



Posudek vedoucího diplomové práce

Bc. Jiří Fatka

Automatická statistická a metrická analýza zdrojového kódu se zaměřením na jazyk ANSI C

Diplomová práce Bc. Jiřího Fatky se orientuje na velmi zajímavou a zvláště v poslední době atraktivní problematiku posuzování autorství děl, v tomto případě softwarových. Cílem jeho práce bylo navrhnout a implementovat nástroje dovolující na základě statistické a metrické analýzy zdrojových kódů programů posoudit a objektivně vyhodnotit podobnost programů a stanovit, nakolik je daný program derivátem jiného, vzniklým nevýznamnou úpravou s cílem nikoliv vylepšit původní program, ale zakrýt původ použitého zdroje.

Jedná se bezesporu o nelehký úkol. Autor se ho zhostil na výbornou a dodané řešení lze považovat za vynikající, velmi hodnotné implementační dílo, kterým autor mimo jakoukoliv pochybnost prokázal nejen excelentní programátorské dovednosti, ale také značný analytický potenciál. **Zadání práce bylo splněno ve všech bodech.**

Předložená práce je poměrně rozsáhlá: Zdrojový kód programového řešení obsahuje 12487 řádek v jazyce C++ (do toho nejsou započítány hlavičkové soubory a testovací zdrojové soubory). Průvodní dokument pak má včetně příloh 77 stránek textu.

Práce působí celkově vynikajícím dojmem – jedná se o značný kus poctivé programátorské práce v klasickém slova smyslu, autor celé řešení naprogramoval sám a „na zelené louce“, bez použití hotových knihoven, komponent, ap. Také výzkum představoval v případě této úlohy svébytný a nikoliv triviální problém.

Vyvinutá aplikace funguje perfektně a zcela podle představ zadavatele. Je snadno přenositelná, nezávislá na cílové platformě a především snadno integrovatelná do validačního frameworku běžícího na serveru `amos.fav.zcu.cz`. Program zároveň pracuje mimořádně rychle, analýza několika set prací doběhne na průměrně výkonném osobním počítači v řádu jednotek sekund.

Implementační část autorovy práce se skládá ze dvou hlavních součástí, které jsou objemově značně nesouměřitelné: První částí je samotný analyzátor zdrojového kódu v jazyce ANSI. Jeho vstupem jsou zdrojové soubory analyzovaného projektu, výstupem je pak tzv. otisk, tj. soubor (textový) s řadou číselných hodnot, které představují souřadnice obrazu projektu v metrickém prostoru analýzy. Druhou – výrazně menší částí – je pak program, který načítá otisky jednotlivých posuzovaných projektů a posuzuje jejich vzdálenosti v obrazovém prostoru, přičemž výsledkem jeho činnosti je procentuální (bez hlubšího syntakticko-sémantického významu) stanovení shody mezi posuzovanými projekty.

Autor použil k tvorbě programového vybavení jazyk C++ za využití vývojového prostředí GCC v operačním systému GNU/Linux. Aplikace je primárně cílena na platformu UNIX/Linux s ohledem na potřebu integrace do validačního frameworku, nicméně je napsána zcela přenositelně a překlad zdrojových kódů je možný bez úprav pomocí prakticky jakéhokoliv překladače jazyka C++.

Ze způsobu zápisu zdrojového kódu aplikace je patrná autorova značná zkušenost a profesionalita. Možnosti jazyka C++ využívá na samou hranici možností, které tento jazyk poskytuje, tj. včetně šablon tříd, což místy vede až ke konstrukcím, které by mohly méně fundovaného programátora stát život. Kultura zdrojového kódu je však úzkostlivě dodržována, kód je zapsán zcela systematicky, za dodržení všech pravidel a doporučení. Komentáře jsou vzorné, objektová dekompozice taktéž.

Překlad a sestavení aplikace ze zdrojových kódů je bezproblémové a přímočaré – autor poskytuje pro obě dvě části dobře promyšlený makefile. Vyvinuté aplikaci nelze vytknout téměř nic, mj. také proto, že autor docházel pravidelně na konzultace a svědomitě plnil pokyny vedoucího práce.

Testování aplikace proběhlo v dostatečném rozsahu, na reálných datech nasbíraných vedoucím práce v průběhu několika let při výuce předmětu Programování v jazyce C a dosažené a prezentované výsledky jsou na velice dobré úrovni (možno říct až nad očekávání dobré). Dosahovaná spolehlivost identifikace derivátů zdrojových kódů je vynikající.

Průvodní dokument je napsán obstojnou technickou češtinou a vysázen v \LaTeX u. Jeho grafická úroveň je vynikající a celkově působí velmi harmonickým dojmem. Text práce je srozumitelný a čtivý, autorovy vyjadřovací schopnosti jsou velice dobré. V dokumentu se sice vyskytují gramatické chyby (autor bohužel při dokončování textu práce evidentně spěchal) a některé anakolutické vazby, případně nepřesná či zavádějící vyjádření, ale celkově zásadně negativně neovlivňují sdělovací a popisnou hodnotu. Anglický abstrakt práce by každopádně snesl více péče a korekturu.

Text je dobře logicky strukturován, poměr jednotlivých částí textu je vyvážený a vhodně pokrývá vykládaná témata od úvodních elementárních informací až po implementační detaily. Autorovo odborné vyjadřování je vyzrálé a nelze mu nic vytknout.

Autor vhodně používá zvýrazňovací řez písma, text je doplněn několika málo obrázky (mohlo by jich ale být více) ve vysoké kvalitě, rozsáhlou řadou výpisů zdrojového kódu a tabulkami, které žádoucím způsobem obohacují text. Sazba normovaných sledů znaků je zvládnuta bez problémů.

Provedení citací je správné, rozsah i výběr literatury je však poněkud méně obsáhlý než by se dalo očekávat. Nicméně autor nastudoval vhodné zdroje a byť je většina z nich pouze v elektronické podobě, nelze to autorovi nijak vyčítat, protože v předmětné problematice je tento způsob publikace naprosto obvyklý.

Závěrem: Jedná se o kvalitní, komplexní a robustní softwarové dílo značného rozsahu. Autor se zadaného úkolu zhostil na výbornou a výsledek představuje špičkový softwarový produkt vhodný k nasazení do praxe. Při řešení složitých problémů spojených s vývojem aplikace autor jednoznačně prokázal vynikající inženýrské schopnosti a značný analytický potenciál.

Všechny výše uvedené výhrady jsou víceméně formálního charakteru, a proto práci rozhodně **doporučuji k obhajobě** a hodnotím klasifikačním stupněm

„výborně“.



Ing. Kamil Ekštejn, Ph.D.
KIV FAV ZČU

V Plzni dne 5. 6. 2012

Doplňující otázky: Nemám žádné doplňující otázky.