

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jan Kleprlík
Název práce Performance and Usability Improvements for Data Lineage Analysis of C# Programs
Rok odevzdání 2024
Studijní program Informatika **Studijní obor** Softwarové systémy

Autor posudku David Bednárek **Role** Oponent
Pracoviště KSI MFF UK

Text posudku:

Náplní posuzované práce je modul pro jazyk C#, který je součástí rozsáhlého komerčního produktu Manta Flow pro data lineage analysis. Tento modul byl vytvořen v rámci předchozích studentských prací, obsahem této práce je odstranění chyb, zlepšení výkonu a několik doplňků, z nichž nejzajímavější je podpora pro analýzu kódu v C#, volaného z jiných prostředí jako databázové či webové servery. Celý systém Manta Flow má komplikovanou architekturu podporující moduly implementované v různých jazycích, které si vzájemně posílají komplikovaná grafová data rozličnými komunikačními kanály - vedlejší součástí této práce byla i výměna Google Protocol Buffers za XML v jednom z těchto komunikačních kanálů. Celkově lze konstatovat, že softwarová část práce sice neobsahuje žádné nové algoritmy a není ani nijak zvlášť objemná, jde však o dílo komplikované zapojením do rozsáhlého systému, který se vyvíjel dávno před vstupem autora této práce a má být dlouhodobě udržovatelný.

Textová část je členěna způsobem, který na první pohled vyhovuje - je zde vysvětlení systému Manta Flow, popis funkce analyzátoru pro C#, formulace jednotlivých cílů práce i kapitola popisující implementaci. Všechny tyto části však mají tentýž prezentační problém: Text neustále přebíhá mezi vysokoúrovňovými architekturními záležitostmi a nízkoúrovňovými detaily typu ukázky CIL kódu nebo rozhraní tříd. Nadbytek detailů zatěžuje čtenáře, přičemž v některých případech ani není na první pohled jasné, zda je popisována třída z analyzovaného kódu nebo třída implementující analyzátor. Pro snazší porozumění textu by bylo lepší, kdyby byly technické detaily nějakým způsobem vyčleněny stranou a jednotlivé sekce měly své úvody shrnující základní fakta.

Komplikovanější části textu jsou psány způsobem, který zřejmě odpovídá firemní kultuře, u diplomových prací je ale neobvyklý - práce se vyhýbá jakýmkoliv matematickým formalismům a prezentace detailů je založena spíše na různých formách diagramů, které sice vzhledově odpovídají zvyklostem softwarového inženýrství, jejich význam ale nemusí být jednoznačný. Výhodou tohoto přístupu je srozumitelnost i pro matematicky nevzdělané čtenáře, nevýhodou je však nedostatečná přesnost, někdy i zbytečně dlouhý popis, který by se dal zkrátit odkazem na vhodnou matematickou strukturu. S absencí formalismu zřejmě souvisí i absence odkazů na literaturu z oblasti překladačů či statické analýzy kódu, většina ze 40 prvků seznamu literatury je tvořena manuály k rozličným technologiím.

Práce evidentně splnila cíle, které plynou z jejího využití v daném komerčním prostředí. Účelem diplomových prací by však měl být i určitý příspěvek v akademické oblasti, tedy

pokus o zobecnění problému nebo alespoň popis problému a řešení tak, aby z něj mohli mít užitek i čtenáři mimo úzký okruh architektů a implementátorů daného produktu. V tomto akademickém smyslu tato práce téměř nic nepřináší.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhuje na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 1. February 2024

Podpis