

Posudek oponenta bakalářské práce

Autor/autorka práce: **Jakub Šmrha**

Název práce: **Automatická těžba dat z veřejných zdrojů pro vizualizaci**

Obsah práce

Cílem práce bylo prostudovat možnosti zpracování částečně strukturovaných dat popisující historické události vytvořit konfigurovatelný nástroj pro získání dat z veřejných zdrojů jako je Wikipedia nebo DB-pedia.

Autor v teoretické části textu práce popisuje rozdělení datových zdrojů z hlediska strukturovanosti dat a veřejně dostupné datové zdroje. V praktické části pak nejprve zvažuje dvě možnosti vizualizace dat s využitím existující knihovny zobrazující časovou osu a následně popisuje návrh, implementaci a testování vytvořeného nástroje.

Kvalita řešení (praktická část bakalářské práce)

V praktické části autor vytvořil nástroj pro zpracování částečně strukturovaných dat, který je plně funkční. Zdrojový kód se skládá z 55 .java souborů o celkové velikosti 213 kB, je standardním způsobem členěn do balíků, je celkem přehledný, ale dokumentační komentáře v mnoha třídách částečně nebo zcela chybí. Testování bylo provedeno pomocí jednotkového testování a manuálního funkčního testování několika testery. Elektronická příloha práce obsahuje kromě textu práce i zdrojové kódy a spustitelnou verzi nástroje se všemi potřebnými knihovnami. Postrádám vygenerovaný JavaDoc.

Kvalita řešení (text bakalářské práce a práce s literaturou)

Text má celkem 86 stran (řádkování cca 1.1) a má celkem logickou strukturu. Poměr teoretické a praktické části je zhruba 1 : 2. Text je celkem přehledně členěn do kapitol a je doplněn výpisy kódu a obrázky. Přílohy práce tvoří výsledky testování, příklady datových formátů a zdrojů a uživatelská a programátorská příručka. V práci postrádám UML diagram případů užití, který by v této práci byl poměrně informativní, a v přílohách podrobnější UML diagram tříd (základní struktura rozhraní a tříd je na Obr. 6.1 na str. 42). Testování mohlo být v práci popsáno více systematicky (např. popsat testovací scénáře, jejich počet, počet testerů, statistiky výsledků atd.). Kladně naopak hodnotím porovnání výsledků s ChatGPT, ač to není požadováno v zadání práce.

K textu mám dále několik drobných výhrad. V textu se občas vyskytuje bílé místo na konci stránky (např. str. 17, 35, 43 a další). Nečíslované nadpisy jsou nekonzistentní a jsou buď přímo v odstavci bez odřádkování (např. v Kap. 6.5, kde se jedná o třetí úroveň) nebo standardně s odřádkováním (např. v Kap. 6.6.4, kde se jedná o čtvrtou úroveň). Některé formulace působí zvláště (např. „Nícméně, implementace API byla druhořadá při tvorbě programu, proto je možné, že nebude tak intuitivní s tím pracovat.“ na str. 47). Množství překlepů je průměrné.

Zdrojů je v práci průměrné množství (konkrétně 21). Zdroje jsou v textu důsledně odkazovány. V několika případech se jedná o stránky Wikipedie, ale v kontextu této práce je to na místě. Reference jsou řazeny abecedně podle prvního autora, což je, vzhledem k tomu, že většina referencí má autora uvedeného, v pořádku.

Splnění zadání

Práce splňuje zadání.

Doplňující informace k bakalářské práci

Nástroj využívá knihovnu pro zobrazení časové osy vyvinutou v rámci několika předchozích bakalářských a diplomových prací.

Dotazy k bakalářské práci

1. Na základě čeho uvádíte pokrytí jednotkovými testy, především pak pokrytí bez GUI a POJO tříd, které uvádíte v rozsahu 85 až 90 %? Nebylo by možné udat přesné číslo?
2. Kolik bylo využito testerů při testování nástroje?

Vzhledem k množství drobných problémům uvedených výše navrhuji známku **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 29.5.2023

Ing. Tomáš Potužák, Ph.D.