



Posudek oponenta závěrečné práce

Oponent práce: Ing. Jaroslav Kuchař, Ph.D.
Student: Bc. Daniil Fedotov
Název práce: Analýza logů serveru
Obor / specializace: Webové a softwarové inženýrství, zaměření Webové inženýrství
Vytvořeno dne: January 24, 2022

Hodnotící kritéria

1. Splnění zadání

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

Cílem práce je řešit problém analýz velkého množství logů z aplikačních serverů. Podívat se na možné metody (zejména z oblasti hledání shluků), způsoby a následně vybrané využít pro ukázkovou analýzu. Práce je zaměřena na znalost z oblasti znalostního i webového inženýrství a jejich propojení. Z tohoto pohledu považuji za zadání jako náročnější.

2. Písemná část práce

75/100 (C)

Práce je psána v anglickém jazyce - obsahuje překlepy, chybějící slova ve větách a v některých částech obraty, které zhoršují pochopení. Je ale obecně srozumitelná a čitelná.

ZP neobsahuje zbytečné části. Bylo by ale vhodné vyvážit informační bohatost u vybraných částí, např. detailní popis GSOM vs metoda postavená na cosine podobnosti - ta má minimalistický popis a není úplně jasný její princip.

U Kapitoly 2 postrádám výsledné zhodnocení, jak uvedené nástroje souvisí nebo je možné využít v rámci řešení ZP.

ZP neobsahuje závažné nedostatky po obsahové stránce nebo z pohledu nepřesností - pouze na straně 19, zdůvodnění nevhodnosti "One-hot encoding" pro jeho vysokou dimenzi a řídkost je podobně platné i pro ostatní zmíněné VSM.

Práce obsahuje mnoho dílčích rozhodnutí, které by bylo vhodné řádně ověřit a podpořit experimenty (jako je zvolení bigramů, konkrétních konstant a metod). Ale chápu, že vzhledem k obsáhlosti problému nebylo možné vše zvládnout.

Bylo by vhodné objasnit proč se např. zvolila metoda GSOM. Zda to bylo na základě

zkušeností apod. Proč se nezkusil např. přímo K-Means a porovnat výsledky. ZP ale popisuje do detailu konkrétní navržené řešení a ukazuje důležité směry práce do budoucna z pohledů výzkumu a vývoje v této oblasti. Z hlediska citací a licenčních podmínek neshledávám závažné nedostatky.

3. Nepísemná část, přílohy 85 /100 (B)

Výstupem práce je implementace zvolených metod a provedení experimentů. Student připravil specifické vlastní implementace, které jsou využitelné i pro budoucí práci. Použité nástroje odpovídají současným trendům a řešením. Z pohledu SI by bylo vhodné doplnit testování na úrovni SW.

4. Hodnocení výsledků, jejich využitelnost 93 /100 (A)

Obsahem této práce je relevantní téma, které se řeší ve výzkumu i v praxi. Výstupy jsou použitelné jako základ, na kterém lze dále stavět. Jedná se i o základ pro publikační výstupy.

Celkové hodnocení 87 /100 (B)

ZP podává teoretické i praktické spojení webového a znalostního inženýrství. Student se podrobně seznámil s vybranými metodami, v praktické části připravil vlastní implementace zvolených algoritmů. Práce by zasloužila rozšířit z pohledu vyhodnocení kvality dílčích přístupů a odůvodnění vybraných postupů. Je to ale dáno komplexností tématu. Zároveň je to cílem obecně výzkumu a vývoje v této oblasti.

Otázky k obhajobě

- Jak je přístup přenositelný na jiné aplikace mimo WebLogic?
- Použila se mimo vektorizací pomocí VSM i vektorizace s využitím embedding (Doc2Vec apod.)?
- Byly provedeny experimenty s existujícími metodami pro hledání clusterů? Jako je např. K-Means (využívá se i v GSOM)?

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.