

# Hodnocení oponenta bakalářské práce

Autor práce: Daniel Stuš  
Název práce: On-line slovník nářečí  
Vedoucí práce: Ing. Roman Mouček, Ph.D. (KIV)

Hodnocena je bakalářská práce, která měla být přepracována po předchozí neúspěšné obhajobě.

## Obsah práce

Cílem byla realizace slovníku nářečí vycházejícího ze slovníku vytvořeného na ZČU KIV před cca 16 roky. Součástí mají být analýza, návrh a implementace slovníku, bezztrátová migrace původních dat, důkladné otestování na netriviální množině dat a diskuze/hodnocení dosažených výsledků. Pro odbornou komunitu má vzniknout srozumitelný tutoriál.

**Rozsah** Text bakalářské práce má 37 normovaných stran dle metodiky KIV.

**Logická struktura** Logická struktura práce doznala spíše drobných změn proti předchozí verzi. Nadále dochází k prolínání různých logických částí napříč textem (např. analýza existujících slovníků v 2.11 zmiňuje specifikaci požadavků a implementaci).

**Obsah** Práce je **podprůměrná**. Některé dílčí části byly doplněny/rozšířeny, ale celkově je text práce bez zásadních změn. Ani nyní autor čtenáře **neseznámí s pojmy** řešené problematiky nebo s prioritami odborníků na nářečí. V textu **nenajdeme identifikaci potřebných změn** struktury dat nebo jejich transformace, ale ani věcné zdůvodnění. V elektronické podobě alespoň přiložil příklad strukturovaného popisu hesel slovníku (sešit tabulkového procesoru).

V druhé kapitole popisuje existující nářeční slovníky. Přidány byly klady a zápory u pěti z uváděných existujících slovníků. Přesto popis zásadního „Nářečního slovníku ZČU“ je velmi neurčitý (např. původní rozsah databáze student popisuje: „... více než 1 700 hesel“, „vyhledávání není zcela funkční“, nebo „struktura dat u jednotlivých hesel není jednotná“) a musím hodnotit za **nedostatečné** i nyní. Další typy<sup>1</sup> slovníků mohly být jistě zajímavou inspirací (propojení/procházení).

Třetí kapitola je o specifikaci požadavků. Přidán byl pouze diagram užití (str. 27), se kterým autor nepracuje. Zachováno je množství nekonkrétních/prázdných frází a textů tvořících vatu, jež jsou **neověřitelné** (např. škálovatelnost, plně vybavená serverovna, nepřetržitá dostupnost serveru).

Čtvrtá kapitola (41–42) ukazuje návrh webového rozhraní shodně s předchozím odevzdáním práce.

Pátá kapitola je pouze o analýze MySQL a MongoDB. Se zdůvodněním vhodnosti MongoDB (kapitola 5.3) se neztotožňují. **Student opět zaměňuje pojmy** „dynamické databázové schéma“ a „nestrukturovaná data“, když **považuje nekompletní záznam za nestrukturovaná data!** Současně mu nečiní problém (v kapitole 6.3) popsat schéma/strukturu pro MongoDB. Opakují se nerelevantní argumenty pro MongoDB jako „schopnost uchovat velké množství nestrukturovaných dat“ a „vhodnější pro aplikace s častým přístupem do databáze, častými úpravami a vhodnější funkce vyhledávání“. Při rozsahu slovníku by neměla problémy žádná databáze.

Šestou kapitolou je implementace, kde student uvádí technologie, a následuje slovní popis migrace dat s vlastním jednoúčelovým nástrojem v C#. Teprve následně je popis nového databázového modelu pro MongoDB (6.3). Rozsah tzv. minimálního záznamu (pro konkrétní slovní druhy) chybí. V neúplné větě (str. 50) zmiňuje XML formát dat od doc. Holuba, ale pravděpodobně měl na mysli XLS/XLSX (sešit tabulkového procesoru) se strukturovaným zápisem (8 sloupců a 18 řádek). V práci popis ani ukázka nejsou, a proto příkládám snímek formuláře k posudku.

Šedemá kapitola je nově rozdělena na testování jiných studentů (7.1) a vlastní end-to-end testování (7.2). Obojí s nulovou informační hodnotou. Testovalo prý šestnáct studentů Slezské univerzity

<sup>1</sup>např. <https://www.nechybujte.cz/> nebo <https://slovník-cizích-slov.abz.cz/>

v Opavě, kteří odhalili „šest drobných implementačních chyb“, ale není zřejmé jaké to byly. Student prý prováděl integrační testy již během vývoje, ale neexistuje nic než tato fráze v textu. Student uvádí šedesát end-to-end testů, jenže při podrobnějším pohledu do zdrojových kódů **netestuje nic vysloveně zásadního**. Samotné formuláře a validaci hesel a mluvčích zcela ignoruje. Testy pro „ověření integrace“ a „integritu dat“ student popsal jako hlavní účel testů. Jenže integrační testy (např. dat původní databáze?) ani jakékoliv testy integrity dat (např. validace pro nově vkládané nebo editované) jednoduše nejsou – jedná se o prázdné fráze v textu práce, ale **reálně neexistují**. Chybný odkaz na webovou aplikaci už překvapením nebyl.

Osmá kapitola měla být diskuzí dosažených výsledků. Student pouze konstatuje, že se jednalo o jeho první takový projekt, vše funguje, je otestováno, ale přesto **výsledný slovník migrovaná data raději ani neobsahuje(!)**. Musel by řešit duplicity. Bohužel jsem nucen opakovat, že **zcela chybí diskuze výsledků**, není kritické věcné zhodnocení ani porovnání s předchozími pracemi a jinými slovníky (výhody, nevýhody, srovnání). V diskuzi není hodnocení původních dat, jejich kvality, ale ani zmínka o úspěšnosti realizované migrace dat. Postrádám vymezení se k analyzovaným slovníkům ze 2. kapitoly, anebo alespoň zhodnocení požadavků, které si stanovil ve 3. kapitole.

Ani v závěrečné kapitole se čtenář nedozví nic konkrétního. Pouze fráze „konzultace s doc. Holubem napomohla poukázat na problémy a vady stávajícího nářečního slovníku . . .“ nebo „. . . s funkčností jsem až na pár výjimek spokojen“ následovaná „ne všechny systémy se povedlo plně nahradit“. Bez bližších detailů. Pro plán budoucích prací, jako realizace výpisu/tisku dat ve formátu výkladových slovníků, se nabízí otázka proč se tak již (od roku 2020 nebo 2021) nestalo?

## Kvalita řešení a dosažených výsledků

**Návrh a implementace** Problémy s návrhem a implementací přetrvávají. Příčinu spatřuji jednoznačně v **nedostatečné analýze, nepochopení řešení domény a původních dat**. Student **neidentifikoval žádné problémy k řešení** pro proces migrace dat. Jednoúčelový triviální nástroj pro migraci dat z exportovaných CSV souborů (tabulky z relační databáze) do JSON je naivní a nepokrývá všechny typy ani existující hodnoty v původních datech (např. sousloví).

Vkládání **duplicitních** slovníkových hesel je zarážející. Na vině je chybný datový model a implementace. Navržený datový model reálně nepokrývá možnou situaci, když může být stejné heslo současně různého slovního druhu. Případné sloučení hesel musí provést administrátor/správce ručně.

