

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Vytvoření Wordpress pluginu pro vyhledávání předků pro Czech-American TV

Originální zadání

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.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne _____

Podpis _____

Abstract

Creating a WordPress Plugin for Ancestry Search for Czech-American TV

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.

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.

Obsah

1. Úvod	1
2. Analytická část	2
2.1. Genealogie a její význam	2
2.2. Technologický základ projektu	2
2.3. Machine Learning v genealogii	3
2.4. Analýza dat a interaktivní funkce	3
2.5. Interaktivní seznamy a filtry	4
2.6. Vizualizace a překladové funkce	4
2.7. Zhodnocení stávajících řešení a doporučení	5

1. Úvod

V dnešní digitální době se stále více lidí zajímá o své kořeny a genealogii. S rostoucí globalizací a migrací se zvyšuje potřeba efektivních nástrojů, které umožňují vyhledávat a analyzovat historická jména a místa. Tato bakalářská práce si klade za cíl rozšířit a vylepšit stávající platformu zaměřenou na genealogické výzkumy a jména měst v Německu.

Jedním z hlavních cílů této práce je vytvoření interaktivního seznamu německých měst, který bude obohacen o vyhledávací funkce a filtry podle regionu a historického období. Tímto způsobem se uživatelům usnadní přístup k informacím o městech, což jim umožní rychleji nalézt relevantní údaje.

Dále se zaměříme na vizualizaci měst, kde se daná jména nachází, pomocí interaktivní mapy. Tento prvek nejenže zlepší uživatelský zážitek, ale také poskytne uživatelům lepší představu o geografickém rozložení měst a jejich historických kontextech.

V rámci práce bude také přidán filtr pro pohlaví a region v sekci „Changing Name“, což umožní uživatelům filtrovat jména podle jejich původu a genderu. Tento prvek přispěje k větší přehlednosti a použitelnosti databáze, která se zaměřuje na genealogická jména.

Dalším klíčovým bodem je přeprogramování stávající genealogické mapy tak, aby efektivně vyhledávala podle měst. Tento prvek je zásadní pro zajištění správné funkčnosti a spolehlivosti aplikace, která slouží uživatelům k objevování jejich rodinných kořenů.

V neposlední řadě plánujeme vytvoření nové sekce „Translation“, která umožní překlad textů z angličtiny do češtiny. Tato funkce bude mít významný přínos pro uživatele, kteří hledají informace v anglickém jazyce a potřebují je převést do češtiny, čímž se rozšíří dostupnost informací pro širší publikum.

Tato bakalářská práce se tedy zaměřuje na zlepšení stávajících funkcionalit a přidání nových, které usnadní uživatelům přístup k historickým a genealogickým informacím. Výsledkem bude komplexní a uživatelsky přívětivý nástroj, který podpoří zájem o genealogii a historická jména.

2. Analytická část

2.1. Genealogie a její význam

Genealogie, tedy studium rodokmenů a rodinné historie, se v posledních letech stala populární nejen mezi historiky, ale také mezi běžnými uživateli, kteří touží po nalezení svých předků a porozumění svému původu. V rámci platformy Czech-American TV je zájem uživatelů o hledání předků a historických vazeb mezi Českou republikou a Spojenými státy obzvlášť silný, a proto je vytvoření interaktivního nástroje pro vyhledávání předků zásadní.

Cíl projektu je umožnit uživatelům intuitivní vyhledávání v databázích českých a amerických předků, což zahrnuje funkce jako filtrování jmen podle genderu a regionu, vizualizaci měst na mapě a možnost překladu.

2.2. Technologický základ projektu

Projekt bude realizován formou pluginu pro redakční systém WordPress, což je populární platforma pro tvorbu a správu webových stránek díky své otevřenosti, široké komunitě a dostupnosti mnoha rozšíření. Plugin bude propojen s databází genealogických dat a využije moderní webové technologie, jako jsou PHP, JavaScript, a AJAX pro dynamické načítání dat.

