

Posudek oponenta diplomové práce

Student: Dušek Vladimír, Bc.
Téma: Integrace, vizualizace a dolování z dat zemí světa (id 25103)
Oponent: Rychlý Marek, RNDr., Ph.D., UIFS FIT VUT

- Náročnost zadání** průměrně obtížné zadání
Jedná se o průměrně obtížné zadání. Použité datové zdroje, postupy předzpracování i dolování jsou známy a tak podstata práce spočívá v jejich použití a návrhu a implementaci výsledné aplikace.
- Splnění požadavků zadání** zadání splněno
Zadání je splněno bez výhrad.
- Rozsah technické zprávy** je v obvyklém rozmezí
Technická zpráva je svým rozsahem v obvyklém rozmezí, od úvodu po závěr obsahuje 64 vysázených stran. Jednotlivé části zprávy mají přiměřený rozsah a jsou svým obsahem informačně bohaté.
- Prezentační úroveň předložené práce** 75 b. (C)
Struktura technické zprávy je logická a odpovídá postupu vývoje výsledného programového řešení. Z obsahu by bylo vhodné vyjmout do samostatné kapitoly experimenty tak, aby bylo vidět členění na jednotlivé úlohy dolování dat (nyní v kap. 5.5 "Dolování znalostí ze získaných dat").
- Formální úprava technické zprávy** 95 b. (A)
Z hlediska formální úpravy je technická zpráva na výborné úrovni a prakticky neobsahuje viditelné nedostatky. Oceňuji množství a kvalitu obrázků vhodně doplňujících text technické zprávy.
- Práce s literaturou** 85 b. (B)
Seznam literatury obsahuje 32 položek z nichž podstatná část jsou odborné zdroje. Použité zdroje jsou relevantní a vhodně pokrývají téma práce. Všechny zdroje ze seznamu jsou v textu technické zprávy vhodně citovány a lze dobře odlišit převzaté prvky od vlastních výsledků a úvah.
- Realizační výstup** 85 b. (B)
Realizačním výstupem je webová aplikace pro vizualizaci zpracovaných dat a získaných znalostí. Oddělenými komponentami řešení je také sada propojených úloh definovaných na platformě Apache Airflow pro získávání a předzpracování dat z mnoha informačních zdrojů a sada Python skriptů pro dolování z dat pomocí knihovny scikit-learn. Řešení je dobře navržené a zdrojový kód je dostatečně komentován. Vytknout lze návrh databázového schéma a způsob využití databáze, kdy je při nahrávání aktuálních dat vždy smazána celá databáze namísto pouhého doplnění chybějících dat.
- Využitelnost výsledků**
Výsledná aplikace je dobře využitelná v praxi pro prezentaci zajímavých statistických dat z různých států.
- Otázky k obhajobě**
 - Jak často probíhá získávání dat, plnění (aktualizace) databáze a následné dolování z dat? Jak dlouho tyto aktivity celkově trvají a jak by bylo možné je optimalizovat?
- Souhrnné hodnocení** 85 b. velmi dobře (B)
Výsledky diplomové práce jsou velmi dobré. Technická zpráva je v pořádku a výsledné programové řešení je dobře navrženo i implementováno a je prakticky použitelné a zajímavé pro širší veřejnost. Navrhuji hodnotit práci stupněm **velmi dobře (B)**.

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

V Brně dne: 1. června 2022

Rychlý Marek, RNDr., Ph.D.
oponent