

# Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

**Student:** Bc. Pavel Řehák  
**Oponent práce:** Ing. Daniel Langr, Ph.D.  
**Název práce:** Paralelní řadící algoritmy  
**Obor:** Počítačové systémy a sítě

**Datum vytvoření:** 26. 1. 2018

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 5:</b>
<b>1. Náročnost a další komentář k zadání</b>	<b>1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání</b>
<b>Popis kritéria:</b> Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
<b>Komentář:</b> Práce se zabývá netriviální problematikou řazení dat ve vícevláknovém prostředí a v prostředí grafických karet. Sekvenční řazení je již poměrně dobře prozkoumané, ale dnes všudypřítomné vícejádrové architektury přinesly do řazení nové výzvy. Totéž v ještě větší míře platí o grafických kartách/akceleračních jednotkách, které jsou dnes čím dál tím více využívány i v oblasti obecných výpočtů. V práci jsou předmětem zkoumání především tři základní řadící algoritmy, konkrétně obě QuickSort a MergeSort, a dále specializovaný RadixSort.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>2. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
<b>Komentář:</b> Zadání práce bylo splněno, práce navíc obsahuje i GPU implementace, které v zadání nebyly požadovány (byl požadován pouze návrh úpravy pro efektivní běh na GPU).	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>3. Rozsah písemné zprávy</b>	<b>1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky</b>
<b>Popis kritéria:</b> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
<b>Komentář:</b> Rozsahem práce odpovídá prezentované problematice a neobsahuje žádné zbytečné části.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>4. Věcná a logická úroveň práce</b>	<b>85 (B)</b>
<b>Popis kritéria:</b> 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.	

#### Komentář:

Po věcné stránce je práce v pořádku, text na sebe navazuje a kapitoly 1-4 obsahují přehledný a informačně poměrně bohatý úvod do problematiky řadících algoritmů a jejich paralelizace pomocí OpenMP a CUDA paradigmat. Moje připomínky:

- V Úvodu autor prakticky jen shrnuje obsah práce a píše, co které kapitoly obsahují. Od Úvodu bych ale očekával jinou funkci - čtenář by měl být uveden do celé řešené problematiky. Konkrétně by Úvod měl obsahovat autorovu motivaci pro řešení prezentovaných problémů, rešerši současného stavu v dané problematice, návaznost a souvislosti se stávajícími řešeními, popis, v čem je autorův přístup jiný, jaké jsou jeho přínosy apod. Úvod by měl maximálně namotivovat čtenáře pro zbytek textu, což tady vůbec neplatí.
- 1.2: "Každý řadící algoritmus má své vlastnosti stejně tak jako i jiný algoritmus." - Zcela nesmyslná a nic neřkající věta.
- Do sekce 1.4 bych doporučoval doplnit Insertionsort, který se z  $O(n^2)$  algoritmů používá v praxi asi nejvíc. Často se kombinuje (na rozdíl od Bubblesortu či Selectsortu) s Quicksortem a Mergesortem, kde se jím řadí malé posloupnosti na konci rekurze. Např. v QS se navíc dá zavolat jen jednou nakonec, protože je efektivní pro částečně seřazené posloupnosti.
- V Kapitole 3 autor čerpá hlavně z přednášek předmětu MI-PDP. Bylo by vhodnější prozkoumat i další zdroje věnující se dané problematice.
- 5.1.1 Absence výběru pivota z mediánu tří prvků je trochu na škodu, protože tato strategie může mít výrazný vliv na dobu řazení (obecně pozitivní i negativní). V případě práce s náhodnými daty lze zajistit deterministický přístup i v tomto případě tím, že se pseudonáhodný generátor inicializuje stejnou "seed" hodnotou, takže pak generuje stejnou posloupnost pseudonáhodných čísel.