2.2.1 Proč WordPress?

WordPress byl zvolen kvůli několika faktorům:

- **Rozšířenost a podpora:** Velká komunita a množství dostupné dokumentace.
- **Jednoduchost vývoje pluginů:** Umožňuje rychlou implementaci uživatelských funkcí.
- **Integrace s existujícími weby:** Czech-American TV již využívá WordPress jako CMS, což usnadňuje integraci nových funkcí.

2.3. Machine Learning v genealogii

Strojové učení má v genealogii velký potenciál, především při analyzování historických dat, rozpoznávání vzorců a predikci možných rodinných vztahů. Jednou z technologií, kterou se v tomto projektu chystám implementovat, je Word2Vec. Tato technologie bude použita pro zlepšení vyhledávání jmen v databázích na základě kontextu, tedy například nalezení příbuzných osob i v případě, že jména prošla historickou změnou (např. z Novák na Nowak).

2.3.1 Word2Vec v praxi

Word2Vec je algoritmus pro analýzu textu, který umožňuje převést slova do matematických vektorů. Tento přístup bude užitečný pro vyhledávání historických jmen, kde může docházet ke změnám ve výslovnosti, pravopisu či překladu. Použití Word2Vec bude přínosné při:

- **Rozpoznávání podobných jmen:** Algoritmus dokáže navrhnout alternativní tvary jmen (např. Novák -> Nowak).
- **Klasifikace jmen podle původu:** Pomocí vektorové reprezentace lze klasifikovat jména podle jejich geografického původu nebo významu.

2.4. Analýza dat a interaktivní funkce

2.4.1 Zpracování RTF souboru

Jedním z počátečních kroků bylo zpracování velkého RTF souboru obsahujícího genealogická data. Tento soubor byl skriptem převeden do strukturované podoby a importován do databáze. Nyní tato data umožňují rychlé a efektivní vyhledávání podle různých kritérií.

2.4.2. Databáze a její struktura

Data jsou uložena v relační databázi, která je navržena tak, aby umožňovala rychlé filtrování a třídění informací o městech, jménech, pohlaví a původu. Klíčové tabulky zahrnují:

- **Tabulka měst:** Obsahuje informace o historických městech a jejich lokalizaci.

- **Tabulka jmen:** Obsahuje jména, jejich původ, gender a další doplňkové údaje.
- **Tabulka regionů:** Slouží k filtrování podle geografických oblastí.

2.5. Interaktivní seznamy a filtry

2.5.1. Filtr pro města a období

Uživatelé budou moci filtrovat města podle historického období či regionu. To umožní například vyhledat města, kde jejich předci žili v určitém časovém období. K tomu využijeme JavaScript a AJAX pro dynamické načítání dat, což zaručí rychlou odezvu systému.

2.5.2. Filtrování jmen podle původu a genderu

Další interaktivní funkcí bude možnost filtrování jmen podle genderu a původu. To uživatelům umožní zaměřit se na konkrétní skupiny jmen, což zlepší přesnost a efektivitu vyhledávání.

2.6. Vizualizace a překladové funkce

2.6.1. Genealogická mapa

Pro zlepšení uživatelské zkušenosti bude vytvořena interaktivní mapa, kde budou vizualizována města spojená s předky uživatele. Tato mapa bude implementována pomocí nástroje Leaflet, který umožňuje jednoduchou a flexibilní práci s geografickými daty. Uživatelé budou moci kliknout na město na mapě a získat další informace o obyvatelích či historických událostech.

2.6.2. Překladové funkce

Pro zajištění lepší dostupnosti informací pro široké publikum bude implementována překladová sekce, která umožní uživatelům překládat genealogická data mezi češtinou a angličtinou. Překlad bude realizován pomocí API nástroje Google Translate, což zjednoduší proces a zajistí spolehlivé překlady.

2.7. Zhodnocení stávajících řešení a doporučení

Existuje několik populárních genealogických platforem, jako jsou Ancestry.com či MyHeritage, které poskytují robustní nástroje pro vytváření rodokmenů a vyhledávání předků. Nicméně tyto platformy často neobsahují detailní data pro specifické oblasti, jako je česká diaspora. Implementace tohoto WordPress pluginu rozšíří možnosti genealogického výzkumu pro české a americké uživatele.