



Posudek oponenta závěrečné práce

Oponent práce:	Ing. Zdeněk Rybola, Ph.D.
Student:	Tomáš Průšek
Název práce:	Anonymizace osobních údajů pro databázi PostgreSQL
Obor / specializace:	Webové a softwarové inženýrství, zaměření Softwarové inženýrství
Vytvořeno dne:	11. června 2023

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání práce považuji za plně splněné.

2. Písemná část práce

80/100 (B)

Písemná část práce je na dobré úrovni.

Autor dokumentuje provedenou analýzu aplikace Winch a její podpory pro anonymizaci dat v databázích, především pak v databázi PostgreSQL. Dále diskutuje možnosti uložení a práce se strukturovanými dokumenty v XML a JSON formátu v této databázi. Tato část je informačně bohatá a přehledná, dobře srozumitelná.

V rámci další kapitoly popisuje autor návrh řešení pro dokončení chybějících částí anonymizace v PostgreSQL databázi a anonymizaci XML a JSON dokumentů v ní uložených. Následuje kapitola Realizace popisující konkrétní výsledky implementace. Obě tyto kapitoly obsahují řadu podrobností o technickém řešení, které jsou však pro čtenáře těžko srozumitelné. Jedná se hutný textový popis. V některých částech by stačil stručnější popis, jinde by pomohl diagram (např. sekvenční diagramy pro komunikaci mezi třídami). Text je také v některých místech nekonzistentní (označení tříd nebo metod běžným fontem, způsob uložení XML/JSON jako text nebo varchar, apod.).

Informace o testování implementovaného řešení je rozprostřena napříč návrhem a realizací, většinou velmi stručně. Větší prostor je věnován pouze testování výkonnosti řešení ve srovnání s alternativním řešením, které bylo zavrhnuto.

3. Nepísemná část, přílohy

80 /100 (B)

Součástí práce je sada zdrojových kódů, které autor v rámci řešení vytvořil. Příložený jsou však pouze zcela nové soubory, nikoli autorem upravené/rozšířené zdrojové kódy. Soubory také neodpovídají svou hierarchií skutečné struktuře řešení. O řadě úprav diskutovaných v rámci textu práce (i takových, které jsou přímou součástí zadání práce) se tak nelze přesvědčit. Aplikaci nelze samostatně spustit a implementované funkčnosti tak vyzkoušet.

Poskytnuté zdrojové kódy však mají vysokou úroveň, obsahují bohaté dokumentační komentáře, jsou vhodně strukturované.

4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Výsledné řešení splňuje zadání, tedy umožňuje plně anonymizovat data v PostgreSQL databázi ve stejném rozsahu, jako umožňuje aplikace Winch pro jiné podporované typy databází. Umožňuje navíc také anonymizovat strukturované XML a JSON dokumenty uložené v této databázi.

Samotná obsluha anonymizace je však velmi nepohodlná. Vyžaduje manuální získání vhodného vzorku dat z databáze, konfiguraci v jedné aplikaci rodiny Winch, následně přenos konfigurace do druhé aplikace, aby se následně v databázi provedla samotná anonymizace. Nicméně, cílem práce byla implementace samotné anonymizace, nikoli její konfigurace. Tento způsob tak lze považovat za dočasné implementačně zjednodušené řešení, které bude následně nahrazeno plnohodnotným řešením pro konfiguraci společnou pro další databáze a které nyní umožňuje autorem implementovanou anonymizaci provádět.

Celkové hodnocení

88 /100 (B)

Dle výše zmíněného navrhuji hodnotit práci výslednou známkou B, velmi dobře. Důvodem je především absence dalších dotčených zdrojových kódů v přílohách práce a místy příliš hutný text práce složitý na sledování a pochopení.

Otázky k obhajobě

Popište možný budoucí způsob konfigurace anonymizace XML a JSON dokumentů v PostgreSQL databázi jednotný se zbytkem konfigurace.

Popište možnosti detekce dat v XML a JSON souborech vhodných k anonymizaci.

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.