Paweł Troka 132334

Zadanie 2. Robot internetowy i graf WWW.

Modelowanie InternetuSprawozdanie

Wybrano domenę htttp://pg.edu.pl, ze względu na to, że jako jedyna z badanych spełniała wymóg ponad 3000 poprawnie linkowanych I nadających się do analizy podstron. Dokładna ich liczba wyniosła około 4500, co oznacza również spory potencjał programu.

Poza główną domeną politechniki analizie podlegały również (zależnie od potrzeb, także w celach czysto testowych).

* <http://cui.pg.edu.pl> (45 podstron, równa 2 średnica)
* <http://pg.edu.pl> (4967 postron)
* <http://www.sportowapolitechnika.pl> (208 postron)
* <http://csa.pg.edu.pl> (89 postron)
* <http://cas.pg.edu.pl> (22 podstron)
* <https://pomoc.pg.gda.pl> (0 podstron, ponieważ cała domena ma disallow dla botów)

# 1. Analiza czasowa w wielowątkowym ściąganiu dokumentów

Ze względu na ograniczony czas jak i ograniczenia sprzętowo-wydajnościowe do realizacji tego zadania wybrano domenę sportowapolitechnika.pl. Również stosunkowo wysoka prędkość pobierania z domeny ułatwiała zadanie.

W analizie postanowiono przetestować dwie rzeczy:

* Szybkość pobierania w zależności od maksymalnej dopuszczalnej liczby wątków

W środowisku .NET całe Task Parallel Library – a więc cała obszerna biblioteka dotycząca wątków, równoległości i zadań korzysta z obiektu statycznego znanego pod nazwą ThreadPool. Modyfikując więc jedną z propercji tego obiektu wpływamy na zachowanie wszystkich mechanizmów całej biblioteki. Postanowiono więc potestować wydajność aplikacji dla róznej liczby wątków.

Po wielokrotnych pomiarach okazało się, że optimum dla pisanego programu oscyluje gdzieś w granicach 150 wątków, dalsze ich zwiększanie raczej zmniejszało wydajność.

* Szybkość pobierania w zależności od stopnia zrównoleglenia pętli obliczeniowej

W środowisku .NET dosyć łatwo można ustawić dla każdej równoległej pętli oliczeniowej opcje równoległości, w tym stopień zrównoleglenia. O ile ilość wątków ustawiana w poprzednim punkcie ma charakter bardziej „globalny” i może źle wpływać na działanie niektórych modułów aplikacji na których działanie wcale nie chcieliśmy wpłynąć o tyle stopień zrównoleglenia pętli możemy ustawiać „bezkarnie” wpływając jedynie na wykonanie tej pętli.

Tutaj z kolei widać bardzo duży wzrost wydajności pomiędzy jednym dopusczalnym wątkiem (czyli tak naprawdę pętla niezrównoleglona) a dwoma. Dalsze zyski wydajności są już znacznie mniejsze ale i tak są znaczne.

Optimum wydaje się być w okolicach 6-10, co jest prawdopodobnie mocno związane z parametrami środowiska testującego (Core i7 720QM, 4 rdzenie, 8 wątków). Zwiększanie powyżej wartości 11 generalnie zmniejszało wydajność – narzut związany z dodatkowymi wątkami nie dawał poprawy wydajności.

# 2. Analiza grafu dla domeny http://pg.edu.pl

Sasasasa

Sasa

Sasas

Sasa

# 3. Rozkłady stopni IN, OUT jako wykres potęgowy