

Posudek oponenta bakalářské práce

Student: Pelantová Lucie
Téma: Automatické zpracování ručně opravených testů (id 18015)
Oponent: Herout Adam, prof. Ing., Ph.D., UPGM FIT VUT

1. **Náročnost zadání** **obtížnější zadání**
Zpracování obrazů není na bakalářském stupni přímo vyučováno a řešitelka tak musela dostudovat množství netriviálních znalostí.
2. **Splnění požadavků zadání** **zadání splněno s drobnými výhradami**
Body 4. a 5. zadání jsou splněny pouze rudimentárně a provedené testy vypovídají opravdu velmi málo o použitelnosti navržených postupů. Ani v textu, ani na přiloženém CD nebyl k nalezení jediný příklad naskenovaného textu, takže opět je těžké si udělat obrázek o tom, zda jsou metody navrženy správně či ne.
3. **Rozsah technické zprávy** **je v obvyklém rozmezí**
4. **Prezentační úroveň předložené práce** **70 b. (C)**
Velká část textu je věnována "teorii", jejíž popis nesvědčí o úplném pochopení souvislostí. Například text začíná velice obsáhlým popisem struktury a vlastností formátu PDF, když přitom řešitelka s PDF přijde do styku pouze prostřednictvím knihovny, která jí PDF převede do rastru, takže znalosti interní struktury PDF jsou pro práci dokonale irelevantní.
Na druhou stranu rozpoznání textu řešitelka nakonec řeší trénováním vlastního klasifikátoru založeného na SVM a v kapitole 2 přitom o SVM a příznakovém rozpoznávání (tak důležitém pro řešení) není ani slovo.
Jak bylo uvedeno výše, testování a prezentace jeho výsledků jsou slabé. Čtenář se nedozví, na jakých datech testování probíhalo, řešitelka nezkoušela pořídit data "snadná", "střední" a "těžká", aby vlastnosti metody byly zřejmé. Výsledkem jsou procentuální úspěšnosti, ale není specifikováno, co je považováno za úspěch, takže čísla vlastně není možné použitelně interpretovat.
5. **Formální úprava technické zprávy** **85 b. (B)**
Text je po formální stránce sepsán poměrně pečlivě, ale obsahuje občasné typografické chyby. Obrázky nejsou vždy jednotné a dostatečně kvalitní (obr. 2.2, 2.5, 3.5).
6. **Práce s literaturou** **75 b. (C)**
Řešitelka cituje velké množství hodnotných zdrojů. Sepsaný text ale nevypovídá o jejich správném pochopení.
7. **Realizační výstup** **70 b. (C)**
Z provedených testů je těžké vyčíst skutečné vlastnosti metody, ale nezdá se, že by vytvořené řešení bylo - byť jednoduchým - použitelným nástrojem řešícím alespoň snadné varianty zadané úlohy.
Funkční a hodnotná část může být detekce tabulek v dokumentech - zdá se ale, že je dosti pomalá.
Na více místech řešení se vyskytuje použití semínkového vyplňování, což neukazuje na dobrou volbu metod a dobré pochopení problematiky.
8. **Využitelnost výsledků**
Cílem práce bylo vytvořit nástroj, který by byl použitelný ve výuce a testování studentů. Dosažené výsledky i uživatelské rozhraní nepůsobí dojmem, že se cíle podařilo v uspokojivé míře dosáhnout. Vytvořený zdrojový kód není velice modulární a znovupoužitelný a metody nebyly vždy zvoleny šťastně (semínkové vyplňování).
9. **Otázky k obhajobě**
-
10. **Souhrnné hodnocení** **70 b. dobře (C)**
Zadání je obtížné a řešitelka musela studovat mnoho znalostí, které přesahují rámec bakalářského studia. Nelze jí tedy vyčítat, že nedosáhla naprosto dokonalého řešení a nepoužila nejnovější a nejsofistikovanější techniky. Bylo by lépe, kdyby se řešitelka zaměřila třeba jen na část celého řetězce zpracování dokumentů a tu vyřešila a hlavně vyhodnotila správně.

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

.....

podpis