



Hodnocení vedoucího závěrečné práce

Vedoucí práce:	Ing. Ivo Lašek, Ph.D.
Student:	Tomáš Machorek
Název práce:	Dotazování nad relačními databázemi s pomocí velkých jazykových modelů (LLM)
Obor / specializace:	Umělá inteligence 2021
Vytvořeno dne:	6. června 2024

Hodnotící kritéria

1. Splnění zadání

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

Student vytvořil požadovanou aplikaci pro dotazování nad daty z relační databáze pomocí přirozeného jazyka s využitím dvou LLM modelů (Llama 3-8B a Mistral 7B). Samotná evaluace modelů se omezuje pouze na jejich základní spuštění nad datovými sadami bez porovnání ladění různých vstupních konfigurací. Sám student shrnuje, že výsledky běhu modelů nejsou ideální, ale chybí pokus o jakékoliv vylepšení jejich výsledků a porovnání různých přístupů. Při testování samotné aplikace oba modely vracejí v mnoha případech nevalidní SQL dotazy, ale v práci chybí jasně zdokumentované limitace vybraných modelů nebo jakékoliv podrobnosti o případných pokusech o jejich vylepšení. Zadání tedy bylo splněno v jeho minimální podobě.

2. Písemná část práce

62 / 100 (D)

Úvodní kapitoly obsahují učebnicový přehled základních pojmů z oblasti neuronových sítí a velkých jazykových modelů. Samotná rešeršní část se omezuje na popis pouze dvou modelů použitých v samotné práci bez porovnání s dalšími možnostmi, které jsou momentálně k dispozici. Původní očekávání od rešeršní části bylo, že student prozkoumá celou škálu modelů a na základě dostupných informací vybere 2-3 nejvhodnější. Toto zdůvodnění, proč byly vybrány zrovna tyto dva modely (Llama a Mistral) a srovnání s dalšími možnostmi v práci téměř úplně chybí. Část zabývající se výběrem datových sad pro evaluaci je adekvátní a vybrané datové sady se také zdají být vhodné pro daný případ. Nejslabší částí práce je sekce týkající se ladění efektivity a evaluace modelů. Není úplně zřejmé, jaké přístupy student přesně použil pro vylepšení modelů. U samotné

evaluace chybí srovnání výsledků s různými konfiguracemi modelů, případně různými typy promptů.

3. Nepísemná část, přílohy 70/100 (C)

Samotná aplikace je poměrně minimalistická, ale splňuje zadání. Pozitivně hodnotím vhodné využití hotových komponent (např. pro vizualizace), které přispívá k jednoduchosti aplikace. Slabší částí je samotné experimentování s modely popsané výše.

4. Hodnocení výsledků, jejich využitelnost 70/100 (C)

Aplikace umožňuje experimentování s jednoduššími verzemi open source velkých jazykových modelů a jako taková poskytuje užitečný framework pro jejich další zkoumání. Samotná experimentální část vzhledem ke svému omezenému rozsahu moc nových poznatků nepřináší.

5. Aktivita studenta

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

Student nevyužil možnosti pravidelných konzultací a celou práci zaslal k revizi až těsně před jejím odevzdáním. To bohužel vede k výše uvedeným výhradám, co se rozsahu splnění zadání týče, kterým se dalo předejít.

6. Samostatnost studenta

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

Student pracoval samostatně v průběhu celého semestru, nicméně některé výstupy obsahovaly zbytečné chyby (např. syntaktická chyba v kódu odevzdané aplikace zabraňující jejímu jednoduchému spuštění). Práci by prospěla důkladnější revize.

Celkové hodnocení 65/100 (D)

Po formální a věcné stránce nemám k práci závažné výhrady, nicméně celkově práce budí dojem, že byla ušita horkou jehlou na poslední chvíli, což bohužel snižuje kvalitu celkového výstupu. Zejména část experimentování a ladění modelů by si zasloužila více pozornosti.

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.