

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Šárka Uramová  
**Název práce** Vyhodnocování překladu textů v obrázcích  
**Rok odevzdání** 2024  
**Studijní program** Informatika  
**Specializace** Umělá inteligence

**Autor posudku** Jindřich Libovický Oponent  
**Pracoviště** Ústav formální a aplikované lingvistiky

## K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání			X	
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X			
<p>Předložená bakalářská práce se věnuje problému překladu v obrázcích jako jsou různá schémata nebo mapy (nikoli tedy např. text na fotografiích z reálného prostředí). Práce představuje jednu z možností, jak vyhodnocovat kvalitu takového překladu (kapitoly 3 a 4) a základní metodu, jak by bylo možné takovou úlohu řešit (kapitola 5).</p> <p>Nejcennějším přínosem práce je získání datasetu pro evaluaci kvality překladu textu v obrázcích za pomoci API k Wikipedii. Dataset obsahuje několik stovek obrázků, které jsou navzájem svými překlady. Práce bohužel nezmiňuje celou řadu statistik, které jsou důležité pro lepší orientaci v datasetu (např. o kolik slov se jedná, jaká je distribuce délek textových segmentů nebo jak často se stane, že řetězec vůbec není potřeba překládat).</p> <p>V další kapitole práce diskutuje možnosti, jak překlad textů v obrázcích evaluovat. Přestože představená metoda zní velmi rozumně, práce nepřináší dostatečné empirické argumenty, které by pomohly posoudit, jak dobře metrika funguje. Takové argumenty by mohly být například: porovnání automatických metrik s lidskou evaluací více systémů nebo například použití více překladových systémů, u kterých známe jejich obecnou kvalitu. Za zásadní metodologickou vadu považují průměrování automatických metrik přes různé cílové jazyky, protože hodnoty metrik BLEU a chrF se v různých jazycích pohybují v různých rozsazích.</p> <p>Podobné výtky lze uplatnit v případě kapitoly 5, která představuje samotný překladový systém, který kombinuje detekci, rozpoznání textu, překlad a vykreslení textu zpět do obrázku. Představená metoda zní opět velmi rozumně (v rámci toho, co lze stihnout za dobu řešení bakalářské práce), ale chybí detailnější empirická evaluace. Celý systém má celou řadu hyperparametrů (nastavení OCR systému, výběr překladového systému, parametry dekódování překladače), které pravděpodobně mají vliv na celkovou kvalitu fungování systému. Evaluace vlivu těchto hyperparametrů by zároveň pomohla demonstrovat efektivitu evaluační metriky.</p>				

## Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>		X		
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>		X		
Analýza		X		
Vývojová dokumentace		X		

Uživatelská dokumentace		X		
<p>Práce je psaná česky, a obsahuje jen zanedbatelné množství gramatických či stylistických nedostatků. Práce má logickou strukturu, která postupuje od získání dat k použití dat k evaluaci komplexního systému.</p> <p>V textu jsou jen drobné nedostatky (např. hned první tvrzení práce o tom, že se v posledních letech zlepšila kvalita automatického překladu by mělo mít oporu v literatuře nebo třeba tvrzení, že Google Lens je end-to-end systém, což možná ani není). Kapitola 2, která představuje základní pojmy práce, působí neuceleným dojmem – uvádí řadu nepodstatných detailů (například datové typy implementace BLEU), za to neuvádí jiné detaily, které by mohly pomoci při orientaci v práci (např. že metrika BLEU dosahuje jen nízké korelace s lidským hodnocením, zatímco COMET velmi vysoké nebo že modely M2M100 a Comet jsou poměrně velké neuronové sítě). Odstavec o systému M2M100 je trochu zavádějící. Je to čtyři roky starý systém, ke kterému existují kvalitnější alternativy (NLLB, MADLAD-400). Místo toho informace, o kolik je lepší než předchozí systémy, by bylo ke čtenáři férovější uvést, o kolik je horší, než současné nejlepší modely (nebo než specializované dvojjazyčné modely, které jsou k dispozici pro většinu jazykových párů, které se vyskytují v obrázcích).</p>				

### Implementační část práce

lepší    OK    horší    nevyhovuje

Kvalita návrhu    ... architektura, struktury a algoritmy, použité technologie		X		
Kvalita zpracování    ... jmenné konvence, formátování, komentáře, testování	X			
Stabilita implementace	X			
<p>Práce je implementovaná v jazyce Python pomocí knihoven, které jsou v oboru standardem. Kód působí přehledně a je přiměřeně čitelný.</p>				

**Celkové hodnocení**    Velmi dobře  
**Práci navrhuji na zvláštní ocenění**    Ne

Datum 26.8. 2024

Podpis