Hodnotící kritérium:

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

### 5. Formální úroveň práce

80 (B)

#### Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

#### Komentář:

K formální stránce práce mám následující připomínky:

- V tabulkách je používána desetinná tečka místo desetinné čárky (jde o český text, nikoliv anglický).
- Čísla v tabulkách by pro přehlednost bylo vhodnější zarovnat na desetinnou čárku (viz např. Tabulka A2) a rovněž nemá žádný smysl uvádět čísla s tak velkou přesností, pro porovnání by úplně stačila např. 2 desetinná místa.
- Abstrakt: "...Na konci je provedeno měření..." - Měření není "provedeno" ve textu práce, měření provedl autor a v textu prezentuje jeho výsledky.
- Úvod: "...Následuje paralelizace... Výsledné zrychlení bude porovnáno..." - nemixovat různé časy, působí to nepřirozeně a rušivě.
- Úvod: "První kapitola pojednává... Pomocí pseudokódu dopomohu..." - mixuje se činný a trpný rod, rovněž působí rušivě.
- Autor mezi odstavce z nějakého důvodu vkládá prázdné řádky, což není zvykem, působí to rušivě a zhoršuje to čitelnost.
- Strana 3: "snažší" - překlep, který musí odhalit každý spellchecker.

Hodnotící kritérium:

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

### 6. Práce se zdroji

95 (A)

#### Popis kritéria:

Vyjádrte se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

#### Komentář:

V práci se zdroji jsem nenašel žádné vážné problémy, jen jednu drobnost:

- Citace [14] odkazuje na slajdy z přednášek, nikoliv na skripta. Tyto slajdy nejsou veřejně dostupné, což by v textu v seznamu literatury mělo být asi uvedeno.

Hodnotící kritérium:

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

### 7. Hodnocení výsledků, publikační výstupy a ocenění

85 (B)

#### Popis kritéria:

Vyjádrte se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvoril sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

#### Komentář:

Naměřené výsledky jsou určitě zajímavé, jen bych uvítal měření s různým typem řazených prvků. Např. řazení prvků, u kterých přesun trvá výrazně déle než jejich porovnání či naopak, či prvků, u kterých není čas porovnání konstantní (jako jsou například řetězce). Jinak se mi líbilo provedení experimentů i na GPU nad rámec zadání.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

### 8. Komentář o využitelnosti výsledků

#### Popis kritéria:

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

#### Komentář:

Z hlediska vícevláknových implementací by určitě pro měření byla vhodnější architektura s více jádry a větším množstvím paměti (např. využití některého uzlu fakultního clusteru Star). Výsledky jsou obecně využitelné, ale pro jejich větší relevanci by bylo potřeba více experimentů, např. s již zmíněnými různými typy dat či na více architekturách.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

## 9. Otázky k obhajobě

Popis kritéria:

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

Otázky:

1) Autor kombnuje Mergesort se Selectsortem (viz 5.1.3). Pokud vím, v praxi se spíše pro tuto kombinaci používá Insertionsort. Jaké důvody vedly k preferenci Selectsortu? Má nějaké výhody?

2) Totéž platí o Quicksortu, ve kterém se často Insertionsort používá, pokud velikost řazené části pole klesne pod určitou úroveň (rekurze je pak "dražší" než  $O(n^2)$  algoritmus). Použil autor takovouto kombinaci, nebo se QS provádí až do konce rekurzivního stromu?

3) Proč autor ukládal vstupní data do souborů? Použitím pseudonáhodného generátoru se stejnou inicializační hodnotou lze snadno generovat stejnou posloupnost pseudonáhodných čísel (a to i vícevláknově) a tudíž např. provádět různá měření nad stejnými "náhodnými" daty. Generování by bylo pravděpodobně i rychlejší oproti načítání ze souboru.

Hodnotící kritérium:

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

## 10. Celkové hodnocení

80 (B)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Práce splňuje zadání a je poměrně kvalitně zpracovaná i dobře čitelná, až na jisté formální nedostatky. Patrně nejvíce mi v práci chyběla určitá motivace pro řešení prezentovaných problémů a souvislosti s aktuálním děním v daném oboru.

Podpis oponenta práce: