

Program na vyhľadávanie reťazca

Šimon Harvan

Špecifikácia

Vytvorte program ktorý bude robiť vyhľadávanie reťazca v textovom súbore sekvenčne aj paralelne (napr. s využitím OpenCV) a porovnajte výsledky (z hľadiska výkonnosti).

Cieľ

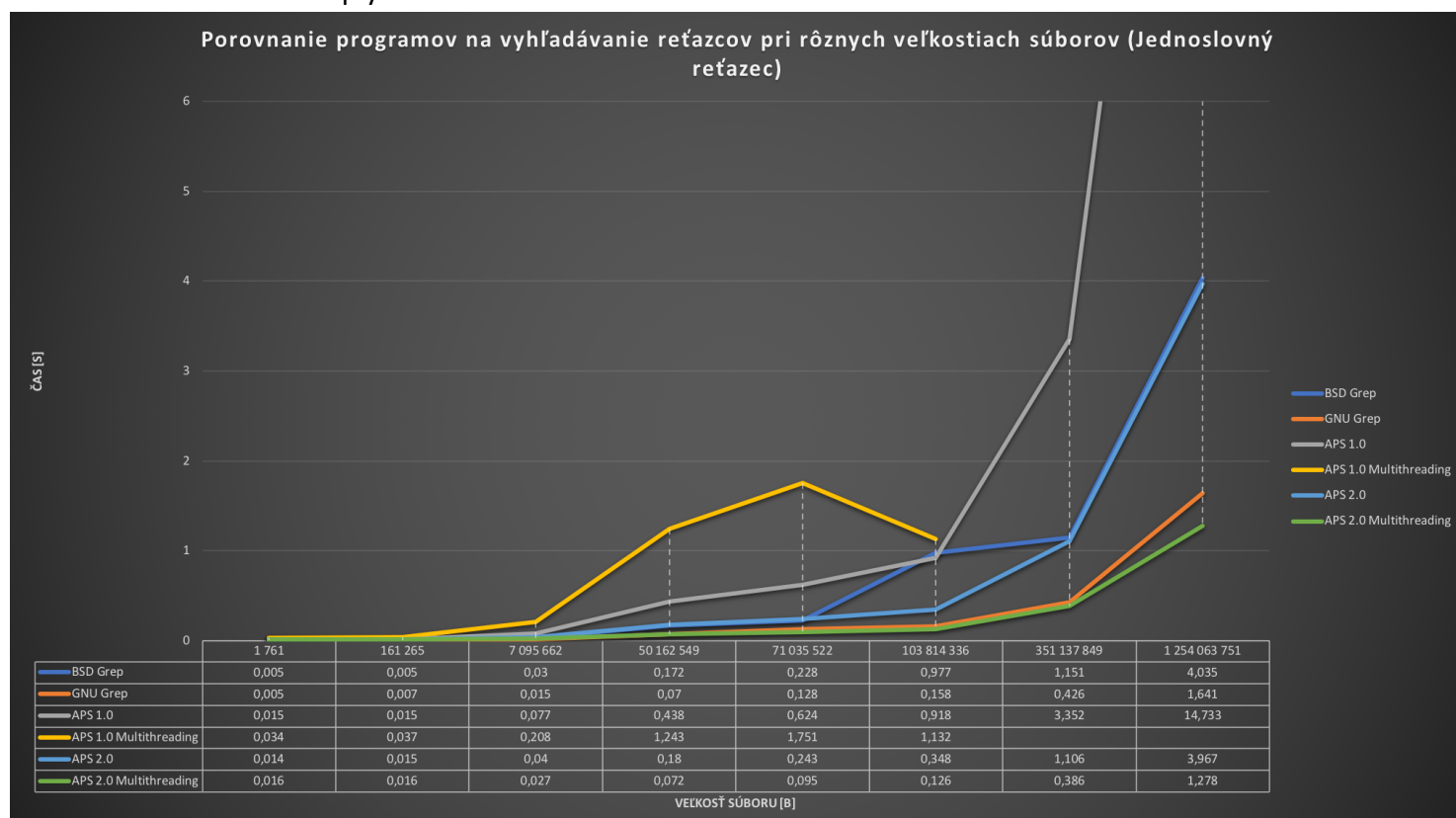
Vytvoriť konzolovú aplikáciu, ktorá bude vedieť na viacerých vláknach vyhľadať reťazec vo veľkých súboroch (rádovo stovky MB), ktorá bude lepšia ako Grep.

Zhodnotenie

Program som otestoval na dvoch rozdielnych zariadeniach s rovnakým procesorom *Intel(R) Core(TM) i5-5257U CPU @ 2.70GHz*. Na oboch zariadeniach program fungoval korektne a vypísal správne výsledky.

Maximálna veľkosť súboru na ktorej sme testovali bol 1.2 GB na ktorom sme preukázali zrýchlenie oproti sekvenčnému spracovaniu. Program sme porovnali aj s prvou verziou programu, ktorý bol vyvinutý minulý rok.

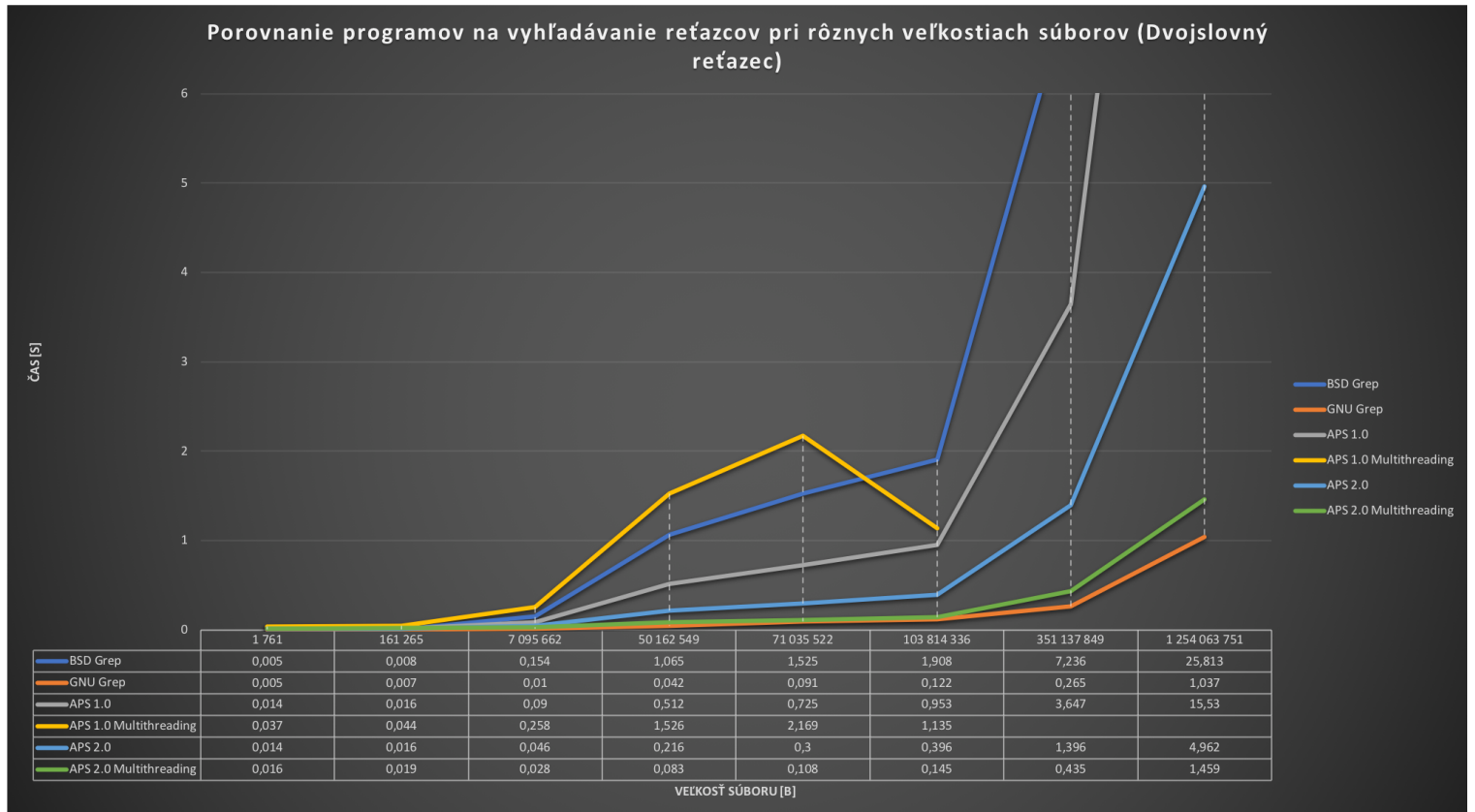
Merania som robil za sebou tak ako sú zapísané a každé meranie som zopakoval 3 krát aby sme obmedzili vplyv cachovania.



Obrázok 1 Porovnanie programov pri rozličných veľkostiach súboru a jednoslovnom reťazci (5 Bytov)

Ako môžeme vidieť na *Obrázok 1*, program APS 1.0 s možnosťou multi-threadingu nemal možnosť prehľadávať súbory väčšie ako 100 MB.

Tieto časy boli namerané programom *time* a boli zaznamenané reálne časy vykonávania (od spustenia po ukončenie volania).



Obrázok 2 Porovnanie programov pre dvojslovný reťazec (22 Bytov)

Na grafe (Obrázok 2) je zaujímavé porovnanie GNU a BSD Grepu, ktorý sa pri veľkom reťazci začína násobne spomaľovať.

Záver

Vytvoril som program, ktorý je rýchlejší ako grep, ale má možnosti na zlepšenie ako napríklad, zvolenie si počtu vlákien, ktoré budú pracovať paralelne alebo veľkosť časti súboru, ktorá bude prehľadávaná na jednom vlákne.

Tiež som narazil na chybu v BSD grepe, ktorý sa s veľkosťou súboru a vyhľadávaného reťazca začína násobne spomaľovať. To by som chcel ďalej adresovať vývojarom tohto programu.