

Posudek oponenta diplomové práce

Autor práce: **Milan Horinek**

Název práce: **Generování jednotkových testů s využitím LLM**

Obsah práce

Jedná se o rozsáhlé dílo, které obsahuje několik vzájemně na sebe navazujících částí, a to celé je výborně technologicky zvládnuté. Diplomant sám naprogramoval několik málo programů v Pythonu, které ale mistrně zařadil do celé sekvence „příprava scénářů – generování testů – čištění testů – spuštění testů – vyhodnocení výsledků“. Pro zautomatizování této činnosti diplomant vybudoval platformově nezávislý stack, který do budoucna umožní vedoucímu práce provádět další experimenty podobného typu téměř automaticky. Z mého osobního pohledu velmi oceňuji, že pro ověření diplomát použil systém TbUIS, tak jak byl zamýšlen. Práce má logickou strukturu, text obsahuje všechny informace potřebné k pochopení problému i k pochopení vytvářeného díla. Diplomant přehledně popsal všechny vytvořené části a zdůvodnil i výběr použitých prostředků.

Svým rozsahem je textová část práce výrazně nadprůměrná (92 stran textu plus 16 stran příloh a seznamů).

Kvalita řešení a dosažených výsledků

V ZIP souboru se nalézají všechny soubory potřebné pro vyzkoušení všech částí práce a tyto soubory jsou v práci popsány. Struktura adresářů je uspokojivě popsána. Zdrojové kódy v Pythonu jsou dobře čitelné a dostatečně komentované.

Jak již bylo zmíněno, těžiště práce je ve vytvořeném stacku. Celá jeho funkcionalita byla oponentovi osobně předvedena a nezbyvá než pochválit propracovanost všech částí. Dobrý obrázek o kvalitě výsledků si lze udělat již pouhým prohlédnutím práce na stranách 57 až 81.

Formální úroveň

Formální úroveň práce je bohužel nejslabším článkem práce. Práce je sice vysazena v LaTeXu, ale evidentně jí chybí závěrečná korektura. Nalezl jsem celkem 179 drobných problematických míst (seznam zaslán diplomantovi). Typicky se jedná o překlepy, typografické prohřešky, chybějící interpunkci a zejména neshodu podmětu s přísudkem, kdy „testy“ a „modely“ jsou povětšinou považovány za životné. Nutno, ale dodat, že zmíněné věci nesnižují pochopitelnost. Naopak velmi oceňuji četné obrázky a úseky kódu, které jednoznačně přispívají k pochopení problému. To, co naopak pochopitelnost snižuje, je poměrně volné používání odborné terminologie z oblasti testování a míchání či zaměňování termínů.

Práce s literaturou

Diplomant udává celkem 62 zdrojů v naprosté většině elektronických, všechny jsou v textu práce použity. Použité zdroje jsou aktuální, odpovídající pro řešení práce a dokladují, s jakou šíří problematiky se musel diplomant seznámit.

Splnění zadání

Zadání považuji za zcela splněné.

Dotazy k práci

1. V práci zmiňujete problémy, které měly použité LLM s generováním čekání na událost. Proč jste místo Selenium Library nepoužil modernější Browser Library, kdy by tento problém zcela odpadl?
2. Jakou verzi Robot Frameworku jste používal? Myslíte si, že je práce citlivá na případnou změnu verze RF?
3. Časová náročnost lokálních modelů (str. 85) je překvapivě vysoká. Jaká byla v tomto kontextu časová náročnost (alespoň odhadem) proprietárních modelů, u kterých tamtéž uvádíte pouze cenu?

Vytvořené dílo považuji za velmi kvalitní s přesahem do vědecké činnosti vedoucího práce. Výše zmíněné prohřešky jsou vzhledem k rozsahu a kvalitě celku zcela minoritní. Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.