

Hodnocení vedoucího bakalářské práce

Martin Soukup

Experimentální server na podporu výzkumu v oblasti EOBÍ dat

Bakalářská práce pana Martina Soukupa popisuje analýzu problematiky a následný vývoj experimentálního serveru, který zpracovává rozsáhlá data poskytnutá společností Deutsche Börse AG ve formátu T7 EOBÍ a na základě jejich zpracování rekonstruuje elektronickou knihu objednávek burzy a v (téměř) reálném čase poskytuje klientům v odpovědi na jejich požadavky informace o stavu burzy ke zvolenému časovému okamžiku.

Autor se musel nejprve poměrně důkladně seznámit s mechanismy obchodování na moderní burze a prostudovat řadu značně komplikovaných dokumentů v anglickém jazyce (dokumentace formátu T7 EOBÍ, T7 EMDI, publikace popisující průběh obchodování na elektronicky řízené burze atp.). Následně provedl řadu testů různých databázových řešení s cílem zjistit, jak bude nejlepší (a nejrychlejší) ukládat a následně vybavovat data o obchodování, potřebná k rekonstrukci elektronické knihy objednávek. Zde je třeba zmínit, že se jedná o data velice rozsáhlá – řádově stovky megabytů pro jeden obchodovaný instrument za jeden den obchodování – a současně vzorkovaná s velmi vysokou frekvencí, kdy rozlišení časových značek v datech je na úrovni nanosekund.

Se všemi těmito náročnými výzvami se autor vyrovnal velmi dobře: Vyvinutý server pracuje víceméně podle požadavků a dosahuje časů vybavování požadovaných informací v řádu nižších stovek milisekund, tj. téměř v reálném čase. Jisté výhody (popsány níže) lze ovšem mít k samotné implementaci softwarového řešení. Text práce dobře dokumentuje provedené práce a pokrývá jak potřebné prerekvizitní teoretické poznatky, tak následně postup implementace softwarového řešení.

Textová část předložené práce je na velmi vysoké úrovni. Autor – přesto, že je jeho vyjadřování velmi úsporné – pečlivě a srozumitelně popsal výchozí teoretické poznatky (formáty dat a API), vysvětlil fungování burzy, smysl a účel elektronické knihy objednávek a postupy její rekonstrukce. Dále zdokumentoval provedené experimenty s různými databázovými stroji (vč. moderních TSDB a NoSQL databází), kdy testoval výkon daného řešení při specifických požadavcích na vybavování uložených dat pro potřeby rekonstrukce knihy objednávek. Následně důkladně rozebral možné postupy této rekonstrukce a navrhl algoritmické řešení vyhovující účelu této práce. Popisu samotné implementace, tedy přímo zdrojového kódu v jazyce Python a jeho rozvržení do modulů, dekompozici algoritmu atd. se autor věnoval také, ale dle mého subjektivního pohledu **mohla být tato část práce poněkud detailnější a rozsáhlejší**.

Poněkud problematičtější je část implementační. Kvalitu zdrojového kódu v jazyce Python lze hodnotit jen velmi obtížně, neboť se jedná vlastně jen o pouhé 4 skripty o celkové délce 651 řádků kódu – o nějakých např. pokusech o optimalizaci časové nebo paměťové náročnosti není vlastně ani jak debatovat (tím spíše, že autor kód v Pythonu používá pouze jako skript volající knihovní funkce). Zdrojový text je povětšinou zapsán ve shodě se zvyklostmi, je přijatelně okomentovaný a celkem dobře čitelný. Práce zjevně není typickou programátorskou kvalifikační prací, její hodnota tkví zejména v průzkumu vhodných nástrojů a postupů, provedených experimentech, zátěžových testech a zejména návrhu efektivního algoritmu rekonstrukce knihy objednávek. Nicméně v souvislosti se zdrojovým kódem je na tomto místě třeba autorovi vytknout užití programátorské techniky naprosto nepřijatelné i v případě tohoto experimentálního prototypu: Celý software je prakticky nekonfigurovatelný, resp. změn konfigurace se dosahuje výhradně zásahy do zdrojového kódu. Takto jsou „natvrdo“ nakódována i např. jména datových souborů, které jsou vstupem pro rekonstrukci knihy objednávek. Také parametry vizualizace (barvy, fonty atp.) nelze měnit jinak, než ručním přeprogramováním příslušného úseku zdrojového kódu.

Po formální stránce je práce na velmi dobré úrovni: Text je napsaný vynikající odbornou češtinou s naprostým minimem chyb, čte se velmi dobře a zejména rychle, protože autorův styl vyjadřování je oproti většině ostatních bakalářských prací s podobným zaměřením mimořádně úsporný, až strohý. Ale to není vlastně na závadu: Autor neunavuje, nenudí, je téměř uspokojivě popisný a rozhodně není epický (což je u některých podobných textů značně na obtíž). Jak již bylo naznačeno výše, pouze kapitola 5. Implementace je dle mého názoru už stručná příliš a chybí mi v ní detailnější popis navrženého algoritmu rekonstrukce knihy objednávek, organizace zdrojového kódu, dekompozice algoritmického řešení, popis alespoň elementárního modelování chování aplikace, identifikace funkčních požadavků a jejich následné uspokojení zapsaným zdrojovým kódem atp.

Estetika dokumentu je naopak bezvadná: Sazba je provedena v L^AT_EXu a výsledek působí velice harmonickým

dojmem. Text je vhodně doplněn diagramy, tabulkami a výpisy zdrojového kódu. Obrázky napomáhají pochopení výkladu a jsou v náležitě technické kvalitě.

Autor využívá v práci 23 zdrojů, jejichž výběr dobře odpovídá tématickému zaměření práce. Jde v drtivé většině o dokumenty dostupné online a dokumentační weby různých softwarových produktů, knihoven atp. Uvážíme-li ovšem specifické zaměření práce, není v podstatě jiná možnost a výběr zdrojů tak lze považovat za vhodný. Požadovat recenzované publikace nebo dokonce tištěné knihy je prakticky vyloučeno, když jen specifikace formátu dat T7 EOBI se za rok změni 2× až 3×.

Zdroje autor cituje korektně a jejich seznam v závěru díla je v souladu s normou ISO 690 a zaužívanými zvyklostmi, až na situace, kdy je jako autor v seznamu literatury uvedena organizace, která předmětný dokument zveřejnila. Následkem téhož jsou pak názvy organizací mnohdy nesprávně kapitalizovány, protože v implicitním nastavení stylu sazby seznamu použité literatury jsou příjmení autorů sázena verzálkami. Nicméně jedná se o pouze technickou chybu, žádná práva autorů použitých zdrojů nebyla pošlapána.

Všechny body zadání práce byly splněny. Práce je nesporně velice užitečným dílem, které zcela určitě bude dále využíváno k provádění vědecko-výzkumných experimentů v oblasti zpracování finančních časových řad, což není na bakalářské úrovni vždy pravidlem. Autor bezpochyby prokázal schopnosti zpracovat poměrně náročné a často mizerně dokumentované teoretické partie, navíc i v oblastech, které nebyly pokryté jeho studijním plánem, tak také na základě načerpaných znalostí vytvořit funkční softwarové řešení, byť s výše popsány nedostatky.

Práci proto každopádně **doporučuji k obhajobě** a s ohledem na uvedené výtky hodnotím klasifikačním stupněm

„velmi dobře“.

Ing. Kamil Ekštejn, Ph.D.
KIV FAV ZČU

V Plzni dne 19. května 2024

K práci nemám žádné doplňující otázky.