

# HODNOCENÍ DIPLOMOVÉ PRÁCE

## Vedoucí DP

Jméno diplomanta: Bc. Václav Voldřich

Garantující katedra: KKY

Název diplomové práce: Metody shlukové analýzy pro klasifikaci dokumentů pro účely následného využití v oblasti automatické adaptace jazykových modelů

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2	Samostatnost zpracování tématu DP	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3	Vhodnost použitých metod	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5	Správnost získaných výsledků	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6	Vlastní přínos	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Doplnění hodnocení, připomínky, dotazy:

Úvodem hodnocení je třeba zdůraznit, že student ve své práci nedodržel termíny stanovené vedoucí, na své práci od jejího zadání nepracoval, nebo jen minimálně a na celé práci začal pracovat až v červenci. Student svou práci poté zpracoval samostatně, nekonzultoval s vedoucí práce a první ukázkou práce přinesl až týden před jejím odevzdáním. Práce svým malým rozsahem i nedostatečným obsahem odpovídá této minimální době studentem jí věnované.

Jazyková úprava práce je zcela nedostatečná, práce obsahuje překlepy, nesmysly (např. použití neexistujících cizích slov), nedokončené věty nebo věty s vynechanými větnými členy. Velké množství textu v práci vůbec nedává smysl (např. : "Pro testování typů příznaků byly parametry metod, u nichž je možné zvolit počet shluků (k-means) byl zadán explicitně."). Grafická úprava práce je dobrá, ale projevuje známky rychlého dokončování práce - místo čísla odkazovaných tabulek jsou v textu otazníky, v seznamu obrázků je špatný název, atd.

Teoretická část práce je zcela nedostačující. Úkolem studenta bylo hned v prvním bodu zadání prostudovat pokročilé metody shlukové analýzy vhodné pro klasifikaci dokumentů. Ze seznamu literatury a z popisovaných metod je zřejmé, že student si neprostudoval žádnou vhodnou literaturu, ani doporučenou v zadání. V teoretické části práce pak popisuje shlukovací metody K-means a jeho varianty (což již bylo součástí jeho bakalářské práce), metodu Mean-shift a metodu PCA a ICA. Metody PCA a ICA přitom vůbec nejsou shlukovací metody. Všechny uvedené metody jsou nedostatečně popsány, není vůbec zřejmé, jakým způsobem byly některé z nich použity ke shlukování.

Co se týče zhodnocení vhodnosti použitých metod, z práce není vůbec zřejmé, zda student udělal nějaký rozbor existujících metod a metody uvedené v práci podle nějakého kritéria vybral. Metodu K-means a její varianty student použil, protože ji měl již zpracovanou z bakalářské práce, dále práce obsahuje už jen metodu Mean-shift. Ostatní metody uvedené jako shlukovací metody (PCA, ICA, metoda ohybu) nejsou shlukovacími metodami. Vybrané metody tedy hodnotím jako nevhodně vybrané a některé i naprosto nevhodně použité.

V práci úplně chybí popis toho, jakým způsobem byly metody zpracovány, nastaveny a použity. V rámci vyhodnocení metod student uvádí F-míru použitou pro vyhodnocení, neuvádí ovšem, jakým způsobem vyhodnotil, který výsledek je správný a který nikoli. Ve výsledcích má poté uvedeno, že některá metoda dosáhla F-míry 0,8 (z rozsahu (0,1)) u principiálně naprosto špatného výsledku (vytvoření jednoho velkého shluku ze všech dat, místo rozdělení do 4 shluků). Tím vytváří pochybnosti o správně počítaném vyhodnocení výsledků metod pomocí F-míry.

V rámci vyhodnocení výsledků student vyhodnocuje úspěšnost shlukování i pro metody, které nejsou shlukovací. Nikde v práci neuvádí, jakým způsobem metody ke shlukování použil, není tedy vůbec jasné jakým způsobem k výsledkům přišel. I samotné ohodnocení získaných výsledků z hlediska úspěšnosti shlukování je nedůvěryhodné, vzhledem k výše uvedenému nevysvětlenému způsobu výpočtu F-míry. Výsledkové tabulky uvedené v práci obsahují značné množství prázdných míst, nedopočtených výsledků. Předpokládám, že to je důsledek minimálního času, který student práci věnoval a už nestihl metody testovat.

Praktická část práce, tedy zdrojové kódy algoritmů odevzdané na přiloženém CD není vůbec v tomto stavu použitelná. Zdrojové kódy nejsou okomentované, chybí jakákoliv dokumentace k tomu jak s kódy pracovat nebo co poskytují za výsledky. Také tři různá jména autorů odevzdaných zdrojových kódů vedou k pochybnosti o autorství studenta.

Student tedy v kapitole 5 své práce nepravdivě uvádí, že práce je připravena na napojení na stávající projekt, práce v této podobě je bohužel naprosto nepoužitelná.

V závěru diplomové práce uvádí student: "Bohužel výzkum byl tak intenzivní, jak mohl být, ovšem ani základní algoritmy, které byly v této práci testované přinesly výsledky." Tato nesrozumitelná věta podle mého vystihuje celý styl této práce. Není zřejmé, zda student chce říci, že provedl intenzivní výzkum a bohužel metody neměly dobré výsledky, nebo zda chce říci, že výzkum neprovedl žádný a přesto má dobré výsledky.

Z práce je vidět, že teoretický výzkum nebyl proveden téměř žádný a výsledky uvedené v práci jsou přinejmenším zpochybnitelné, jak z hlediska použitých metod tak i z hlediska jejich vyhodnocení.

Práce v této podobě nesplňuje zadání a podle mého názoru neodpovídá rozsahem ani obsahem diplomové práci a tudíž práci hodnotím jako nevyhovující a nedoporučuji ji k obhajobě.

Splnění bodů zadání	<input type="checkbox"/> úplně	<input type="checkbox"/> částečně	<input checked="" type="checkbox"/> nesplněno	
Doporučení práce k obhajobě	<input type="checkbox"/> ano		<input checked="" type="checkbox"/> ne	
<b>Celkové hodnocení práce</b>	<input type="checkbox"/> výborně	<input type="checkbox"/> velmi dobře	<input type="checkbox"/> dobře	<input checked="" type="checkbox"/> nevyhověl
Jméno, příjmení, titul vedoucího DP: Lucie Skorkovská, Ing.				
Pracoviště vedoucího DP: KKY				

6.9.2013

Datum

Podpis