Proti předchozí odevzdané práci je zaveden číselník obcí (včetně kraje a okresu) používaný jen pro určení bydliště mluvčích. Atribut *oblast* pro určení místa, kde je slovo používáno, je pouze volný text – nebyla snaha realizovat označení oblastí srozumitelnější (např. rozklíčovat zkratky a vhodně použít číselník obcí. Při více oblastech je nutné zapisovat vše do stejného textového pole a oddělovat hodnoty čárkou. Existují i další atributy (např. příklady užití slova), které vyžadují, aby data „strukturoval“ (odřádkováním) uživatel místo použití vhodného typu formuláře. Ostatní atributy, kde by byl očekáván číselník, jsou přímo textové hodnoty<sup>2</sup>. Výklad významu hesla realizuje tzv. „křížové odkazy“ (synonyma, antonyma a frazémy), ale datový model poskytuje prosté pole Stringů, nikoliv jednoznačné propojení/vazbu na další hesla slovníku na úrovni databáze.

Při řešení se student zaměřuje na **nepodstatné** části (např. profilová fotografie, správa session) místo s tématem souvisejících funkcí (např. upload audio souboru). Profilová fotografie i správa session jsou podporovány použitým frameworkem, ale upload audio souborů by znamenal více úsilí.

Implementován je nevhodný způsob nápovědy atributů „s průvodcem“, kde nelze vyvolat nápovědu pro konkrétní pole přímo. Do uživatelské přívětivosti a pro praktické použití má webová aplikace daleko, když je nutné používat funkci zpět prohlížeče – z přehledů a především z editačních formulářů nelze plynule přejít zpět ani provedené změny zahodit.

Popis členění zdrojových souborů do adresářů a souborů, je na třech listech v kapitole 6.4. Nově byl přidán pouze obecný diagram (str. 56), na který se v textu neodkazuje. Zdrojové soubory komentovány nejsou. Objevuje se četné vnořování cyklů nebo if-else konstrukcí.

<sup>2</sup>Hodnoty „Dokonavý“ nebo „Nedokonavý“ pro slovesný vid a např. „Podstatné jméno“ nebo „Sloveso“ pro slovní druh.

Student opět konstatuje soulad s GDPR, ale problematikou se v textu nezabýval. Správce osobních údajů by se měl mít na pozoru. Veřejně (aniž bych byl přihlášený) je viditelné celé jméno, rok i místo narození mluvčí (např. slovo „fičet“)! Pochybná může být prokazatelnost uděleného souhlasu u mluvčích, které může vkládat libovolný registrovaný uživatel.

**Migrace dat – bezztrátový přenos dat** Na první pohled jsou navržený datový model i samotná migrace dat problematické. **Chybí analýza dat** původní relační databáze, ověření konzistence a úplnosti existujících záznamů. Ani minimalistickou popisnou charakteristiku před ani po transformaci dat nenajdeme, např. počty hodnot napříč atributy nebo slovními druhy. Jaké chyby v datech byly identifikovány student neuvádí. Vedle prostých slov uměl původní model zachytit také například *sousloví*, které je při transformaci jednoduše ignorováno – zřejmě z neznalosti dat. Transformovaná data se stávají nekonzistentní, když nadále obsahují původní hodnotu „*soussl.*“ (příp. jakoukoliv její obdobu). Při bližším pohledu je zřejmé, že *sousloví* studentův nový model ani webová aplikace **nepodporují**, když formulář striktně předpokládá určení slovního druhu, kde hodnota pro *sousloví* není. Troufám si tvrdit, že **výsledek postrádá smysl**. Chybí doklad o kvalitě dat, o úspěšnosti transformace, ale i popis vypořádání se s nedostatky původních dat. Splnění **bezztrátového přenosu dat student neprokázal**.

**Testování** Část testování jsem popisoval výše. To **podstatné** – formulář pro editaci/vkládání hesel a mluvčích – vůbec **netestuje**. V aplikaci realizovaná validace dat je omezena na ověření existence pár povinných hodnot (heslo, slovní druh), ale nic víc. Poměrně triviální formuláře obsahují **řadu nedostatků** a chyb. Rok narození mluvčích je bez validace, když akceptuje 1800 nebo 2050. U atributů *bydliště* a *původem* je nefunkční editace, když prostě procházení (tabulátorem) vyplněnou obec automaticky odstraní.

