

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jan Pavelka
Název práce Object layout in a 2D room based on text description
Rok odevzdání 2024
Studijní program Informatika
Specializace Umělá inteligence

Autor posudku Zdeněk Kasner Oponent
Pracoviště Ústav formální a aplikované lingvistiky

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X	X		
Splnění zadání		X		
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X	X		

Práce Jana Pavelky se zabývá detekcí rozložení objektů v místnosti na základě textového popisu. Motivací pro tento projekt jsou deskové RPG (role-playing game) hry, při kterých vypravěč potřebuje vygenerovat plán prostředí pro konkrétní situaci. Vstupem je věta v přirozeném jazyce, v tomto případě v češtině, např. „V místnosti stojí dvě čalouněné židle kolem bohatě prostřeného stolu.“. Výstupem jsou souřadnice místnosti a objektů v ní, které odpovídají zadanému rozložení a jsou vizualizovány jako 2D schéma.

Přístup, který student zvolil, se sestává ze dvou částí: 1) *lingvistická analýza*, při které jsou ze zadaného textu vyextrahovány jednotlivé objekty a relace mezi nimi, a 2) *generování rozmístění objektů*, při kterém je na základě zjištěných objektů a relací vygenerován odpovídající plán místnosti. Převážná část práce se věnuje lingvistické analýze, při které (narozdíl od druhé části) se student nemohl opřít o existující přístupy a musel vytvořit doménově specifické řešení s pomocí základních nástrojů. Hlavním takovým nástrojem je syntaktický parser UDPipe, pomocí kterého student extrahuje objekty a relace mezi nimi. Využívá k tomu syntaktické role relevantních slov ve větě – podstatná jména jsou potenciální objekty, kvantifikátory modifikují počet objektů a předložky určují relace mezi nimi. Student se přitom musel vypořádat s řadou specifik českého jazyka a s variabilitou potenciálních popisů. Na generování samotného plánu místnosti student využívá genetický algoritmus, jehož cílem je vytvořit řešení odpovídající zadaným podmínkám (s tím, že čím lépe řešení odpovídá podmínkám, tím je fitness řešení vyšší).

Zadání je netriviální už tím, že vyžaduje kombinaci dvou velmi odlišných paradigmat, a celkově ho považuji za splněné. I přesto mám i k práci určité výtky, které lze zároveň brát jako náměty k obhajobě:

- Práce byla testovaná na velmi malém vzorku dat. Ačkoliv jsem přesné množství použitých dat nenašel uvedené přímo v práci, z dat v příloze se dá usuzovat, že se jednalo o 7 konkrétních objektů, 3 typy místností, a něco mezi 10-20 vstupních popisů. Zatímco objekty i místnosti se v zásadě se liší jen tvarem a velikostí obdélníkového průmětu, více vstupů by umožnilo otestovat robustnost řešení. V práci (§2.1) je dokonce i navrženo, že další výstupy by šlo generovat pomocí ChatGPT, ale pouze pokud by šlo o anglický jazyk. S tím si dovoluji nesouhlasit – s využitím dostupných příkladů jako ukázky požadovaného výstupu by určitě šlo získat další příklady i v češtině.

(pokračování na další straně)

- V souvislosti s předchozím bodem v práci chybí jakékoliv kvantitativní vyhodnocení kvality navrženého řešení. Jediné vyhodnocení se nachází v kapitole 8, kde student komentuje několik konkrétních příkladů, ale ohledně robustnosti řešení nechává čtenáře na pochybách: „*The quality of the layout depends partially on luck. It was quite common that some objects fled away from the room (see figures 8.2(b), 8.3(a)) but this problem should be fixed in the very last version of the program.*“ Dostupná data navíc nejsou pro vyhodnocení v práci použita nijak systematicky, takže není jasné, jestli byly všechny dostupné vstupy v programu otestovány.
- Na několika místech jsou anekdotická pozorování (bez zmínky o velikosti vzorku a konkrétních výsledcích) zaměňovány za experimenty, např. v sekci 3.4.4: „*I performed few real-life experiments to choose the best approach. I pointed at real objects and asked people where the object is placed with respect to them, to me or other objects in the space. The experiments support the choice of the third approach.*“ nebo v kapitole 8: „*Both distance-based and intersection-based constraints seems to work (see figures 8.1(b), 8.2(b)) but the distance-based are significantly better.*“.
- Stejně tak jsou občas používány technické pojmy pouze ilustrativně, např. v sekci 5.1.4: „*In our approach, they are based on fuzzy logic which means that the result of their evaluation is not a binary value but a real number from [0, 1].*“ (tato věta by se obešla bez zmínky o fuzzy logice) nebo v sekci 3.4.5: „*From the candidates, the most probable one is chosen. It is not clear which candidate is the most probable one if there are multiple candidates. My choosing strategy is based on what I have seen in the data and my experience.*“ (kandidát je tedy nejspíš vybrán ne na základě pravděpodobnosti, ale heuristikou).

Oceňuji naopak to, že:

- Vzorek dat pochází od lidí zkušených v hraní deskových RPG her.
- Část práce věnující se lingvistické analýze je velmi podrobná a je zřejmé, že se jí student detailně zabýval.
- Práce navrhuje postupy ke zlepšení nedostatků práce do budoucna.

Textová část práce

lepší OK horší nevyhovuje

Formální úprava	... jazyková úroveň, typografická úroveň, citace			X	
Struktura textu	... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu		X	X	
Analýza			X		
Vývojová dokumentace			X		
Uživatelská dokumentace			X		

Práce je psaná v angličtině a její hlavní část má 48 stran. Práce je typograficky zdařilá a obrázky s příklady vhodně ilustrují okolní text. Rozsah textu mi přijde optimální a obsahem text pokrývá celou náplň práce.

(pokračování na další straně)

Výtky a náměty k obhajobě:

- Práce obsahuje velké množství překlepů a je zřejmé, že neprošla jakýmkoliv nástrojem pro kontrolu pravopisu. Příklady z úvodních kapitol: *mainstays of* (intro), *descriptions*, *whith* (§1.1), *charcters*, *machine translation* s malým M na začátku věty (§2.1), *initilaization* (§2.2), *for mor* (footnote 5), *sptial* (§3.4.3), *depenends* (3.4.4), *suare* (Figure 3.5), *perserve* (§3.4.4), *semmantic* (§3.5), *as a hole* (§3.4.5), atd.
- Čtení práce dále znesnadňuje velké množství stylistických chyb – neobratných výrazů, vět následujících českou gramatiku místo anglické, apod.: *besides it* (§1.1.1), *I presume to genralize this* (§3.2), *Defaultly, the face faces to the bottom of the map.* (§3.4.4), *which assigns an evaluation to the pair of objects* (§4.1), atd.. Častokrát je ve větě chybně množné číslo, případně v ní chybí čárky. Práci by určitě pomohlo využití nástroje na kontrolu stylu a gramatiky, např. u Grammarly mohou studenti MFF za určitých podmínek využívat i prémiovou verzi (a většinu problémů by vyřešila i základní verze).
- Odkazy na obrázky jsou psané s malým písmenem (*figure*) a občas jsou neúplné (*see figure* (intro), *see 3.1*) nebo naopak obrázky nejsou zmíněné v textu (Figure 2). Figure 2 má navíc obsah téměř identický s kódem na konci sekce 1.2.2, který ovšem není v plovoucím prostředí. V seznamu obrázků na konci práce se dvě položky jmenují *short* a tři *Results*, ze seznamu zkratk funguje jako odkazy zhruba polovina.
- Teorie je v práci průběžně míchána s implementací. Kapitola 1 obsahuje související práci, specifikaci problému, a ukázkou konkrétního výstupu z programu, kapitola 2 obsahuje popis zdroje dat a část návrhu řešení, kapitoly 3 a 5 obsahují obě návrhy a konkrétní detaily řešení lingvistické analýzy, a kapitola 4 obsahuje vše ohledně genetického algoritmu od specifikace problému, souvisejících řešení a detailů zvoleného přístupu. Kapitola 4 působí obzvlášť zmateně – zatímco sekce 4.1 obsahuje formální definici problému, sekce 4.3 už tuto definici nijak neuvažuje a popisuje genetický algoritmus neformálně v jednom odstavci. Porovnání algoritmů v sekci 4.2 je pak velmi stručné, u jednoho z algoritmů chybí jakékoliv detaily.
- Práce obsahuje pouze 10 referencí, z toho 6 na literaturu. Vzhledem k implementačnímu zaměření práce to nevidím jako velký problém, ale například u UDPipe a Korektoru mohla být kromě odkazu i citace na původní článek.

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu	... architektura, struktury a algoritmy, použité technologie	X	X		
Kvalita zpracování	... jmenné konvence, formátování, komentáře, testování		X	X	
Stabilita implementace			X		
<p>K práci je přiložený spustitelný kód v jazyce Python. Kód je vhodně členěný a je založen na modulech, které poskytují rozhraní k jednotlivým funkcím. Kód obsahuje minimum komentářů a žádné README, působí ale přehledně – díky jmenným konvencím jsou moduly sebedokumentující a každý z modulů obsahuje pouze několik řádků kódu. Věřím proto, že by projekt šel po doplnění dokumentace smysluplně dále rozšiřovat.</p> <p>(pokračování na další straně)</p>					

Spuštění kódu probíhá jednoduše: po příkazu `python main.py` může uživatel zadat větu, na základě které je mu vygenerovaný plánec. Po zadání věty bohužel kód v průběhu loguje minimum informací: v zásadě jde pouze o výstupy parseru UDPipe. To je problém především proto, že program v současné podobě není použitelný v reálném čase. Vygenerování plánu pro ukázkovou větu „*Vejdete do čtvercové místnosti. Kolem stolu stojí dvě židle.*“ trvalo na mém počítači cca 30 vteřin, což je poměrně přijatelné, zato pro jeden z dalších dostupných ukázkových vstupů „*Za dveřmi se nachází veliká místnost s postelemi po obou stranách. Celkem napočítáte 12 lůžek. Vždy mezi dvěma postelemi se nachází svícen připevněný ke stěně.*“ trvalo generování 44 minut, během kterých program neposkytoval žádné informace o hledaném řešení nebo zbývajícím čase. Pro tento vstup navíc výstup z parseru obsahuje jen první větu a plánec obsahuje velké množství vzájemně se překrývajících objektů, které příliš neodpovídají vzájemnému vstupu.

Problém řešený v práci je ale (při dostatečně obecném řešení) náročný a nedá se očekávat, že řešení bude plně odladěné pro použití v praxi. Oceňuji, že kód je srozumitelný, funkční a dává výstupy i pro uživatelem zadané věty.

Celkové hodnocení	Velmi dobře
Práci navrhuji na zvláštní ocenění	Ne

Datum 26.1.2024

Podpis