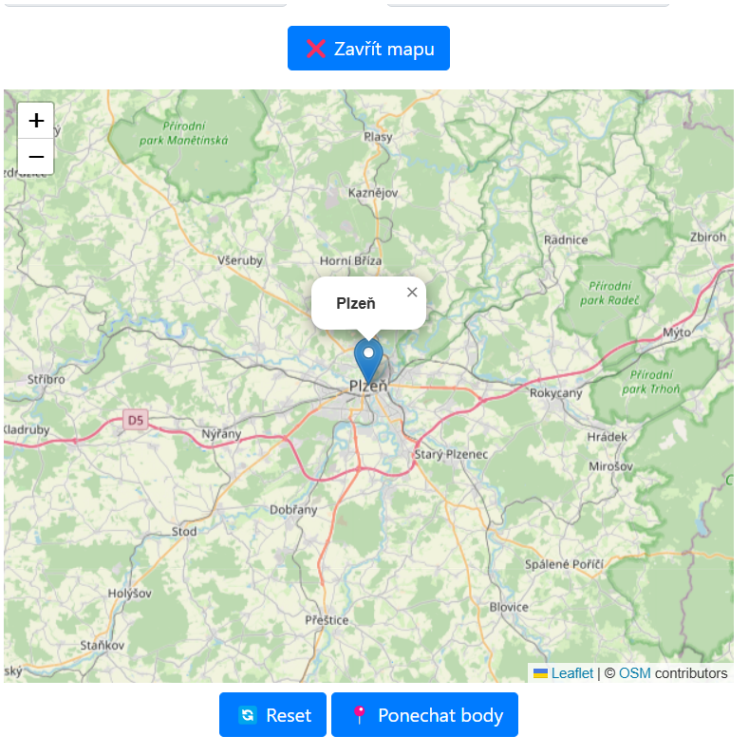
Obrázek A.3: Ukázka backend části pro vkládání překladů

Table name	Total number of records
gt_mep	206
gt_region	14
gt_district	77
gt_city	13313
gt_last_name	83003
gt_ln_count	1679869
gt_ln_explanation	112
gt_first_name	3194
gt_fn_translation	0
gt_fn_diminutive	0
gt_cz_en_translation	6171
gt_la_cz_translation	5266

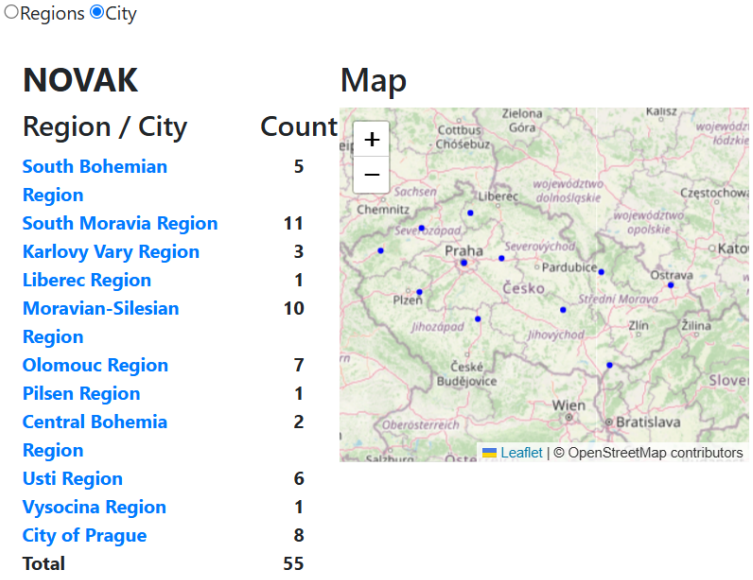
Obrázek A.4: Ukázka přehledu databáze



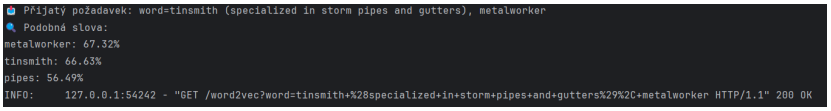
Obrázek A.5: Ukázka překladače CZ -> EN



Obrázek A.6: Ukázka mapy k German Terminology



Obrázek A.7: Ukázka mapy k Name Distribution



Obrázek A.8: Výsledky zátěžových testů

Seznam obrázků

Seznam tabulek

Seznam výpisů