

Posudek oponenta bakalářské práce

Student: Švaňa Miloš

Téma: Čištění, extrakce textu a převod webových stránek do vertikálního formátu (id 18729)

Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno
Zadání práce bylo splněno, student navrhl a vytvořil nástroj, který umožní čištění a vertikalizaci velkého množství dat z webu.
3. **Rozsah technické zprávy** je v obvyklém rozmezí
Rozsah technické zprávy je v obvyklém rozmezí. Jednotlivé kapitoly práce jsou informačně bohaté a pro práci nezbytné.
4. **Prezentační úroveň předložené práce** 85 b. (B)
Logická struktura technické zprávy je velmi dobrá. Rozsahy a návaznosti jednotlivých kapitol jsou také na velmi dobré úrovni. Práce je dobře pochopitelná pro čtenáře. Vytkl bych snad jen to, že práci trochu chybí širší nadhled a nezkoumá jiné existující přístupy pro vertikalizaci.
5. **Formální úprava technické zprávy** 90 b. (A)
Po typografické a jazykové stránce nelze práci téměř nic vytknout. Snad jen občasné chybějící URL odkazy na zmíněné nástroje.
6. **Práce s literaturou** 90 b. (A)
Student použil rozumné množství zdrojů a v práci se na ně vhodně odkazuje. Student u vědeckých článků uvádí, kdy je online navštívil. Tato informace však není nutná.
7. **Realizační výstup** 90 b. (A)
Realizační výstup práce je velmi zdařilý. Studentovi se povedlo odstranit většinu nedostatků stávající verze. Hlavní přínos studenta vidím ve snížení nároků vertikalizátoru na operační paměť.
8. **Využitelnost výsledků**
Jedná se o práci kompilačního charakteru. Výsledky práce budou využity ke zpracování velkých dat ve Výzkumné skupině znalostních technologií. Student také v práci slibuje, že jednotlivé části vertikalizátoru v nejbližší době zveřejní na serveru github.com či v repozitáři Python Package Index.
9. **Otázky k obhajobě**
 1. Na straně 14, v sekci popisující způsoby porovnávání nástrojů Justext a Boilerpipe píšete, že když jeden z nástrojů selže, bod je přidělen automaticky jeho konkurentovi. Co když ale bude výstup konkurenčního nástroje zcela špatný? Taká mu přidělíte bod?
 2. Na straně 16 uvádíte, že celý vertikalizátor je navržen jako sada nástrojů, které mezi sebou komunikují podobně jako roury v Unixu. Jak přesně jednotlivé nástroje v proudu dat rozpoznají hranice mezi jednotlivými dokumenty?
10. **Souhrnné hodnocení** 90 b. výborně (A)
Předložená práce se zabývá tvorbou vertikalizátoru (nejen) pro Výzkumnou skupinu znalostních technologií na FIT VUT. Vytvořený vertikalizátor odstranil nedostatky předchozí verze a bude dále využíván výzkumnou skupinou. Technická zpráva je na velmi dobré úrovni. Práce je čtivá a dobře popisuje danou problematiku. Z těchto důvodů navrhuji hodnocení stupněm A (90b).

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

V Brně dne: 1. června 2016

.....
podpis