

Posudek oponenta diplomové práce

Student: Matoušek Martin, Bc.

Téma: Paralelní a distribuované zpracování rozsáhlých textových dat (id 17871)

Oponent: Dytrych Jaroslav, Ing., UPGM FIT VUT

1. **Náročnost zadání** **průměrně obtížné zadání**
Zadání bylo možné pojmout i jako obtížné, ale student dle mého názoru zůstal u průměrného pojetí.
2. **Splnění požadavků zadání** **zadání splněno**
Zadání bylo splněno, i když vytvořený plakát, který považuji spíše za formalitu, nemá vyhovující úroveň.
3. **Rozsah technické zprávy** **splňuje pouze minimální požadavky**
Technická zpráva má cca 42 normostran textu a necelých 7 tištěných stran obrázků. Všechny části technické zprávy jsou informačně bohaté a většina je jich pro práci nezbytná.
4. **Prezentační úroveň předložené práce** **60 b. (D)**
Technická zpráva sice jako celek má logickou strukturu a jednotlivé kapitoly na sebe navazují, ale členění na podkapitoly dle mého názoru není vhodné. Některé podkapitoly 2. úrovně neobsahují žádný text, ale pouze číslované podkapitoly 3. úrovně, které jsou často hodně krátké. Místy se pracuje s pojmem, který je vysvětlen později. Význam některých částí teorie pro řešenou problematiku není zcela jasný - např. podkapitola 2.2.4 či popisy plánovacích algoritmů, se kterými se později nepracuje. U používaných systému pro plánování bych pak očekával srovnání. Návrh je dle mého názoru příliš stručný. Popis implementace je naopak na příliš nízké úrovni. U využitých technologií by bylo vhodné uvést jejich verze. Popis jednotlivých metod tříd patří spíše do programové dokumentace než do technické zprávy.
5. **Formální úprava technické zprávy** **69 b. (D)**
V práci je celá řada formálních chyb. Překlepy a drobné chyby se vyskytují v celé práci už od anglické verze abstraktu. Často se v práci vyskytuje anglický slovosled (např.: "ISO8601 notace" či "CPU jader"), nepřesnosti v terminologii a nešikovné obraty (např.: "rychlostí doby zpracování"). Některé obrázky a tabulky nejsou odkazovány z textu. Po typografické stránce je práce lehce nadprůměrná, ale místy se vyskytují chyby jako využívání spojovníků místo pomlček.
6. **Práce s literaturou** **98 b. (A)**
Literatura obsahuje tištěné knihy, vědecké články i elektronické zdroje. Výběr literatury je adekvátní pro dané téma. Práci s literaturou považuji za nadprůměrnou.
7. **Realizační výstup** **82 b. (B)**
Vytvořené řešení je plně funkční na serverech skupiny KnoT i na superpočítači Salomon. V obou případech dochází k výraznému zrychlení zpracování oproti předchozímu řešení. Za nevýhodu považuji menší uživatelskou přívětivost a absenci globálních statistik zpracování na serverech KnoT. Stránka s celkovými statistikami místo celkových statistik obsahuje počet zpracovaných souborů a doporučené nastavení pro další běh zpracování dat.
8. **Využitelnost výsledků**
Aplikace je prakticky využitelná pro zpracování velkých dat na superpočítači Salomon. Pro využití na serverech výzkumné skupiny KnoT by dle mého názoru byl potřebný další vývoj za cílem snížení pracnosti konfigurace a spouštění.
9. **Otázky k obhajobě**
 - Prováděl jste i testy s více servery výzkumné skupiny KnoT, nebo pouze s 1?
 - Zabýval jste se režii na měření spotřeby paměti s využitím vzorkování?
 - Jaký zásah do aplikace by byl nutný pro měření zátěže procesem SEC a proč?
10. **Souhrnné hodnocení** **78 b. dobře (C)**
Vytvořené řešení je plně funkční na dvou diametrálně rozdílných platformách a má jasné přínosy pro rychlost zpracování a efektivitu využití zdrojů. Text technické zprávy je po prezentační a formální stránce podprůměrný a návrhová a experimentální část práce měly být popsány lépe. Celkové řešení hodnotím jako průměrné a navrhuji hodnocení stupněm C.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

.....

podpis