

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Improving the plagiarism detection in BRUTE
Jméno autora:	Tomáš Votroubek
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Ing. Petr Pošík, Ph.D.
Pracoviště vedoucího práce:	Katedra kybernetiky, FEL ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání <i>Hodnocení náročnosti zadání závěrečné práce.</i>	průměrně náročné
--	-------------------------

Splnění zadání <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	splněno
Testování nad zcela reálnými daty se ukázalo být příliš ambiciózním: data obsahovala jen málo potvrzených plagiátů (TP), téměř jistě obsahovala relativně hodně nedetekovaných plagiátů (FN). Jakékoli porovnání s tímto chybným stavem světa by bylo samo chybné. Proto student využil sice reálná data, ale plagiáty z nich vytvořil sám, a to jasně definovanými, automaticky proveditelnými a opakovatelnými úpravami.	
Nad rámec zadání student implementoval dva vizualizační nástroje využívající výstupy vytvořeného detektoru, které mohou být využity jako pomůcka k rozhodování učitele, zda se jedná o plagiát nebo ne.	

Aktivita a samostatnost při zpracování práce <i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	A - výborně
Student pracoval samostatně a pravidelně docházel na konzultace, kde byl vždy dobře připraven. Bylo zřejmé, že na tématu pracuje průběžně.	

Odborná úroveň <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	A - výborně
Práci považuji odborně za velice zdařilou. Student nastudoval principy existujících detektorů plagiátů, i principy hodnocení detektorů pomocí Precision-Recall křivky. Experimenty nad vytvořenými datovými sadami jsou pečlivě zpracované.	

Formální a jazyková úroveň, rozsah práce <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	A - výborně
Po formální stránce nemám k práci významnější výhrady. Práce je logicky členěna a jednotlivé části na sebe dobře navazují. Výsledky experimentů jsou prezentovány především formou grafů, což považuji za adekvátní.	

Výběr zdrojů, korektnost citací <i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	B - velmi dobře
---	------------------------

Student řádně cituje převzaté prvky, odkazuje se na 20 jiných pramenů. Trochu mi v práci chybí srovnání výsledků studie s výsledky jiných studií porovnávajících detektory plagiátů známých z literatury.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Výsledky porovnání ukazují, že student vytvořil detektor, který je do značné míry nezávislý na programovacím jazyku, a přesto se jeho výsledky blíží k výsledkům detektorů, které znalosti o konkrétní prog. jazyku používají. Navíc je detektor řádově rychlejší, než doposud používaný Perlovský modul Text::Similarity.

I jako neznalý jazyka F# se opovažuji soudit, že styl, jímž student detektor implementoval, je velmi čistý a čitelný.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Celkové hodnocení práce zakládám především na následujících aspektech:

- Student definoval a vytvořil postupy pro vytvoření testovacích datových sad obsahujících různé typy změn strojového kódu.
- Pomocí těchto testovacích datových sad provedl porovnávací studii existujících detektorů plagiátů, včetně aktuálně používaného modulu Text::Similarity.
- Navrhl a implementoval vlastní detektor plagiátů, který je aplikovatelný jak na text, tak na kódy v různých programovacích jazycích, přičemž se detektor svou účinností blíží výsledkům specializovaných detektorů a v rychlosti předčí aktuálně používaný Text::Similarity.
- Nad rámec zadání implementoval dva podpůrné vizualizační nástroje.

Předloženou závěrečnou práci proto hodnotím klasifikačním stupněm **A - výborně**.

Datum: 29.1.2018

Podpis: Ing. Petr Pošík, Ph.D.