FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

# Hodnocení vedoucího závěrečné práce

| | |
|---|---|
| **Vedoucí práce:** | Ing. Ondřej Guth, Ph.D. |
| **Student:** | Andrey Olos |
| **Název práce:** | Zranitelnost Javy na odepření služby při vyhledávání s regulárními výrazy |
| **Obor / specializace:** | Bezpečnost a informační technologie |
| **Vytvořeno dne:** | 4. srpna 2023 |

## Hodnotící kritéria

### 1. Splnění zadání

▸ [1] **zadání splněno**
[2] zadání splněno s menšími výhradami
[3] zadání splněno s většími výhradami
[4] zadání nesplněno

Téma hodnotím jako náročné vzhledem k nutnosti seznámit se s koncepty, které nejsou v bakalářském (ani jiném) programu FIT vyučované: útok ReDoS, algoritmy na vyhledávání s regulárními výrazy (a jejich rozšířeními) používané v praxi. Oborově téma přesahuje z bezpečnosti do teoretické informatiky a asi i do softwarového inženýrství.

Student zadání splnil, navíc oproti zadání provedl popis (reverzní inženýrství) stávající implementace a návrhy na opravy (formou popisu, nikoli zdrojového kódu).

### 2. Písemná část práce 84 /100 (B)

Text práce působí lehce rozporuplným dojmem. Po věcné stránce a z hlediska prezentace hlavních výsledků je vše v pořádku (až na drobnosti, viz níže). Práce se zdroji dělá smíšený dojem: v práci se nevyskytuje případ porušení citační etiky; uvádění citovaných zdrojů je v některých případech nepěkné (viz seznam níže), občas nesprávné; použitých zdrojů mohlo být více a v kapitole 3 to působí jako nejasné odlišení vlastních poznatků od převzatých (nedošlo ale k vydávání cizích výsledků za vlastní). Kvalita textu rozhodně trpí nedostatkem času na závěrečné psaní a začištění: toto je nejslabším aspektem práce.

Kapitolu 2, která je klíčová ke splnění zadání a která obsahuje hlavní výsledky, hodnotím velmi dobře, zvlášť dobře promyšlenou (a následně splněnou) metodiku. Kapitoly 3 a 4 jsou oproti zadání (a naší původní dohodě) navíc, jde čistě o iniciativu studenta. V kapitole 3 jde z velké části o převzetí jednoho z citovaných zdrojů a navázání na jeho výstupy, zčásti jde o vlastní výsledky studenta (reverzní inženýrství implementace v OpenJDK). Zde je jen škoda, že student nestihl zjistit, na základě čeho se nastavuje parametr

posIndex řídící de facto časovou náročnost vyhledávání, to ale není výtka. Kapitola 4 obsahuje obecné návrhy na vyrovnání se zranitelností.

K převzatým zdrojům: je škoda, že nebylo zpracováno zdrojů více. Další zdroje (ale i ty přímo citované) obsahují velké množství vzorků přímo označených jako zranitelné (nebo aspoň zkoumané s potenciálem být), i když ne přímo v OpenJDK. Není nakonec jasné, podle jakého kritéria byly vzorky zkoumané přímo v této práci vybrané. Toto ale není výtka, vzhledem k rozsahu a časové náročnosti jsou výsledky i záběr přebíraných zdrojů dostatečné.

