

# Oponentský posudok habilitačnej práce

**Ing. Patrik Hrkút, PhD.**

Téma habilitačnej práce: **Metódy analýzy zdrojového kódu**

Odbor habilitačného konania a inauguračného konania: **aplikovaná informatika**

Pracovisko: **Fakulta riadenia a informatiky, Žilinská univerzita v Žiline**

Posudzovaná habilitačná práca sa zameriava na detailné riešenie komplexného problému odhaľovania plagiátov v zdrojových kódach. Aj keď táto téma nie je výskumne nová, je stále aktuálna, s dostatočným priestorom na nové vedecké prínosy a spolu so silným aplikačným zameraním predloženej habilitačnej práce sa výborne hodí do odboru aplikovaná informatika.

Autor po krátkom úvode postupne predstavuje jednotlivé publikované práce v predmetnej problematike, ktoré sú vhodne zostavené tak, aby prezentovali dôležité fázy postupu a hlavné úlohy potrebné pri tvorbe finálneho systému na detekciu plagiátov zdrojového kódu. Jednotlivé prezentované práce prinášajú zaujímavé vedecké ale aj aplikačné výsledky, nakoľko výsledný systém bol implementovaný a je aktívne využívaný na domovskej fakulte. Takéto prepojenie vlastných výskumných výsledkov s ich dotiahnutím cez implementačnú fázu až k nasadeniu a prevádzke v reálnom prostredí veľmi oceňujem.

V nasledujúcej časti oponentského posudku sa postupne stručne vyjadrujem k jednotlivým prezentovaným prácam, vrátane priebežných otázok. Jednotlivé prezentované publikácie majú často iné poradie autorov než je tomu vo zverejnenej publikácii. Chýba tiež informácia o percentuálnom podiele autorov, ale tieto informácie som našiel v habilitačnom spise.

Článok 1 je spracovaný pomerne povrchno, teoretická analýza miestami mohla byť lepšie odôvodnená pomocou experimentálnych výsledkov iných autorov. Podobne je to u článku 2, kde autori predkladajú hypotézu, ale nie analýzy experimentálnych výsledkov, takže nemožno objektivizovať niektoré ich tvrdenia. Zišli by sa aspoň minimálne informácie o výsledkoch experimentov iných autorov, o ktoré sa opierajú kľúčové tvrdenia, ako napr. preferencia AST štruktúry. Otázkou tiež vyvoláva záver článku v ktorom sa zrazu spomínajú informačné systémy pre podporu dopravy, aj keď článok analyzuje metódy na odhaľovanie plagiátov.

- Na str. 34, v poslednej vete kapitoly 4 tvrdíte, že príkladom metódy založenej na fuzzy logike je SOM. Môžete toto tvrdenie zdôvodniť?

Článok 3 prináša vlastný experiment s vhodne navrhnutou metodikou a spôsobom vyhodnotenia. Mám však k nemu dve menšie technické otázky:

- Použitie operácie zjednotenia v súvislosti s vektormi je dosť neštandardné. Čo presne to znamená?
- V časti 4.3 na str. 45 má byť v menovateľoch oboch vzťahov pre výpočet podobnosti naozaj minimum? Na to aby bola výsledná hodnota celého zlomku nie väčšia ako 1 by tam malo byť skôr maximum.

Nasledujúca kapitola už popisuje vlastný antiplagiátorský systém založený na reprezentácii typu abstraktné syntaktické stromy (AST) pomocou charakteristických vektorov, ktoré sa ukládajú spolu s ďalšími dôležitými informáciami do relačnej databázy. To podporuje škálovateľnosť

navrhnutého riešenia spolu s ďalšou originálnou vlastnosťou, ktorou je aplikácia inkrementálneho zhlukovania na postupne pridávané zdrojové kódy. Miera presnosti na str. 52 je definovaná neštandardným spôsobom. Prvá veta definície navyše neobsahuje informáciu o druhom činiteli pomeru (menovateľ).

Články 4 a 5 podrobnejšie prezentujú návrh vlastného antiplagiátorského systému, pričom každý rozpisuje detailnejšie iné vybrané časti a aspekty vlastného riešenia. Článok 4 obsahuje čiastočnú verifikáciu formou manuálneho porovnania vybranej množiny výsledkov a stručného porovnania s výsledkami MOSS systému. Experimentálna časť tohto článku by si zaslúžila podrobnejšie spracovanie. Článok 5 má experimentálnu časť spracovanú detailnejšie. V oboch prípadoch však platí, že sa len v obmedzenej miere využívajú štandardné metriky definované autorom ako úplnosť a presnosť.

Článok 6 sa podrobne venuje zhlukovacej časti systému a článok 7 opisuje a experimentálne overuje niekoľko vylepšení inkrementálneho zhlukovania, zásluhou ktorých autori dosiahli až 30-násobné zrýchlenie výsledného programu oproti jeho základnej verzii naprogramovanej pomocou dostupných .NET knižníc. Článok 6 bol publikovaný v kvalitnom časopise na hornej hranici Q3 v JCR. V SJR je tento časopis v rámci jednej z kategórií v Q2.

- Pri ladení hyperparametrov k-means ste uvažovali aj nad inou ako Euklidovskou vzdialenosťou? Ako ste riešili inicializáciu centroidov? Skúšali ste nejaké opatrenia na to, aby ste predišli nevhodnej voľbe centroidov?

Jednoznačne najkvalitnejšie spracovaným je článok 8, publikovaný v renomovanom časopise IEEE Access (aktuálne Q2 podľa JCR a Q1 podľa SJR). Obsahuje rozšírenú analýzu súčasného stavu, detailnú prezentáciu vlastných algoritmov aj s konkrétnymi príkladmi a podrobné experimentálne vyhodnotenie, ktoré má prevažne kvalitatívny charakter, ale dáva dobrý prehľad o vlastnostiach vytvoreného systému a jeho kvalitách v porovnaní s dvoma najčastejšie používanými systémami vo svete.

Po formálnej stránke je práca napísaná veľmi kvalitne, aj keď sa v nej občas vyskytnú gramatické chyby. Sprievodný text vhodne uvádza a sumarizuje hlavné výsledky zaradených publikovaných prác.

Záverom konštatujem, že habilitačná práca Ing. Patrika Hrkúta, PhD. je na veľmi dobrej úrovni a z informácií uvedených v habilitačnom spise vyplýva, že Ing. Patrik Hrkút, PhD. spĺňa všetky kritériá stanovené Fakultou riadenia a informatiky Žilinskej univerzity v Žiline pre získanie vedecko-pedagogického titulu docent. V zmysle vyššie uvedeného preto

#### **odporúčam,**

aby bol Ing. Patrik Hrkút, PhD. vymenovaný za docenta v odbore aplikovaná informatika.

V Košiciach, 9. novembra 2022

prof. Ing. Ján Paralič, PhD.  
Katedra kybernetiky a umelej inteligencie  
Technická univerzita v Košiciach  
Letná 9, 042 00 Košice