

Posudek oponenta bakalářské práce

Student: Marhefka Adam
Téma: Generování rodokmenů z matričních záznamů (id 24660)
Oponent: Šátek Václav, Ing., Ph.D., UIT S FIT VUT

- Náročnost zadání** průměrně obtížné zadání
Student navazuje na předchozí práce zabývající se zpracováním matričních záznamů. Tyto práce rozšiřuje (doplnění databáze o nové události - sňatky a úmrtí) a optimalizuje (především eliminuje načítání z grafové databáze).
- Splnění požadavků zadání** zadání splněno
Zadání bylo ve všech bodech splněno.
- Rozsah technické zprávy** je v obvyklém rozmezí
Předložená práce je obvyklého rozsahu.
- Prezentační úroveň předložené práce** 75 b. (C)
Práce je logicky vystavena a jednotlivé kapitoly na sebe navazují. Drobná výtka: pro čtenáře je "nepříjemné" rozdělení tabulek (např. 3.3, 3.4 a 3.5) na 2 stránky.
- Formální úprava technické zprávy** 85 b. (B)
Práce je napsána ve slovenském jazyce v sázecím prostředí LaTeX.
- Práce s literaturou** 65 b. (D)
Student použil především "online" zdroje (9 z 12ti zdrojů jsou url odkazy). U reference [9] a [12] nelze určit o jaký typ odkazu se jedná - článek, skripta, BP/DP...? Nebyla citována (asi ani použita) literatura doporučená v zadání práce.
- Realizační výstup** 78 b. (C)
Výsledná aplikace je implementována v Pythonu. Student si nejprve nastudoval stávající aplikaci z předchozích prací. Již při procházení kódu narazil na některé nedostatky, které odstranil/vylepšil. Poté rozšířil aplikaci o nové záznamy sňatků a úmrtí osob. Největší zrychlení aplikace získal novou implementací dat reprezentovaných přímo jako objekty v Pythonu namísto předchozí grafové databáze, která značně zpomalovala práci s daty.
- Využitelnost výsledků**
Práce navazuje a rozšiřuje předchozí aplikaci pro práci s matričními záznamy. Práce je v praxi využitelná při generování podrobnějších rodokmenů.
- Otázky k obhajobě**
 - V práci jste navrhl nové metriky pro porovnávání záznamů (Jarova a Jaro-Winklerová vzdálenost). Co Vás k použití těchto metrik vedlo a jsou nějaké případy, kdy by mohlo být jejich použití v aplikaci výhodnější než použití stávajících (Levenšteinových) metrik?
 - V kapitole 5.1 jste testoval efektivitu Vašich úprav na 2 různých "podobných" sadách vzorků. Proč jste netestoval na stejné sadě vzorků?
- Souhrnné hodnocení** 78 b. dobře (C)
Práce byla průměrně obtížná, student splnil všechny body zadání. Práci **doporučuji k obhajobě** a hodnotím **stupněm dobře** (C, 78 bodů).

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 1. června 2022

Šátek Václav, Ing., Ph.D.
oponent