



Posudek oponenta závěrečné práce

Student: Andrej Taňkoš
Oponent práce: Mgr. Jiří Toušek
Název práce: Návrh a implementace analýzy datových toků pro jazyk Pig v projektu Manta
Obor: Webové a softwarové inženýrství

Datum vytvoření: 5. 6. 2018

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> 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.	
<i>Komentář:</i> Práce splnila zadání.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	80 (B)
<i>Popis kritéria:</i> 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. 26/2017, 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.	

Komentář:

Rozsah práce je přiměřený, práce je dobře strukturovaná a srozumitelná, neobsahuje zbytečné části.

Text práce obsahuje značné množství překlepů a drobných chyb, namátkou: "automaticky" místo "analýzou" kapitole 2.4; malé písmeno na začátku věty v kapitole 4.2.2; "vytváraní uzlov pre revolvovanú entitu" v kapitole 4.3.3; odkaz na "Následující obrázok 2.1", který se však v tetu kapitoly 2.3.1 nachází před tímto odkazem.

V kapitole 2.3 je použita zkratka UDF, která není v seznamu zkratk a je definovaná až v kapitole 2.3.4.

V kapitole 2.4 je uvedeno, že uzel QUERY má potomky typu QUERY (správně STATEMENT).

V kapitole 3.1.1 je chybný odkaz na diagram 3.2 ("Ako je možné vidiet z diagramu 3.2, ..."), má být 3.3. Následně je na diagram 3.3 odkázáno při popisu stromové struktury reprezentující identifikátor, která je ovšem znázorněna v diagramu 3.4.

V téže kapitole je uvedeno jméno třídy AstCopeNode místo AstScopeNode.

V kapitole 4.3.3 je uveden odkaz na požadavek NF2 místo NF1, navíc je tento požadavek označen za funkční.

V příloze B se AST zapisuje do proměnné "result", avšak následně se čte z proměnné "tree".

V kapitole 2.4.1 autor zvolil použití parseru z projektu Apache Pig ve formě Java kódu vygenerovaného lexeru a parseru. Jako výhodu uvádí možnost použít aktualizovanou verzi parseru/lexeru. Nikde však neuvádí, jakým způsobem by se taková aktualizace provedla - přiložený Java kód parseru a lexeru zcela jistě není pouze vygenerovaný z gramatiky Apache Pig; autor neuvádí, jaké úpravy je nutné v gramatice nebo ve vygenerovaném Java kódu provést.

V kapitole 3.1.1 mám za to, že obecný text popisující transformaci stromu se zčásti slil s popisem reprezentace identifikátorů. Dále bych v této kapitole ocenil příklad takové nevhodné struktury původního AST, zdůvodnění její nevhodnosti a ukázkou její transformace na vhodnější tvar.

Typografická úroveň práce je vyhovující. Práce se zdroji je v pořádku.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

3. Nepísemná část, přílohy

70 (C)

Popis kritéria:

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

Komentář:

Prototypová implementace řeší téměř kompletně problém analýzy datových toků v Apache Pig, funkčnost prototypu je ověřena testy.

Největší výhradu mám ke způsobu integrace prototypu s projektem Manta, kdy autor místo maximálního využití stávající infrastruktury zvolil cestu velmi volné integrace, kdy často pouze implementuje požadovaná rozhraní a vytváří jejich nové, odlišné implementace. Tím je zásadně ztížena možnost reálně integrovat modul do projektu: jednak by pak bylo nutné v projektu udržovat více implementací téže funkcionality, druhak autor ve svých implementacích mnohé aspekty použitých rozhraní vůbec neimplementoval. Spokojil se totiž s implementací jen těch částí rozhraní, které pro svůj modul nutně potřeboval, zbytek rozsáhlé infrastruktury ponechal bez implementace. Toto navíc učinil velmi nevhodným způsobem, kdy neimplementované metody místo vyhození výjimky buď vrátí null, nebo nedělají nic. Takovou implementaci však nelze použít mimo hranice modulu - jakákoliv metoda, kterou autor nepotřeboval přímo pro funkci modulu, totiž nejen nebude správně fungovat, ale volající se to ani nemusí dozvědět.

Velké množství tříd a rozhraní (možná většina) postrádá dokumentaci ve formě Javadoc, ač autor v závěru své práce tvrdí opak.

Kód místy nevhodně pracuje s výjimkami (odchytávání obecné Exception, odchytávání IndexOutOfBoundsException).

V kódu se vyskytuje třída PigParserNode, jejíž význam mi není vůbec jasný. Javadoc nesděljuje žádnou užitečnou informaci, text práce tuto třídu nezmiňuje, v kódu modulu jsem nenašel žádné použití, které by mi význam třídy a jejích položek ozřejmilo.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

80 (B)

Popis kritéria:

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.

Komentář:

Implementované prototypové řešení prokazuje teoretickou správnost i praktickou použitelnost navržených postupů. Výsledky práce jsou přímo uplatnitelné v praxi.

Implementované prototypové řešení bohužel vzhledem k popsáním problémům vyplývajícím ze zvolené formy integrace nebude možné přímo využít, bude nutné jej značně upravit nebo pouze využít některé jeho části či teoretické závěry.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

- Zvažoval jste možnost použít parser a lexer z projektu Apache Pig v podobě zkompileovaných knihoven místo převzetí zdrojového kódu? Pokud ano, co vás vedlo k volbě použít raději zdrojový kód?
- V kapitole 4.1.4 uvádíte pravidla, podle nichž nástroj Pig pojmenovává objekty bez explicitního jména. Lze na takto automaticky pojmenované objekty následně odkazovat jejich generovaným jménem?
- Prototyp implementuje resolving identifikátorů, který by v případě více možných kandidátů na odkazovaný objekt mezi nimi vybral nedeterministicky. Může v Apache Pig nastat situace, kdy jednomu jménu odpovídá více možných objektů a je mezi nimi potřeba rozhodnout podle nějakého dalšího kritéria?

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

80 (B)

Popis kritéria:

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.

Text hodnocení:

Autor se úspěšně vypořádal s celou problematikou analýzy datových toků v nástroji Apache Pig. Ke zvládnutí tohoto úkolu musel prostudovat možnosti analyzovaného jazyka Pig Latin, místy experimentálně ověřovat chování nástroje a musel alespoň částečně prostudovat i jeho zdrojové kódy.

Oceňuji důkladně provedenou analýzu analyzovaného jazyka, pokrývající všechny jeho významné aspekty. Kvalitu písemné části práce bohužel snižuje množství zbytečných překlepů a chyb.

Prototypová implementace řeší téměř kompletně úkol analýzy datových toků z Apache Pig, korektnost implementace je ověřena testy. Požadavek na integraci modulu s projektem Manta je bohužel splněn jen formálně, kvůli zvolenému přístupu k integraci nebude přímé použití možné bez rozsáhlých úprav.

I přesto jsou však teoretické závěry práce a části prototypové implementace přímo uplatnitelné v praxi.

Podpis oponenta práce: