

# Posudek oponenta bakalářské práce

Autor/autorka práce: **Karel Šilhavý**

Název práce: **Konfigurovatelné získávání grafových dat z otevřených zdrojů**

## Obsah práce

Cílem práce bylo prostudovat možnosti získání grafových dat z částečně strukturovaného textu a vytvořit systém pro získání těchto dat z infoboxů Wikipedie.

Autor v teoretické části textu práce popisuje rozdělení dat z hlediska jejich strukturovanosti a otevřené zdroje informací. V praktické části textu práce pak popisuje konkrétní výběr použitého zdroje dat a realizaci systému včetně testování na několika testovacích případech.

## Kvalita řešení (praktická část bakalářské práce)

V praktické části autor vytvořil systém skládající se ze dvou samostatných aplikací – DatabaseInserter a LinkingApp. První aplikace slouží ke zpracování dumpu Wikipedie a uložení výstupu do databáze. Druhá aplikace s GUI umožňuje uživateli data z databáze spojovat do výsledných grafových dat, které je možno exportovat a následně zobrazit (zobrazení není součástí zadání, viz níže). Obě aplikace jsou funkční. Ovládání GUI aplikace není příliš intuitivní a velice stručná uživatelská příručka, ve které o ovládání GUI nic není, nezavěšenému uživateli příliš nepomůže. Zdrojový kód se dohromady skládá z 20 .java souborů o celkové velikosti 70 kB. Samotné dělení systému na dvě samostatné aplikace se jeví jako zbytečné. Obě aplikace jsou členěny do několika balíčků, kořenový balík v obou případech chybí. JavaDoc komentáře jsou u menší části metod, u tříd a jejich atributů téměř zcela chybí. CD obsahuje kromě textu práce i zdrojové kódy, spustitelné .jar soubory pro obě aplikace, skripty pro inicializaci databáze a naplnění testovacími daty a příklady výstupních grafových dat. Postrádám vygenerovaný JavaDoc.

## Kvalita řešení (text bakalářské práce a práce s literaturou)

Text má celkem 43 stran (řádkování cca 1.3) a má celkem logickou strukturu. Poměr teoretické a praktické části je zhruba 1 : 1. Text je celkem přehledně členěn do kapitol a je doplněn výpisy, obrázky a tabulkami. V práci postrádám UML diagramy případů užití a tříd. Přílohu práce tvoří velice stručná (mírně přes 1 stranu) uživatelská příručka, která se však týká pouze aplikace Database-Inserter. GUI aplikace LinkingApp z uživatelského hlediska popsáno není (je zde pouze odkaz na kapitolu v popisu implementace). V některých částech textu jsou poměrně vágní formulace nebo na sebe text zdánlivě nenasazuje (např. Kap. 2.2 odstavec 1 a 2).

Hlavním problémem textu práce je však její teoretická část, tedy první dva body zadání. V práci jsou pouze stručně popsány typy dat (strukturovaná, nestrukturovaná a částečně strukturovaná) a následně otevřené zdroje informací se zaměřením na Wikipedii. To však zcela neodpovídá bodům 1 a 2 zadání, např. analýza možnosti získávání grafových dat z dostupných zdrojů (bod 2) z velké části chybí (částečně se nachází v Kap. 4.1).

K textu mám dále několik drobných výhrad. Úvod je až zbytečně obsáhlý a popisuje na první pohled nesouvisející věci. Motivace a přesný cíl práce tak z úvodu tak nejsou příliš jasné. V obsahu nejsou uvedeny přílohy. Množství překlepů je průměrné.

Zdrojů je v práci průměrné až lehce nadprůměrné 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.

## **Splnění zadání**

Body 3 až 5 zadání považuji za splněné. Body 1 a 2 zadání považuji za částečně splněné.

## **Doplňující informace k bakalářské práci**

Systém exportuje data do formátu zobrazitelného aplikací TimeLine, což je časová osa vyvinutá na KIV v rámci několika předchozích diplomových a bakalářských prací.

## **Dotazy k bakalářské práci**

1. Proč se vytvořený systém skládá ze dvou samostatných aplikací, které jsou propojeny jen daty v databázi? Nebylo by uživatelsky přívětivější dát extrakci dat z dumpu do databáze jako jednu z funkcionalit Linking App?
2. Proč je u třídy `JgraphDataHandler` uvedena jako autor i Gabriela Hessová?

Vzhledem k problémům uvedeným výše navrhuji známku **dobře** a práci doporučuji k obhajobě.

V Plzni 9.8.2021

Ing. Tomáš Potužák, Ph.D.