

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 porovnajete výsledky (z hľadiska výkonnosti).

Implementácia

Program sme vytvorili v jazyku C s frameworkom OpenCL, ktorý umožňuje paralelne spracovanie. Program sme navrhli tak, že východia možnosť je spracovávať text sekvenčne.

Program používa algoritmus Knuth-Morris-Pratt (KMP), ktorý má časovú náročnosť uvedenú v tabuľke:

Predspracovanie	Vykonávanie	Pamäť
$\Theta(m)$	$\Theta(n)$	$\Theta(m)$

- m – dĺžka hľadaného reťazca
- n – dĺžka súboru v ktorom hľadáme

Paralelné spracovanie je možné vďaka frameworku OpenCL, ktorý poskytuje rozhranie do ktorého program (kernel) a data, ktoré vie vykonať ako SPMD (Single Program Multiple Data).

Pri paralelnom spracovaní sme museli rozdeliť text na niekoľko častí. Aby sme zabezpečili nájdenie reťazca aj v predelenej časti vždy sme predĺžili prehľadávanú časť $+ m - 1$ znakov. Na každej časti sme následne vykonávali algoritmus KMP a zapisovali výsledky do poľa.

Používateľská príručka

```
aps [-tldh] [-p pattern] [-f file]
-t multithreading
-l vypíše číslo riadku a celý riadok
-o vypíše offset bytov
-d vypíše rozšírený výpis na konci
-p vyhľadávaný reťazec
-f súbor v ktorom sa vyhľadáva
-h vypíše pomoc
```

Zdrojový kód

Zdrojové kódy nájdete aj na: <https://github.com/simonharvan/aps-search-string>