

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucího

posudek oponenta

Autor/ka: *Bohuslav Macháč*

Název práce: *Načítání dat z tištěných dokladů*

Studijní program a obor: *Informatika, programování*

Rok odevzdání: *2011*

Jméno a tituly vedoucího/opponenta: *RNDr. Václav Krajiček*

Pracoviště: **KSVI MFF UK**

	e x c e l e n t n í	o d p o v í d a j í c í	s l a b š í	n e v y h o v u j í c í
Náročnost zadaného tématu		X		
Míra splnění zadání			X	
Rozsah práce		X		
Struktura textové části práce			X	
Analýza			X	
Vývojová dokumentace		X		
Uživatelská dokumentace		X		
Jazyková a typografická úroveň			X	
Návrh a design implementace			X	
Kvalita zpracování softwarové části		X		
Stabilita aplikace			X	X

Nejvýznamnější klady:

- Uživatelské rozhraní je zpracováno na dobré úrovni.
- Autor myslel na rozšiřitelnost aplikace. Možnosti rozšíření dobře zdokumentoval.

Nejzávažnější nedostatky:

- Co je asi největší slabinou celé práce, je absence jakýchkoliv testů, měření a porovnání, které by samy o sobě obhájily správnost zvoleného postupu. Bez nich jde jen o možné řešení problému bez možnosti zhodnocení. Chybí širší testovací množina. Autor v závěru pouze uvádí, že program zkoušel na dalších typech dokladů, ale bez uvedení konkrétních výsledků. Toto všechno jsou základy vědecké práce, které by si zde měl autor osvojit.
- Analýza problému je provedena nedbale. Autor sice popisuje rozhodnutí, která musel udělat, ale nepodpořil je žádnými ilustracemi nebo názornými příklady, které by podpořily jeho tvrzení.
- Klíčový problém, kterým se autor v práci zabývá, tzv. automatic document zoning není v rešerši vůbec zmíněn.
- Autor opomenul uvést, jestli podobné nástroje již existují. Uvádí jedinou referenci zabývající se podobným problémem, který v práci řeší (Cattoni et al 1998). Což se dá jen stěží nazvat bohatou, nebo dokonce vyčerpávající rešerší.
- Z formálních nedostatků je nejzávažnější fakt, že práce místy nedodrжуje vědecký styl.
- Aplikace na přiloženém mediu je nestabilní jak v režimu zpracování dokumentu, tak v režimu tvorby šablony. Později se zjistilo, že důvodem jsou chybějící metadata pro použitý OCR engine.
- Automatické hledání polí funguje na jediném přiloženém reálném příkladu vcelku uspokojivě, nastavení parametrů však není intuitivní, ani příliš interaktivní.

Další poznámky:

- Celá práce na mě dělá dojem, že se autor příliš soustředil na podružné věci jako uživatelské rozhraní a ukládání do XML, místo zpracování samotného problému a prezentace jeho úspěšného vyřešení formou testů a srovnání.
- Bylo by vhodné pracovat v metrickém systému (místo pixelů) a zavést krok kalibrace. Rastrové obrázky mnohdy obsahují údaj pro převod na metrické jednotky (DPI, apod).

	v ý b o r n ě	v e l i d o b ř e	d o b ř e	n e p r o s p ě l / a
Návrh známky			X	X

Datum: 13.6.2011

Podpis: