



Posudek oponenta závěrečné práce

Oponent práce:	Ing. Zdeněk Rybola, Ph.D.
Student:	Bc. Jakub Doležal
Název práce:	Anonymizace osobních údajů ve strukturovaných dokumentech
Obor / specializace:	Webové a softwarové inženýrství, zaměření Softwarové inženýrství
Vytvořeno dne:	29. května 2022

Hodnotící kritéria

1. Splnění zadání

[1] zadání splněno

► [2] zadání splněno s menšími výhradami

[3] zadání splněno s většími výhradami

[4] zadání nesplněno

Většinu požadavků zadání práce považuji za splněné, s výjimkou diskuze možných přístupů k definici anonymizačního modelu. V práci je představen jen jediný konečný návrh bez jakékoli diskuze možností. Výtku mám také k otestování výsledného řešení, které diskutuje pouze test výkonnosti a jednotkové testy aplikace jako takové zmiňuje jen v pár větách. Také množina podporovaných anonymizačních funkcí není nikde diskutována.

2. Písemná část práce

68/100 (D)

Písemná práce je rozdělena do dvou hlavních částí. První část je věnována analýze a návrhu. V rámci popisu aplikace Winch postrádám vysvětlení užší souvislosti se samotnou anonymizací sdílenou jak aplikací pro anonymizaci databází, tak aplikací pro anonymizaci tabulkových souborů.

Specifikaci požadavků považuji za dostatečnou, stejně jako rozbor formátů XML a JSON a způsobů definice jejich struktury.

Analýza podobných řešení se omezuje pouze na velmi stručnou analýzu způsobů vizualizace stromové struktury.

Následující kapitola vhodně rozebírá technologie pro zobrazení struktury XML a JSON souborů, dotazovací jazyky a existující parsery.

Kapitola 2.8 by měla diskutovat možnosti definice anonymizačního modelu pro provedení anonymizace souborů. Obsahuje však jen popis výsledného návrhu, žádná diskuze a nastínění různých možností se zde nenachází.

Následuje kapitola 2.9 popisující architekturu řešení. Text kapitoly je velmi hutný, popisuje jednotlivé balíčky v kódu, v nich obsažené třídy. Navíc se zde míchá popis stávajícího řešení a nově přidávaných tříd pro podporu nové funkčnosti. Díky tomu se v textu velmi špatně orientuje. Vhodnější by bylo zestručnit popis stávajícího kódu/architektury a popis novinek oddělit zvláště do návrhové kapitoly. Navíc se zde ztrácí rozdíly mezi kontextem, konfigurací, vzory, apod.

Kapitola 3 stručně shrnuje provedenou realizaci a odchylky od návrhu z předchozí kapitoly. Zakončena je výčtem neimplementovaných funkcí bez jakéhokoli zdůvodnění. Mezi ně patří mimo jiné konfigurace anonymizace na základě XSD nebo JSON Schema nebo podpora anonymizace více souborů naráz prostřednictvím cesty se zástupnými symboly požadované v požadavcích.

Kapitola 4 zmiňuje použití jednotkových testů k ověření implementace, ale bez jakýchkoli podrobností nebo kvantifikace. Autor se tak zde zaměřuje spíše na testy výkonu kvůli zvolenému způsobu načítání XML a JSON dokumentů.

Po formální stránce je práce na přiměřené úrovni. Jazykově je práce a obsahuje přiměřené množství gramatických chyb a překlepů. Nevhodné je použití zkratk (např. "anonym. třídu" na str. 28). Z textu jsou vhodně citovány zdroje, jejich seznam na konci práce však obsahuje nedostatky (např. chybějící datum čerpání z online zdrojů).

3. Nepísemná část, přílohy

75/100 (C)

Výsledkem práce je nová verze aplikace Winch pro anonymizaci souborů, která mimo původní podpory pro tabulkové soubory nyní nabízí také podporu pro anonymizaci strukturovaných souborů ve formátu XML a JSON. Oproti požadavků však aplikace neumožňuje výběr souborů pomocí zástupných symbolů (dle masky či proměnlivé cesty), podporuje pouze minimum anonymizačních tříd a neumožňuje konfiguraci anonymizace podle externě definovaného schématu (pomocí XSD, DTD či JSON Schema).

Přiložené zdrojové kódy představují rozšířené zdrojové kódu původní verze aplikace. Jejich architektura a členění tak odpovídá dříve zavedenému stylu.

Přestože je v textu práce zmiňována uživatelská příručka (půl stránky popisu konfigurace anonymizace v novém UI aplikace) a příručka je přiložena jako příloha práce, jedná se pouze o původní příručku s anonymizací tabulkových souborů (XLS a XLSX) a o anonymizaci XML nebo JSON tam není ani zmínka.

4. Hodnocení výsledků, jejich využitelnost

70/100 (C)

Výsledkem práce je nová verze aplikace Winch pro anonymizaci souborů podporující nově také anonymizaci strukturovaných souborů ve formátu XML nebo JSON. Pro reálné použití je však rozšíření nedotažené, protože nepodporuje několik klíčových funkcí. Z nich nejpodstatnější je podpora minimálního množství anonymizačních funkcí bez možnosti jejich další konfigurace. Dále pak chybí podpora konfigurace anonymizace na základě schématu obsahu souborů (XSD, apod.) nebo požadovaná možnost hromadné anonymizace více souborů pomocí cesty se zástupnými znaky.

Celkové hodnocení

70/100 (C)

Celkově hodnotím práci jako průměrnou. Některé části práce jsou na dobré úrovni, ale řada nedostatků celkovou kvalitu práce snižuje. Navrhuji tedy hodnotit stupněm C.

Otázky k obhajobě

V které části systému Winch jsou implementovány samotné anonymizační funkce? Proč jich aplikace ve stávající verzi podporuje tak málo? Co je třeba udělat, aby aplikace dokázala použít všechny anonymizační funkce dostupné třeba pro anonymizaci v databázích nebo tabulkových souborech?

Jaké jsou důvody pro neimplementování zmiňovaných požadavků? Jak náročné (kvantifikujte) by bylo je doimplementovat?

Instrukce

Splnění zadání

Posudte, 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, posudte 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 posudte, 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. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, 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.

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.