**Komunitně srozumitelný a veřejně dostupný tutoriál** Dle mého není tutoriálem uživatelská dokumentace z přílohy bakalářské práce. Zejména, protože neobsahuje to co od tutoriálu očekávám (tj. názorně a krok za krokem ukázat příklady konkrétního použití a vyplnění pro jednotlivé slovní druhy nebo slovní spojení). To by přitom mohlo současně sloužit i jako scénář pro otestování aplikace, když by bylo uvedeno co a jak vyplnit/nevyplnit. Uživatelsky přívětivá není ani nápověda skrytá pod jediným tlačítkem (otazníku) na začátku formuláře (pro všechna pole). Potřebuji-li poradit s konkrétním formulářovým prvkem, musím vždy nejprve proklikat průvodce přes ty předchozí. Ani toto však nepovažuji za požadovaný tutoriál.

## Formální a jazyková úroveň

Formální a jazykovou úroveň považuji za **nedostatečnou**. Neúměrná je četnost všeobecných, neurčitých a zavádějících tvrzení, duplicitních informací, parazitních slov, ale i množství faktických chyb a nepřesností nebo kostrbatých a neúplných vět. Rozsah je uměle natahovaný nekonkrétním a duplicitním textem. V textu je nadměrné množství gramatických chyb, překlepů nebo chybějících písmen (např. „analýzi“, „sloupcých“, „kvatitativní“, „údzbu“), které odhalí i nástroje jazykové kontroly. Chybí logická a především věcná **argumentace**. V textu jsou faktické chyby – např. z 3.1.4 cituji: „*multiplatformnost, tedy zachování svou čitelnosti nezávisle na velikosti zobrazovacího zařízení*“; nebo v 4.2 je v textu citovaný průzkum „*pro přelom roku 2019/2022*“, který byl ve skutečnosti 2021/2022, jak uvádí seznam literatury. Odkud pochází hodnota 53% v další větě (pro rok 2019) už zřejmě není. Text není konzistentní. Např. diagram na obrázku 3.1 uvádí třídu (rolí uživatele) „*Návštěvník*“, ale v textu 3.2.3 je to „*Běžný (neregistrovaný) uživatel*“. Diagramy/obrázky zabírají celou stránku (na str. 49 a 56). **Typografická pravidla** student zcela ignoruje – střídavě používá české i anglické uvozovky, oddělovač tisíců jen někdy, a objevuje se nevhodné dělení čísel i slov. Součástí dokumentu jsou seznamy obrázků, literatury a zkratk a uživatelská dokumentace, zdrojové soubory webové aplikace, obsah původní databáze, triviální zdrojový soubor pro transformaci dat a data studentů v XLS(X).

## Práce s literaturou

Práce s prameny je **nedostatečná**. Celkově bylo použito 19 elektronických zdrojů. Kvalifikační práce vychází z práce jiných autorů, a podstatným nedostatkem je uvedení řádné citace pouze u dvou autorů (až v kapitole 6.3). Další autoři jsou pouze vyjmenováni v textu. V textu je množství tvrzení nepodložených uvedením použitého pramene.

## Doplňující informace k práci

Studenta jsem požádal o předvedení migrace dat a zodpovězení několika otázek k bodům 3–6 zásad pro vypracování. Na moji žádost poskytl (9. 8. 2022) dodatečně data z databáze původní, aktuální, ale i data po sloučení obou databází. O znalosti problematiky ani o správnosti a vhodnosti předloženého výsledku mne však nepřesvědčil. Data přiložená k práci sám prohlásil za chybná (kolekce *districts.json* přiložena není, soubor *po\_migraci.json* neobsahoval data po transformaci).

