

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Kategorizace uživatelů na základě historie stahovaných webových dokumentů
Jméno autora:	Dušan Jenčík
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Oponent práce:	Mgr. Viliam Lisý, Ph.D.
Pracoviště oponenta práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Študent sa musel zoznámiť s novými technikami nad rámec učiva jeho bakalárskeho programu a musel si vhodne došpecifikovať zadanie, ktoré mu umožňovalo pristúpiť k problému rôznymi spôsobmi.	

Splnění zadání	splněno s většími výhradami
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Miesto kategorizácie užívateľov do kategórií podľa veku a pohlavia, ako naznačovalo zadanie, študent kategorizoval užívateľov podľa záujmov. Väčší problém bol, že si študent jasne nestanovil kritériá kvality kategorizácie (bod 5 zadania). Z tohoto dôvodu nebolo veľmi jasné, či sú študentove rozhodnutia ohľadne metód spracovania dát správne ani ako dobrú kategorizáciu napokon našiel. Zadanie ale stále považujem za splnené, pretože práca obsahuje aspoň kvalitatívne vyhodnotenie nájdených zhlukov na základe príkladov. Tak isto súhlasím so študentom v tom, že nájdené kategórie záujmov užívateľov by nebolo ťažké transformovať na ich iné charakteristiky.	

Zvolený postup řešení	částečně vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Za veľký metodický problém považujem, že si práca na začiatku jasne nestanovila konkrétne parametre hľadaných zhlukov. Pomohla by napríklad konkrétna motivácia, k čomu majú nájdené zhluky slúžiť alebo konkrétne kritérium, ako kvalitu nájdených zhlukov vyhodnotiť. Toto sa mierne dalo vytušiť v závere práce pri diskusii jednotlivých výsledkov, ale väčšina rozhodnutí v práci nebola a nemohla byť kvôli absencii týchto kritérií jasne odôvodnená. Napríklad, študent z dát hneď na začiatku odstránil 10% najčastejšie používaných adries s tým, že tie nedokážu diskriminovať užívateľov. Bez podrobnejšieho vysvetlenia o tom nie som vôbec presvedčený. Relatívna frekvencia návštev Facebooku a napríklad CNN by mohol byť silný indikátor veku. Z práce to pôsobí tak, že študent robil podobné rozhodnutia na základe dojmov a nie rigoróznejšieho štatistickej analýzy.	
Podobný problém spôsobila aj nejasná špecifikácia očakávaných výpočetných zdrojov, ktoré by mali byť na spracovanie dát použité. Študent diskutoval ktoré metódy sú výpočetne únosné a ktoré nie dlho pred tým, ako v experimentoch došlo na to, či chce pre výpočet použiť laptop alebo superpočítač. Jasné a včasné definovanie požiadavkov na výpočetné prostriedky a parametre výsledku by umožnili oveľa presnejšiu diskusiu jednotlivých rozhodnutí. Takto je práca popisom toho čo študent spravil, ale málo a nejasne sa venuje tomu prečo to tak spravil a prečo vylúčil prípadné alternatívy.	
Pozitívne hodnotím využitie paralely medzi kategorizáciou dokumentov a riešeným problémom, čo študentovi pomohlo použiť existujúce algoritmy a s dodaním dodatočnej informácie z katalógu webových stránok vytvorí dokonca pravdepodobnostný model záujmov užívateľov. Mám pocit, že toto je už len malý krok od odhadovania veku či pohlavia a študent tieto svoje výsledky v diskusii hodnotil zbytočne negatívne.	

Odborná úroveň	C - dobre
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat</i>	

získaných z praxe.

Ako je jasné z predošlého, študent k práci podľa môjho názoru nepristupoval dostatočne rigorózne. Na druhej strane, samotný výsledok práce v podobe vektoru miery záujmu o sadu abstraktných tém pre každého užívateľa by pre poskytovateľa dát mohol byť veľmi užitočný. Pri optimalizácii paralelného algoritmu pLSA musel študent tento netriviálny algoritmus pochopiť a naimplementovať so svojimi vylepšeniami. Tým študent preukázal schopnosť samostatnej práce na zložitejšom probléme.

Formální a jazyková úroveň, rozsah práce

A - výborně

Posuďte správnosť používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Po formálnej a jazykovej stránke nemám k práci zásadnejšie výhrady. Text je bez preklepov. Jeho čitateľnosť mierne znižuje veľké množstvo footnotov, ktoré človeka nútia skákať pri čítaní na spodok stránky. Ďalej by som osobne uvítal prezentáciu pozitívnych výsledkov pred negatívnymi, aby sa dali lepšie diskutovať dôvody neúspechu iných prístupov, ale aj zvolená štruktúra dáva zmysel.

Výběr zdrojů, korektnost citací

C - dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Práca obsahuje nadpriemerný počet citácií, ale samotná kapitola 2 popisujúca príbuzné prístupy má len necelé 2 strany. Keďže práca neskôr pracuje s analógiou riešeného problému a kategorizácie dokumentov, očakával by som minimálne ucelenejší prehľad tejto oblasti. V práci je jasné čo je príspevok študenta a čo sú existujúce výsledky a zoznam citácií je asi korektný, aj keď nie úplne štandardný. Dost' má táce mi napríklad príde nazvať rok vydania archívnych časopiseckých publikácií „Poslední aktualizace“.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Student spravil adekvátne množstvo práce a dospel k zaujímavým výsledkom, ktoré by pravdepodobne mohli byť použiteľné v praxi a preto si myslím, že by mal prácu obhájiť. Na druhej strane, keďže ide o prácu skôr výskumného ako implementačného rázu, očakával by som rigoróznejšiu definíciu skúmaného problému a podrobnejšiu analýzu vplyvu jednotlivých rozhodnutí na finálny výsledok. Väčšina argumentov je veľmi vágnych, nepodložených konkrétnymi dátami ani výsledkami experimentov. Myslím si, že by to zlepšilo nie len čitateľnosť práce, ale aj užitočnosť finálneho výsledku.

Predloženou záverečnou prácu hodnotím klasifikačným stupňom **C - dobre**.

Otazka: V tabuľke 4.1. na strane 33 porovnávate výpočetnú zložitost' algoritmov pomocou spodného asymptotického odhadu. Mohli by ste vysvetliť prečo ste zvolili tento neštandardný spôsob?

Datum: 3.6.2015

Podpis: