

IDS
Projekt
-
Agregace a třídění flow dat

Zvolené algoritmy

Pro realizaci projektu byl zvolen programovací jazyk C++ ve standartu C++0x. Tento jazyk byl zvolen především pro jeho schopnost manipulovat s daty na nízké úrovni, tedy i podstatně rychleji nežli jazyky vysokoúrovňové.

Pro agregaci byl zvolen kontejner `std::unordered_map` a pro steřízení agregovaných dat bylo použito kontejneru `std::multimap`, jelikož třídící funkce kontejneru `multimap` byla rychlejší, než mnou napsaný algoritmus quicksort. Podle všeho by však měl být kontejner náročnější na paměť, což jsem považoval za přijatelné.

Rozbor algoritmu

Vlastní algoritmus je pouze sekvenční a je rozdělen do několika fází:

1. Rekurzivně se projde složka předaná argumentem a zpracují se data, která se zároveň i agregují a vloží do kontejneru `std::unordered_map`.
2. Následně se vytvoří kontejner `std::multimap` a obsah `std::unordered_map` se do něj vloží a automaticky setřídí na základě argumentu. V jednu chvíli jsou tedy v paměti data 2x -> paměťová náročnost. Po přenesení dat do kontejneru `std::multimap` jsou však data ze `std::unordered_map` smazána.
3. V třetí fázi dochází již k výpisu agregovaných dat na standartní výstup.

Možné optimalizace

Algoritmus je sekvenční, tedy platí možnosti optimalizace, které byly zmíněny v zadání. Konkrétně jde o paralelizaci čtení binárních souborů a použití paralelních algoritmů pro agregaci a třídění dat. Zároveň by bylo možné optimalizovat prostorovou náročnost programu pomocí postupného promazávání kontejneru `std::unordered_map`.