## Dotazy k práci

1. Jaký je původ, rozsah a celkově význam nestrukturovaných dat v bakalářské práci, když v analýze databázových systémů (str. 44) uvádíte, cituji: „*Hlavní výhodou MongoDB oproti SQL databázím je schopnost uchovat velké množství nestrukturovaných dat, což je zejména vhodné pro ukládání ne zcela kompletních záznamů, např. právě hesel ve slovníku.*“?

## Splnění zadání

Zadání se stanovenými zásadami pro vypracování bakalářské práce student **celkově nesplnil**.

První a druhý bod zásad vypracování byly splněny bez výhrad. **S menšími výhradami** byl splněn třetí bod (není zřetelná role odborníka na nářečí). **S velkými výhradami** byl splněn bod 5. (aplikace nebyla studentem otestována důkladně). **Nesplněné** byly body zadání: 4. (studentem nebylo prokázáno požadované zabezpečení bezztrátového přenosu původních dat nebo zdůvodněna změna schéma), 6. (prostá uživatelská dokumentace ani nepraktický průvodce s nápovědou k formulářovým polím nepovažuji za komunitně srozumitelný a veřejně dostupný tutoriál) a 7. (hodnocení výsledků obsahuje pouze obecné fráze).

## Navržené hodnocení a doporučení k obhajobě

Navrhuji hodnocení známkou *nevyhověl* a práci *nedoporučuji k obhajobě*. Některé části student doplnil nebo mírně upravil, ale i přes předchozí podrobný posudek zůstala významná část nedostatků a výtek bez povšimnutí. Pro případ neúspěšné obhajoby navrhuji komisi požadovat, v souladu s čl. 49 platného Studijního a zkušebního řádu ZČU, vypracování nové kvalifikační práce jako *podmínkou pro opakování obhajoby*.

V Plzni 12. 8. 2022

Ing. Petr Včelák, Ph.D.  
KIV, ZČU

## Příloha

Obrázek představuje strukturovaný popis/kartu jednoho slovníkového hesla (list tabulkového procesoru). Zdroj: sešit přiložený na CD hodnocené bakalářské práce.

	A	B	C	D	E	F	G	H
1								
2	Tab.							
3								
4	č.	základní podoba	výslovnost	substantiva	adjektiva, zájmena, číslovky	slovesa	neohesbná slova	slovní spojení
5				koncovka 2. pádu sg. nebo pl.	určení slovního druhu	určení vidu	určení slovního druhu	běžné sousloví
6				jmenný rod		vazba		frazém (určení typu)
7				určení pomn.				termín (určení typu)
8				volitelná rubrika	volitelná rubrika	volitelná rubrika	volitelná rubrika	volitelná rubrika
9		výklad významu	kontext/exemplifikace		doplňující určení slova či spojení	křížové odkazy	odkaz na rodilého mluvčího	jiné údaje
10			kontext		realie a souvislosti	synonyma	příjmení, jméno	
11	1.		kontext		původ slova	antonyma	zkratka jména, rok narození	
12			exemplifikace		jiná forma slova	frazémy	profese, vzdělání	
13			exemplifikace		časové, stylistické, teritoriální, jiné	jiné	obec (název, zkratka, okres, kraj)	
14			kontext		realie a souvislosti	synonyma	příjmení, jméno	
15	2.		kontext		původ slova	antonyma	zkratka jména, rok narození	
16			exemplifikace		jiná forma slova	frazémy	profese, vzdělání	
17			exemplifikace		časové, stylistické, teritoriální, jiné	jiné	obec (název, zkratka, okres, kraj)	
18			kontext		realie a souvislosti	synonyma	příjmení, jméno	
19	3.		kontext		původ slova	antonyma	zkratka jména, rok narození	
20			exemplifikace		jiná forma slova	frazémy	profese, vzdělání	
21			exemplifikace		časové, stylistické, teritoriální, jiné	jiné	obec (název, zkratka, okres, kraj)	