

УДК 004.032.24

С.С. Заверуха

(Тернопільський національний технічний університет імені Івана Пулюя)

ВИКОРИСТАННЯ ЗАСОБІВ БАГАТОПОТОКОВОГО ПРОГРАМУВАННЯ ДЛЯ ПРИШВИДШЕННЯ ПОБУДОВИ МАТРИЦІ ПОДІБНОСТЕЙ N-ВИМІРНИХ ВЕКТОРІВ

UDC 004.032.24

S.S. Zaverukha

USING MULTITHREADED PROGRAMMING TO ACCELERATE THE CONSTRUCTION OF A MATRIX OF SIMILARITIES OF N-DIMENSIONAL VECTORS

Починаючи з кінця 90-х років ХХ століття в зв'язку з необхідністю підвищення ефективності виконання операцій обчислювальною технікою для вирішення завдань в різних сферах набуває популярності парадигма мультипотокowego програмування. Під мультипотокком розуміється можливість програмного забезпечення підтримувати виконання кількох потоків одночасно.

При розробці програмного забезпечення часто виникає проблема вибору інструменту розробки. Одним з можливих рішень даної проблеми є використання мови програмування Java та бібліотеки (паketу) Concurrency, що дозволяє будувати власний застосунок використовуючи всю потужність багатоядерних систем. Попри це, мультипотоккове програмування є доволі складним процесом що має низку проблем, такі як: велика складність, непередбачуваність швидкості виділення потоку, ресурсний голод, стан гонки чи взаємне блокування потоків.

Можливими варіантами виходу з проблеми є використання атомарних типів, синхронізація методів, модифікатора volatile, заміна змінних (mutable) об'єктів на не змінні (immutable). Особливу увагу необхідно звернути на використання незмінних об'єктів, даний підхід є ефективним рішенням у випадку створення малих об'єктів, проте при створенні великих об'єктів це надзвичайно сильно завантажує підсистему пам'яті спочатку при створенні а потім при видаленні.

Основною метою даного дослідження є поліпшення архаїчного методу ієрархічної кластеризації, що складається з таких кроків: 1) пошук матриці відповідності та 2) попарне об'єднання вузлів в агломератному підході.

Завдання пошуку матриці відповідності векторів є доволі простим але затратним по часу при виконанні в одному потоці, однак його можна виконати швидко та безпечно з допомогою бібліотеки concurrency.

Паралельне програмування скорочує час виконання розпаралеленої частини, проте пришвидшення не приймає лінійної залежності від кількості виділених потоків. Виділення та запуск нового процесу забирає певний період машинного часу, окрім цього потокам необхідна синхронізація та програмний бар'єр для зберігання цілісності логіки.

Отже, використання паралельного програмування є ефективним при розробці додатків що обробляють великі масиви даних, проте вони накладають певні обмеження та складності.

Література:

1. Goetz, Brian; Joshua Bloch; Joseph Bowbeer; Doug Lea; David Holmes; Tim Peierls (2006). Java Concurrency in Practice. Addison Wesley. - 384 p.
2. Witten, Ian H. (2011). Data Mining - Practical Machine Learning Tools and Techniques with JAVA Implementations. Elsevier. - 416 p.