

## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Jméno studenta:** Daniel Vondra

**Název práce:** Parsování CV

**Autor posudku:** Pavel Kříž

**Cíl práce:** Základem práce bude vytvoření aplikace pro parsování životopisů ve formátu DOC, DOCX, RTF a PDF. Aplikace bude schopná detekovat jméno, příjmení, telefonní číslo, email, popřípadě titulu dané osoby z CV.

Povinná kritéria hodnocení práce	Stupeň hodnocení					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	X					
Vymezení cíle a jeho naplnění				X		
Zpracování teoretických aspektů tématu				X		
Zpracování praktických aspektů tématu				X		
Adekvátnost použitých metod				X		
Hloubka a správnost provedené analýzy				X		
Práce s literaturou				X		
Logická stavba a členění práce		X				
Jazyková a terminologická úroveň				X		
Formální úprava a náležitosti práce		X				
Vlastní přínos studenta			X			
Využitelnost výsledků práce v teorii (praxi)				X		

### Vyjádření k výsledku anti-plagiátorské kontroly:

Práce v rámci kontroly anti-plagiátorským systémem nevykazuje žádné nedostatky.

### Dílní připomínky a náměty:

- Text obsahuje překlepy a terminologické nepřesnosti (např. Collback, REEDME.sql).
- Tvzení „Server, kde bude aplikace nahrána, musí podporovat zvolený jazyk, skriptovací jazyk, kaskádové styly pro design aplikace a také databázi.“ není správné. Jak by si autor představoval „podporu kaskádových stylů serverem“?
- Na straně 3 jsou hned v prvním odstavci věty, které až na několik slov představují doslovnou citaci z uvedeného zdroje. Dle mého názoru nejde o parafrázování a měly by být označeny jako doslovná citace.
- Regulární výraz  $(19|20)\{d\}$  na str. 4 neodpovídá rokům 1900 až 2000, jak uvádí autor.
- Zkratka VMS není vysvětlena.
- Tvzení „Operační systém Linux oproti Windows podporuje funkce nutné pro převádění jednotlivých typů souborů, ve kterých se může životopis nacházet.“ není nijak podloženo a dle mého názoru není pravdivé. S ohledem na použité nástroje (PHP, LibreOffice) není mezi uvedenými oper. systémy valný rozdíl.
- Autor se místy v textu opakuje.
- Bylo by vhodné uvést, co je obsahem elektronické přílohy.

- Tvzení „Největší chybovost byla odstraněna a aplikace funguje s velikou pravděpodobností správných informací.“ není nijak podloženo.
- Kvalita a udržovatelnost výsledného kódu aplikace je spíše podprůměrná (nepoužit Autoloading vlastních tříd, kombinace práce s databází pomocí knihovny Dibi a nízkourovňových mysqli\_\* funkcí, často „špagetový kód“ absntující vzor Model-View-Controller).

### **Celkové posouzení práce a zdůvodnění výsledné známky:**

Student Daniel Vondra zpracoval bakalářskou práci v rozsahu 45 stran textu, ve které navrhl a vyvinul aplikaci pro automatickou extrakci klíčových informací ze životopisů. V první části práce definoval cíl a shrnul dvě existující řešení. Následně autor popsal vlastní řešení a využití technologie. V závěry uvedl problémy, které vyvoj provázely a shrnul výsledky.

Práce má jasně dané zadání, které je relativně ambiciózní. Nicméně v rámci bak. práce proveditelné s měřitelnou úspěšností/chybovostí. K jakémukoliv kvantitativnímu zhodnocení úspěšnosti extrakce však zřejmě vůbec nedošlo o reálná chybovost je otázkou. To výrazně snižuje kvalitu práce. Také rešeršní část je poměrně stručná, přestože zdroje na téma „Resume parsing“ lze dohledat. Kód aplikace připomíná spíše „první funkční řešení,“ než udržovatelný a přehledný zdrojový kód.

Přes uvedené výhrady konstatuji, že cíl práce byl rámcově splněn a výsledná aplikace může poskytovat jistou pomoc při zpracování velkého množství CV.

### **Otázky k obhajobě:**

- Na kolika různých reálných životopisech bylo výsledné (finální) řešení testováno a s jakou úspěšností?
- Jaké prostředky Vaše řešení nabízí k opravě případných chyb – je možné informace ke konkrétnímu CV manuálně doplnit nebo to plánované workflow nevyžaduje?
- Jakému rozsahu let odpovídá regulární výraz  $(19|20)\{2\}$  ?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: D**

**V Hradci Králové, dne 27. května 2019**

---

podpis