

Hodnocení vedoucího bakalářské práce

Autor/autorka práce: **Gabriela Hessová**

Název práce: **Automatické získání historických údajů z webových zdrojů**

Aktivita studenta

Studentka na práci aktivně pracovala během celého roku a dosažené výsledky pravidelně prezentovala a konzultovala. Účastnila se také aktivně všech schůzek týkajících se vizualizace dat.

Spolupráce s vedoucím

Na konzultace přicházela studentka připravená, s vlastními návrhy postupů a řešení problémů které se během řešení objevily.

Původnost práce a práce související

V úvodní části práce je popsána řada zdrojů, které mohou sloužit pro získání dat o událostech. Z nich je především pro svůj rozsah zvolena Wikipedie. Ačkoliv existuje řada nástrojů pro automatickou analýzu wikipedie, nepodařilo se najít žádný, který by umožnil zpracování údajů z infoboxů pro potřeby vizualizace událostí na časové ose bez nutnosti značného přepracování. Práci považuji za původní a dále rozšiřovatelný nástroj pro práci s Wikipedií.

Kvalita řešení

Text práce je přehledný a dobře čitelný. Teoretická část srovnává několik zdrojů historických dat a následně i několik nástrojů pro automatické zpracování wikipedie. Protože žádný z nich nebyl vhodný pro požadovanou úlohu, vytvořila studentka vlastní aplikaci, která je v druhé části detailně popsána. Vytvořená sada nástrojů není uživatelsky příliš přívětivá, ale z hlediska požadavků na ni je plně funkční – umožňuje zpracovat dump celé Wikipedie, získat z něj potřebné údaje a ty vložit do databáze historických událostí. Pro zjednodušení pracuje jen s historickými osobnostmi, kromě získání údajů o jejich narození a smrti generuje i provázání příbuzenskými vztahy. Zdrojové texty jsou komentované a dobře strukturované. Výhradu mám k tabulkám 7.2 a 7.4 (testování úplnosti a přesnosti) které nejsou příliš přehledné.

Využitelnost dosažených výsledků

Výsledná data jsou použitelná pro další vizualizaci. V rámci práce vznikly kromě požadovaných dat zajímavé statistiky o výskytu různých druhů infoboxů ve Wikipedii. Vytvořený nástroj je navíc možné snadno upravit pro hledání libovolných dat v infoboxech.

Splnění zadání

Všechny body zadání považuji za splněné.

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 13.8.2015

Ing. Richard Lipka, Ph.D.

