



# Hodnocení vedoucího závěrečné práce

**Vedoucí práce:** Ing. Jiří Novák, Ph.D.  
**Student:** Milan Vu  
**Název práce:** Extrakce strukturovaných dat z českých faktur  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 16. července 2021

## Hodnotící kritéria

### 1. Splnění zadání

- [1] zadání splněno
- ▶ [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání práce bylo splněno, i když aplikace by mohla být lépe otestována z pohledu rychlosti (resp. výsledky by mohly být lépe shrnuty např. formou tabulky) a také z pohledu použitelnosti (vzhledem k oboru práce to však nepovažuji za závažný problém). V kapitole 6 by mohly být explicitně shrnuty výhody/nevýhody nového řešení vůči existujícím metodám, aby si to čtenář nemusel sám domýšlet na základě popisů existujících řešení v analýze a popisu nového řešení v návrhové a implementační části. Implementace by mohla podporovat uživatelské účty.

### 2. Písemná část práce

80/100 (B)

Rozsah práce splňuje požadavky kladené na bakalářskou práci. Přestože se práce týká českých faktur, je napsaná v anglickém jazyce. I když to v tomto případě může být trochu matoucí, oceňuji snahu napsat práci v angličtině. Práce by mohla být o něco lépe strukturovaná. Kapitoly 3 a 5 jsou krátké, bylo by vhodnější je spojit s kapitolou 4 do jedné kapitoly nazvané např. Design and Implementation. Sekce 2.1, 2.3.1, aj. - když uvádím výčet metod v nějakém pořadí, a pak je v další sekci popisují, je vhodné toto pořadí zachovat. Některé části sekce 2.1.2 se dají považovat za zbytečné - popis RGB, binary image, apod. Generování podsekcí hloubky 4 a vyšší (např. 2.1.2.1) už práci trochu znepráhledňuje. Struktura sekce 2.2.1 není úplně vhodná - popis Pythonu a OpenCV je zamíchán mezi CRAFT a EAST. Popis Levenshteinovy vzdálenosti na str. 22 je celkem zbytečný, stačilo by uvést referenci. V sekci 2.5 by bylo vhodné doplnit reference. Seznam zkratk by mohl být setříděný podle abecedy.

### 3. Nepísemná část, přílohy

85 /100 (B)

V rámci práce byl implementován labelovací nástroj pro označování důležitých částí faktur - celková cena, datum splatnosti, IČO, apod. S jeho použitím byla získána trénovací data pro několik machine learning modelů - logistic regression, kNN, SVM, naive bayes, random forest, gradient boosting, které byly následně využity pro rozpoznávání důležitých míst v nově nahraných fakturách. Modely byly porovnány z pohledu přesnosti v experimentální části (kapitola 6). Informace z označených částí jsou extrahovány pomocí OCR engine a převedeny do strukturované formy obsahující umístění oblasti v obrázku (souřadnice x, y, výšku a šířku), typ textového pole a text. Není úplně zřejmé, jakou finální konfiguraci používá nasazená verze aplikace.

### 4. Hodnocení výsledků, jejich využitelnost

75 /100 (C)

Aplikace je prakticky použitelná pro analýzu faktur. Použitelnost je však do značné míry limitována tím, že nejsou podporovány uživatelské účty. Mohla by být také řešena nějaká možnost dalšího zpracování dat uživatelem (např. export do Excelu přes csv soubor, apod.).

### 5. Aktivita studenta

- [1] výborná aktivita
- [2] velmi dobrá aktivita
- ▶ [3] **průměrná aktivita**
- [4] slabší, ale ještě dostatečná aktivita
- [5] nedostatečná aktivita

Student by při řešení projektu mohl být aktivnější a své řešení mohl častěji a průběžně konzultovat.

### 6. Samostatnost studenta

- ▶ [1] **výborná samostatnost**
- [2] velmi dobrá samostatnost
- [3] průměrná samostatnost
- [4] slabší, ale ještě dostatečná samostatnost
- [5] nedostatečná samostatnost

Student je schopen samostatné tvůrčí práce.

## Celkové hodnocení

80 /100 (B)

Z výše uvedených důvodů práci doporučuji k obhajobě a hodnotím známkou B.

## Instrukce

### Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

### Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

### Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

### Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

### Aktivita studenta

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

### Samostatnost studenta

V souvislosti s průběhem a výsledkem práce posudte schopnost studenta samostatně tvůrčí práce.

### Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.