

Prof. RNDr. Mária Lucká, PhD.

Fakulta matematiky, fyziky a informatiky

Univerzity Komenského

Mlynská dolina F1

842 48 Bratislava

### **Posudok oponenta habilitačnej práce**

Autor:	<b>Ing. Patrik Hrkút, PhD.</b>
Téma:	<b>Metódy analýzy zdrojového kódu</b>
Odbor:	<b>Aplikovaná informatika</b>

#### 1. Aktuálnosť témy a štruktúra práce

Habilitačná práca Ing. Patrika Hrkúta, PhD. s názvom „Metódy analýzy zdrojového kódu“ sa zaoberá zaujímavou a v súčasnosti veľmi aktuálnou témou hľadania podobnosti zdrojových kódov. Práca má formu súboru vybraných publikovaných článkov, ktoré sú uvedené oddelene v samostatných kapitolách s krátkym sprievodným textom. Zdrojové texty majú na rozdiel od bežných textov jednoduchšiu formu a význam ich spracovania súvisí s viacerými praktickými aplikáciami, akými sú napr. zisťovanie kvality zdrojového kódu, zisťovanie chýb v zdrojovom kóde, či určovanie plagiátorstva. Významnosť zvolenej témy umocnila dostupnosť internetu a tým aj možnosť neoprávneného použitia dostupného softvéru.

#### 2. Vytýčenie cieľov práce a ich splnenie

Cieľom práce bola tvorba softvérového nástroja vhodného na vyhľadávanie podobnosti zdrojových kódov, ktorý by bol použiteľný aj na zisťovanie plagiátorstva pri tvorbe a kontrole zdrojových kódov v rámci zvolenej skupiny jeho tvorcov. Aj keď u nás a vo svete existuje viacero efektívnych nástrojov na zisťovanie plagiátorstva týkajúceho sa hľadania podobnosti publikovaných textov, existencia nástrojov na zisťovanie podobnosti vytvorených zdrojových kódov nie je taká bežná a pritom je neodmysliteľná napr. dôležitá pri kontrole samostatnosti tvorby riešení napr. ročníkových prác študentov informatických fakúlt. Cieľom práce bolo hľadanie možnosti kontroly plagiátorstva vo veľkých dátových súboroch a teda hľadanie podobnosti viacerých súborov súčasne, čo sa habilitantovi aj podarilo.

#### 3. Hodnotenie obsahovej stránky

Práca je veľmi dobre logicky vystavaná. Okrem úvodu a záveru je rozdelená do šiestich nosných kapitol, ktoré načrtávajú genézu problematiky návrhu a riešenia úloh, ktoré boli potrebné pri tvorbe antiplagiátorského systému. Každá kapitola najprv stručne opisuje riešenú problematiku,

ktorá je následne podrobne riešená v priložených publikáciách. Všetky kapitoly, okrem štvrtej a siedmej, obsahujú po jednom vedeckom článku dokumentujúcom riešenie opísaného problému, pričom spomínané dve kapitoly obsahujú po dva vedecké články. Autor habilitačnej tak chronologicky načrtáva a postupne prezentuje všetky problémy, ktoré bolo potrebné vyriešiť pri tvorbe ním vytvoreného antiplagiátorského systému. Začína hľadaním podobnosti a rozdielov spracovania bežného textu a zdrojového textu s cieľom aplikovať možné postupy a algoritmy z oblasti spracovania prirodzeného jazyka. Kvôli práci s veľkým množstvom dát hľadá vhodnú dátovú štruktúru tak, aby umožňovala spracovanie veľkého množstva dát, ktorou je abstraktný syntaktický strom (AST). Pre zjednodušenie procesu vyhľadávania autor práce využíva metódu zhukovania, aby tým vylúčil zo spracovania úplne odlišné kódy a znížil objem prehľadávaných dát. Venuje sa tiež hľadaniu vhodnej metódy zhukovania pre zvolený typ problému. Pretože prehľadávanie stromov je výpočtovo náročná procedúra, autor pri svojej práci využíva rôzne metódy hešovania. Pri hľadaní efektívnej reprezentácie zdrojového kódu ukázal, že na určenie miery podobnosti je výhodnejšie použiť charakteristické vektory ako linearizovať AST stromy, nakoľko je možné takto získať presnejšie výsledky a metóda je výpočtovo rýchlejšia. Práca sa venuje aj problémom škálovateľnosti navrhovaných riešení. Pri experimentovaní porovnáva svoje riešenie s existujúcimi riešeniami, akými sú softvérové systémy napr. MOSS a JPlag. Podarilo sa mu nájsť riešenie, ktoré je porovnateľné a v niektorých parametroch lepšie ako existujúce riešenia. Jeho výhodou je to, že je modulárne a rozšíriteľné tak, aby ho bolo možné rozšíriť tak, aby bolo použiteľné pre viaceré programovacie jazyky. V súčasnosti však existujú viaceré systémy na hľadanie podobnosti zdrojových kódov, ktoré využívajú metódy umelej inteligencie a sú založené napríklad na metódach strojového učenia, neurónových sieťach, metódach hlbokého učenia. Pokladám za nedostatok, že v práci tieto metódy nie sú spomenuté, a to napriek tomu, že s nimi priamo habilitant nepracoval. Celkove by kvalite práce bolo prospelo, keby obsahovala všeobecný úvod s prehľadom existujúcich metód a až potom sa venovala opisu vlastných riešení. Pozitívne ale hodnotím záver práce, ktorý naznačuje možné zlepšenia existujúceho systému a ukazuje možné smery ďalšieho výskumu. Celkove hodnotím prácu ako veľmi dobrú, o kvalite dosiahnutých výsledkov svedčia aj kvalitné publikácie a hodnota vytvoreného antiplagiátorského systému.

#### 4. Otázky k habilitačnej práci

- a) Vo vašej práci spracováate veľké objemy dát. Skúšali ste použiť niekde paralelné metódy spracovania, prípadne kde vidíte možnosti ich použiť a ako?
- b) Pri hľadaní miery zhody zdrojových kódov hľadáte iba presné zhody charakteristických vektorov. Skúšali ste aplikovať pradedpodobnostný prístup na nájdenie podobných vektorov? Ako úspešne?
- c) Mohli by ste povedať, či poznáte nejaké antiplagiátorské systémy založené na metódach umelej inteligencie? Akým spôsobom pracujú? Porovnávali ste vaše riešenia aj s takýmito systémami? Dala by sa vo vami navrhovaných riešeniach použiť umelá inteligencia a ak áno, tak na riešenie ktorých problémov?

## 5. Celkový záver

- Na záver konštatujem, že predložená práca spĺňa požiadavky kladené na habilitačnú prácu a odporúčam ju k obhajobe.
- Na základe preštudovania habilitačného spisu v časti Kritériá na získanie titulu docent na FRI ŽU v Žiline v odbore aplikovaná informatika priradeného k odboru informatika konštatujem, že habilitant nielen spĺňa, ale vo väčšine kritérií vysoko prevyšuje väčšinu požadovaných minimálnych kritérií potrebných k schváleniu žiadosti o habilitáciu.
- Po úspešnej obhajobe habilitačnej práce navrhujem, aby mu bol udelený titul docent v odbore Aplikovaná informatika.

V Bratislave, dňa 11.11.2022

Prof. RNDr. Mária Lucká, PhD.