Seznam jednotlivých nedostatků:
- (překlep) s. IX: Virtula → Virtual
- (drobnost) s. 1, 9: nekonzistentní velikost písmen ve zkratkách DOS a ReDoS oproti seznamu zkratek
- (návaznost) s. 6: pojmy použité před jejich definicí: DFA, regex, PDA, LBA, gramatika, TM
- (překlep) s. 6: automat → automaton
- (citace) s. 35: bibliografická položka 7 pro OpenJDK (cit. na s. 7) je ošklivá (chybí název, působí neúplně)
- (citace) s. 35: bibliografická položka 9 pro API dokumentaci třídy Pattern (cit. na s. 7) je ošklivá (chybí název, působí neúplně)
- (překlep): s. 7 příklad 1.16: [xyz] → [xyz]+ [(asi)]
- (jazyk): s. 7 příklad 1.16: match → matches
- (citace) s. 35: bibliografická položka 10 (cit. na s. 7) je ošklivá (chybí název, působí neúplně)
- (faktická nepřesnost): s. 8: zpětná reference se v syntaxi Java regex zadává 1 zpětným lomítkem následovaným číslem, nikoli 2 zpětnými lomítky (autor patrně myslel zadání formou literálu řetězce v Javě)
- (mikrotypografie) s. 8: "\\1" → "\\1"
- (faktická nepřesnost): s. 8 příklad 1.24: uvedený regex popisuje jiný jazyk: {(ab)^n c^m d c^m : n > 1}, zatímco uvedený jazyk lze hledat např. regexem (abab)+(c*)d\2
- (nejasná formulace): s. 8, definice 1.27: u časové složitosti není jasné, zda se jedná konstrukce NFA nebo samotného vyhledávání a dále není jasné, které řetězce se týká parametr n, zda vzorku a nebo vstupního textu
- (faktická nepřesnost) s. 9: zpětné reference umožňují hledat kontextové, nikoli bezkontextové jazyky
- (citace) s. 35: bibliografická položka 13 (cit. na s. 9) je ošklivá (chybí název, působí neúplně)
- (citace) s. 9+35: citování Wikipedie pro definici klíčového pojmu této práce (ReDoS) působí nevhodně, přitom vhodným a přístupným zdrojem je např. stránka na OWASP, která je mimochodem v této práci rovněž citovaná
- (nadbytečná část) s. 11, podkapitola 2.1, 1. odst. v podstatě shrnuje úvod práce
- (citace) s. 11, položka 11: text se odkazuje na OpenJDK bug tracker, avšak položka 11 je článek publikovaný na arXiv preprints
- (zapomenutý text) s. 14, 2. odstavec: The difference
- (překlep) s. 19 a 27: (a+)1,100 → (a+){1,100}
- (mikrotypografie: spojovník místo pomlčky) s. 21, 24, 27, např. package - java.util.regex → package – java.util.regex
- (překlep) s. 22: matche → matcher
- (překlep) s. 22: Patcher → Matcher
- (překlep) s. 22: IntellijIDE → IntelliJ IDEA
- (citace) s. 36: bibliografická položka 20 (cit. na s. 22) je ošklivá (chybí název, působí

neúplně)
- (typografie) s. 27: mírný přesah textu do pravého okraje
- (mikrotypografie) s. 27: spojovník místo mínus: -1 → –1
- (překlep) s. 30: posted on there → posted on their
- (reference) s. 30: figure 1 → algorithm 1

### 3. Nepísemná část, přílohy 100 /100 (A)

Nepísemná část není součástí zadání. Příloha obsahuje zdrojové kódy použité v experimentech a výstupy experimentů. Prostředky použité pro experimenty hodnotím jako přiměřené.

### 4. Hodnocení výsledků, jejich využitelnost 95 /100 (A)

Práce z velké části shrnuje již známé i publikované poznatky, které formuluje jinak, avšak prezentuje i nové. Práce lze chápat jako výzkumnou. Věřím, že bude možné (po mírném dopracování a doladění nedostatků) výsledky publikovat formou (vědeckého) článku. Což je vzhledem ke skutečnosti, že jde o bakalářskou práci, skvělý výsledek.

### 5. Aktivita studenta

[1] výborná aktivita
[2] velmi dobrá aktivita
▸ **[3] průměrná aktivita**
[4] slabší, ale ještě dostatečná aktivita
[5] nedostatečná aktivita

Student se práci věnoval průběžně, s dodržováním termínů nebyl problém, stejně jako s připraveností na konzultace. Škoda, že psaní a začišťování textu nevěnoval více času, avšak chápu, že šlo o maximum možného.

### 6. Samostatnost studenta

[1] výborná samostatnost
▸ **[2] velmi dobrá samostatnost**
[3] průměrná samostatnost
[4] slabší, ale ještě dostatečná samostatnost
[5] nedostatečná samostatnost

Velká část práce je samostatnou iniciativou studenta (2 rozsáhlé kapitoly). Hledání zdrojů a hlavní výsledky práce jsou rovněž jeho zásluhou. Student od začátku zvolil systematický přístup, kterého se celou dobu držel. Jediná slabina byla v závěrečné části psaní textu, kdy samostatnosti bylo méně, avšak nic, co by vybočovalo z běžného stavu bakalářského stupně.

## Celkové hodnocení 89 /100 (B)

Práce dobře shrnuje známé výsledky a dokonce přináší nové. Celkový dojem z jinak dobré výzkumné práce sráží kvalita zpracování textu: větší množství drobných nedostatků. To je důvodem sníženého bodového hodnocení.

# Instrukce

## Splnění zadání

Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

## Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

## Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

## Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

## Aktivita studenta

V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

## Samostatnost studenta

V souvislosti s průběhem a výsledkem práce posuďte schopnost studenta samostatné tvůrčí práce.

## Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.