

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Autor práce: Bc. Filip Mihalovič
Název práce: Distribuovaná analýza genomických dat založená na cloud computingu
Vedoucí: Doc. Ing. Jiří Kléma, Ph.D.

Diplomová práce pana Mihaloviče se zabývá možnostmi paralelního sestavování genomických sekvencí z velkého množství krátkých čtení generovaných technikami sekvenování nové generace. Tento nový způsob sekvenování díky masové paralelizaci výrazně urychluje a zlevňuje proces sekvenování, současně ale generuje kratší čtení. To výpočetně komplikuje proces sestavování sekvencí, který následuje po samotném sekvenování. Sestavování lze zjednodušeně chápat jako proces nalezení nejkratší nadsekvence obsahující všechna čtení, o rozhodovací verzi tohoto problému je známo, že je NP-úplná. Základní myšlenkou práce je tedy snaha o maximální paralelizaci a zrychlení procesu sestavování s využitím metod cloud computingu.

Text práce začíná biologickým úvodem, dále obsahuje základní přehled použitelných distribuovaných systémů. Poté autor rekapituluje stávající algoritmy sestavování sekvencí založené na grafech překrytí a de Bruijnových grafech. Ve dvou klíčových kapitolách pan Mihalovič nejprve navrhuje vlastní paralelní algoritmus vycházející ze sekvenčního SAGE algoritmu a následně jej implementuje a testuje v prostředí Apache Spark. V důsledku se podařilo vytvořit funkční prototyp paralelního algoritmu sestavování sekvencí. I když prototyp, pokud jde o výpočetní časy, nepřekonává svůj sekvenční vzor, jsem s výsledkem spokojen. Prototypem se podařilo demonstrovat zásadní možnosti a omezení paralelizace sestavování sekvencí. Cílem práce nemohla být důsledná implementace všech optimalizačních heuristik, které řeší zejména mnoho dílčích problémů vzniklých nedostatků samotného sekvenování a redundancí genomu.

Diplomant řešil své téma dlouhodobě a zodpovědně. Aktivně se podílel na formulaci dílčích úkolů, v oblasti distribuovaných systémů prokázal velkou samostatnost. Pan Mihalovič reagoval po celou dobu naší spolupráce na kritické připomínky, i v poslední fázi řešení se mu podařilo zapracovat všechny výhrady, které jsem k textu práce měl. Oceňuji také obrovský skok, který student udělal od bakalářské práce, již jsem také vedl. Souhrnně lze říci, že práce splňuje všechny cíle, které jsme si na počátku vytkli.

Pan Mihalovič ve své diplomové práci naplnil všechny body zadání a prokázal schopnost samostatné práce. Práci doporučuji k obhajobě a hodnotím ji známkou

A — výborně.

V Praze, dne 25. května 2016

Doc. Ing. Jiří Kléma, Ph.D., vedoucí