

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

КАФЕДРА ІНФОРМАТИКИ ТА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

Курсова робота з освітнього компоненту

«Технології паралельних обчислень. Курсова робота»

Тема: Алгоритм сортування злиттям та його паралельна реалізація мовою програмування Java

|  |  |
| --- | --- |
| **Керівник**:  ст. викладач  Дифучин Антон Юрійович  «Допущено до захисту»  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2024 р.  Захищено з оцінкою  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Члени комісії:  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **Виконавець**:  Бондаренко Максим Вікторович  студент групи ІП-13  залікова книжка № ІП-1304  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «26» травня 2024 р.  Інна СТЕЦЕНКО  Антон ДИФУЧИН |

**Київ – 2024**

**ЗАВДАННЯ**

1. Виконати огляд існуючих реалізацій алгоритму, послідовних та паралельних, з відповідними посиланнями на джерела інформації (статті, книги, електронні ресурси). Зробити висновок про актуальність дослідження.
2. Виконати розробку послідовного алгоритму у відповідності до варіанту завдання та обраного програмного забезпечення для реалізації. Опис алгоритму забезпечити у вигляді псевдокоду. Провести тестування алгоритму та зробити висновок про коректність розробленого алгоритму. Дослідити швидкодію алгоритму при зростанні складності обчислень та зробити висновки про необхідність паралельної реалізації алгоритму.
3. Виконати розробку паралельного алгоритму у відповідності до обраного завдання та обраного програмного забезпечення для реалізації. Опис алгоритму забезпечити у вигляді псевдокоду. Забезпечити ініціалізацію даних при будь-якому великому заданому параметрі кількості даних.
4. Виконати тестування алгоритму, що доводить коректність результатів обчислень. Тестування алгоритму обов’язково проводити на великій кількості даних. Коректність перевіряти порівнянням з результатами послідовного алгоритму.
5. Виконати дослідження швидкодії алгоритму при зростанні кількості даних для обчислень.
6. Виконати експериментальне дослідження прискорення розробленого алгоритму при зростанні кількості даних для обчислень. Реалізація алгоритму вважається успішною, якщо прискорення не менше 1,2.
7. Дослідити вплив параметрів паралельного алгоритму на отримуване прискорення. Один з таких параметрів – це кількість підзадач, на які поділена задача при розпаралелюванні її виконання.
8. Зробити висновки про переваги паралельної реалізації обчислень для алгоритму, що розглядається у курсовій роботі, та програмних засобів, які використовувались.

**АНОТАЦІЯ**

**Структура та обсяг роботи.** Пояснювальна записка курсової роботи складається з N сторінок, 5 розділів, містить 21 рисунків, 44 таблиць, 4 додатки, 17 джерел.

**Мета.** Реалізувати та проаналізувати паралельну реалізацію алгоритму сортування злиттям на мові Java з використанням ForkJoin Framework, що допоможе оптимізувати роботу алгоритму у виконанні задачі сортування.

Робота складається з п’яти розділів, в яких описується алгоритм та його відомі паралельні реалізації, розробляється послідовний алгоритм та аналізується його швидкодія, обирається та описується програмне забезпечення для розробки паралельних обчислень, розробляється паралельний алгоритм з використанням обраного ПЗ, досліджується ефективність паралельних обчислень алгоритму.

КЛЮЧОВІ СЛОВА: СОРТУВАННЯ ЗЛИТТЯМ, FORKJOIN FRAMEWORK, JAVA, ПАРАЛЕЛЬНИЙ АЛГОРИТМ

ЗМІСТ

[ВСТУП 5](#_Toc167488147)

[1 ОПИС АЛГОРИТМУ ТА ЙОГО ВІДОМИХ ПАРАЛЕЛЬНИХ РЕАЛІЗАЦІЙ 6](#_Toc167488148)

[1.1 Опис алгоритму 6](#_Toc167488149)

[1.2 Опис відомих паралельних реалізацій алгоритму 8](#_Toc167488150)

[2 РОЗРОБКА ПОСЛІДОВНОГО АЛГОРИТМУ ТА АНАЛІЗ ЙОГО ШВИДКОДІЇ 9](#_Toc167488151)

[3 ВИБІР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ РОЗРОБКИ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ ТА ЙОГО КОРОТКИЙ ОПИС 10](#_Toc167488152)

[4 РОЗРОБКА ПАРАЛЕЛЬНОГО АЛГОРИТМУ З ВИКОРИСТАННЯМ ОБРАНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ: ПРОЄКТУВАННЯ, РЕАЛІЗАЦІЯ, ТЕСТУВАННЯ 11](#_Toc167488153)

[5 ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ АЛГОРИТМУ 12](#_Toc167488154)

[ВИСНОВКИ 13](#_Toc167488155)

[СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 14](#_Toc167488156)

[ДОДАТКИ 15](#_Toc167488157)

[Додаток А. Назва додатку 15](#_Toc167488158)

[Додаток Б. Назва додатку 16](#_Toc167488159)

[Додаток В. Назва додатку 17](#_Toc167488160)

# ВСТУП

Алгоритми сортування лежать в основі багатьох операцій з обробки даних. Одним із найпоширеніших алгоритмів є сортування злиттям, який характеризується своєю стабільністю та ефективністю. Даний алгоритм базується на концепції «розділяй і володарюй», бо він розділяє масиви на менші підмасиви, сортує їх окремо, а потім зливає в один відсортований масив. Стабільним та ефективним його робить часова складність *O*(*n*log*n*), яка зберігається у найгіршому, середньому та найкращому випадках.

Однак, навіть ефективні алгоритми потребують оптимізації, коли мова йде про серйозне збільшення розмірів даних та суворі вимоги до продуктивності. Ефективним рішенням даної проблеми є розпаралелення алгоритмів. Паралельні обчислення дозволяють розподіляти навантаження між кількома потоками, які виконують задачу одночасно, що сильно зменшує загальний час виконання задачі. У контексті нашої задачі розпаралелення може значно прискорити процес виконання, оскільки різні частини масиву можуть сортуватись одночасно.

У сучасному світі все більше комп’ютерів мають багатоядерні процесори, які здатні виконувати багато задач одночасно. Також великі обсяги даних стають дедалі частішою реальністю в сучасних додатках. Усім цим обумовлена доцільність розпаралелення алгоритму сортування злиттям, оскільки це допоможе використовувати ресурси більш ефективно, зменшуючи при цьому час виконання.

# 1 ОПИС АЛГОРИТМУ ТА ЙОГО ВІДОМИХ ПАРАЛЕЛЬНИХ РЕАЛІЗАЦІЙ

## 1.1 Опис алгоритму

Алгоритм сортування злиттям (merge sort) є класичним алгоритм, що працює за принципом «розділяй і володарюй». Основна ідея полягає в рекурсивному поділі масиву на менші підмасиви до тих пір, поки кожен з них не буде уявляти з себе масив довжиною 1, після ці підмасиви зливаються назад в один відсортований масив.

Алгоритм сортування складається з трьох основних етапів:

1. Розділення. Розбиття масиву на дві частини.
2. Рекурсивне сортування. Рекурсивно застосовуємо алгоритм до розділених частин.
3. Злиття. Зливаємо дві частини в один відсортований масив.

Нижче наведено псевдокод алгоритму сортування злиттям. Функція mergeSort(arr) – головна функція алгоритму, яка виконує всю його логіку. В якості вхідного параметру вона приймає масив arr. Якщо довжина масиву менша або дорівнює 1, то він повертається без змін, оскільки це є базовим випадком рекурсії. В іншому випадку масив розбивається на дві частини за індексом, який визначає середину масиву: , де mid – індекс середнього елементу, len – довжина масиву. Обидві частини сортуються рекурсивно за допомогою викликів відповідних методів mergeSort(arr[0:середина]) та mergeSort(arr[середина:довжина arr]). Після сортування частин вони зливаються в один відсортований масив за допомогою функції merge(left, right), яка приймає обидві частини в якості параметрів. Функція merge(left, right) – є допоміжною функцією для злиття двох відсортованих масивів в один масив. Вона почергово порівнює елементи обох масивів і додає до результуючого масиву менший елемент. Коли елементи одного масиву закінчуються, до результуючого масиву додаються всі елементи іншого, які залишились нерозподіленими.

Псевдокод алгоритму:

***функція*** *mergeSort(arr):*

***якщо*** *довжина arr ≤ 1:*

***повернути*** *arr*

*mid = довжина arr / 2*

*left = mergeSort(arr[0:середина])*

*right = mergeSort(arr[середина:довжина arr])*

***повернути*** *merge(left, right)*

***функція*** *merge(left, right):*

*result = []*

*i = 0, j = 0*

***поки*** *i < довжина left і j < довжина right:*

***якщо*** *left[i] ≤ right[j]:*

*додати left[i] до result*

*i = i + 1*

***інакше****:*

*додати right[j] до result*

*j = j + 1*

***поки*** *i < довжина left:*

*додати left [i] до result*

*i = i + 1*

***поки*** *j < довжина right:*

*додати right[j] до result*

*j = j + 1*

***повернути*** *result*

Часова складність алгоритму становить *O*(*n*log*n*), що робить його достатньо ефективним для великих обсягів даних. Просторова складність становить *O*(*n*), оскільки потрібен простір для злиття підмасивів.

## 1.2 Опис відомих паралельних реалізацій алгоритму

Паралельні реалізації алгоритму сортування злиттям дозволяють прискорити сортування за рахунок розподілу навантаження між кількома процесорами або ядрами. Для виконання задач розпаралелювання даного алгоритму існує кілька підходів. Розглянемо основні відомі підходи до вирішення цієї задачі.

Бібліотека Fork/Join в Java дозволяє реалізовувати паралельні обчислення шляхом рекурсивного розподілу завдань на підзадачі. Кожна задача розбивається на дві підзадачі, які виконуються паралельно, після чого результати зливаються.

Інший підхід до паралельного сортування полягає у використанні звичайних потоків. У цьому випадку для кожної рекурсивної задачі створюється новий потік, який виконує сортування підмасиву.

Також існує підхід, який застосовується до розподілених систем з великою кількістю процесорів, з використанням MPI. У даному випадку різні процесори виконують сортування окремих частин масиву та обмінюються даними для злиття результатів.

# 2 РОЗРОБКА ПОСЛІДОВНОГО АЛГОРИТМУ ТА АНАЛІЗ ЙОГО ШВИДКОДІЇ

## 2.1 Розробка послідовного алгоритму

При розробці паралельного алгоритму за допомогою Fork/Join наявність окремої реалізації послідовного алгоритму є обов’язковою, оскільки при розділенні на підзадачі ми маємо порогове значення нижче якого розділення не може відбуватись, а підмасиви розміром менше цього значення будуть сортуватись послідовним сортуванням, оскільки розділення маленьких задач на підзадачі тільки сповільнить виконання роботи.

Нижче наведено код послідовного алгоритму сортування злиттям (рис. 2.1)

Зображення, що містить текст, знімок екрана, програмне забезпечення

Автоматично згенерований опис

Рисунок 2.1.1 – код послідовного алгоритму сортування злиттям.

Алгоритм починається з перевірки, щоб індекс початку масиву був меншим за індекс кінця, інакше це масив з одного елементу, який і так є відсортованим. Якщо індекс початку все-таки менше, то рекурсивно викликається сортування для двох підмасивів, на які було поділено початковий масив за індексом mid. Після сортування обидва підмасива зливаються у відсортований масив.

Злиття в свою чергу починається з виділення тимчасового масиву для збереження даних при злитті. Потім визначаємо індекси початку обох підмасивів та індекс для ітерації по тимчасовому масиву. Маємо цикл, який проходиться одночасно по двом підмасивам та порівнює між собою елементи цих підмасивів, менший з яких записуючи в результуючий масив, після чого здвигає індекси належним чином. Коли один з підмасивів повністю розподілиться, запишемо в тимчасовий масив усі елементи іншого підмасиву, які залишились. Вміст тимчасового масиву копіюється назад у вихідний масив.

## 2.2 Аналіз швидкодії послідовного алгоритму

Проведемо аналіз швидкодії послідовного алгоритму сортування злиттям. Для цього проведемо декілька замірів на різних розмірностях масиву та проаналізуємо отримані результати. У якості масиву для сортування буде виступати масив об’єктів будівель, а в якості компаратора буде використовуватись ціна кожної будівлі, що міститься в кожному окремому об’єкті.

Важливим етапом перед тестуванням є прогрівання, це допоможе отримати більш точні та стабільні результати вимірювань. Прогрівання дозволяє уникнути впливу різних факторів, які можуть спотворювати перші запуски програми, тому в нашому випадку перед проведенням замірів будемо обов’язково проводити прогрівання п’ятьма ітераціями запуску алгоритму.

Наступним етапом є проведення основного тестування, де ми маємо запустити наш алгоритм 20 разів на копіях одного й того самого масиву задля об’єктивності. На кожній ітерації відпрацювання програми отримуємо час виконання нашого алгоритму та інформацію про правильність сортування. У підсумку час виконання усіх ітерацій складається та ділиться на кількість цих ітерація, завдяки чому ми отримуємо середній час виконання сортування одного масиву, що робить наші дані для експерименту більш об’єктивними. Нижче наведено приклад виконання подібного експерименту (рис. 2.2.1) та таблиця з результатами експериментів для різної кількості елементів масиву (табл. 2.2.1).

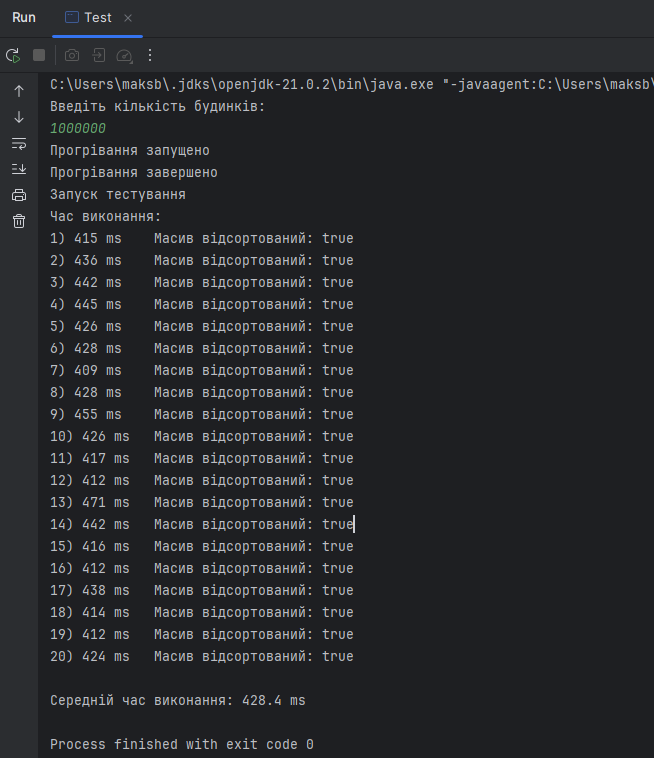


Рисунок 2.2.1 – Приклад обчислення середнього часу виконання.

Таблиця 2.2.1 – Результати експериментів для різної розмірності масивів.

|  |  |
| --- | --- |
| Кількість елементів | Середній час послідовного алгоритму, мілісекунд |
| 100000 | 28.05 |
| 250000 | 81.65 |
| 500000 | 191.15 |
| 1000000 | 428.4 |
| 1500000 | 713.65 |
| 2000000 | 966.7 |
| 3000000 | 1595.2 |
| 5000000 | 3094.5 |
| 10000000 | 6286.25 |

Судячи з результатів у таблиці бачимо очікуваний результат: час виконання зростає зі збільшенням розмірності масивів, але таблиці недостатньо для більш глибокого аналізу, тому побудуємо графіки, на яких буде більш наглядно видно, як змінюється швидкість виконання в залежності від розміру.

Для початку побудуємо графік залежності середнього часу виконання від кількості елементів (рис. 2.2.2).

Зображення, що містить текст, ряд, схема, Графік

Автоматично згенерований опис

Рисунок 2.2.2 – Графік залежності середнього часу від розмірності масиву.

Графік демонструє, як змінюється час виконання послідовного алгоритму при збільшенні розміру вхідних даних. Одразу бачимо, що графік має лінійно-логарифмічну залежність, що відповідає зазначеній часовій складності алгоритма: *O*(*n*log*n*). З ростом кількості елементів час виконання зростає більше, ніж лінійно, але менше, ніж квадратично. На прикладі 1000000 та 2000000 елементів, розмір збільшився вдвічі, а час збільшився в 2.257 рази. Також на графіку не спостерігається значних відхилень чи аномалій, що свідчить про стабільну роботу алгоритму на різних обсягах даних та відсутність суттєвих затримок чи надмірних витрат на конкретних етапах сортування. На великих обсягах даних (10 мільйонів) час виконання алгоритму становить більше 6 секунд, що свідчить про потребу в розпаралеленні алгоритму для оптимізації роботи на великих розмірностях.

# 3 ВИБІР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ РОЗРОБКИ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ ТА ЙОГО КОРОТКИЙ ОПИС

Для реалізації паралельних обчислень у рамках даної курсової роботи було обрано мову програмування Java та її технологію Fork/Join, яка надає інструменти для пришвидшення паралельних обчислень шляхом спроби використовувати всі доступні ядра процесора, що досягається за допомогою принципу «розділяй та володарюй», що ідеально підходить до нашого алгоритму.

Java є однією з найпопулярніших мов програмування, яка вирізняється стабільністю та активною підтримкою розробників з боку спільноти. До основних переваг можна віднести те, що програми можуть виконуватись на будь-якій платформі з встановленою JVM, Java забезпечує високий рівень безпеки і стабільності програми, має вбудовану підтримку багатопотоковості, що сильно допомагає в паралельних обчислення, має набір бібліотек, які допомагають у розробці додатків з паралельним обчисленням.

Бібліотека Fork/Join є інструментом для реалізації паралельних обчислень за принципом «розділяй і володарюй». Дана бібліотека дозволяє розбивати задачі на менші підзадачі, які виконуються паралельно, що сприяє пришвидшенню виконання алгоритму на багатоядерних процесорах. З основних переваг бібліотеки можна виділити автоматичне управління потоками, оскільки Fork/Join сам керує пулом потоків, ефективне використання ресурсів, що забезпечує високу продуктивність. Алгоритм сортування злиттям ідеально підходить до паралельної реалізації за допомогою Fork/Join завдяки рекурсивній природі та принципу «розділяй і володарюй».

В якості середовища розробки було обрано IntelliJ IDEA. Воно широко використовується для розробки програм на Java завдяки великій кількості інструментів, зручному інтерфейсу та підтримці великої кількості бібліотек.

# 4 РОЗРОБКА ПАРАЛЕЛЬНОГО АЛГОРИТМУ З ВИКОРИСТАННЯМ ОБРАНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ: ПРОЄКТУВАННЯ, РЕАЛІЗАЦІЯ, ТЕСТУВАННЯ

## 4.1 Проєктування

Оскільки ми використовуємо бібліотеку ForkJoin для паралельного виконання завдань, основною ідеєю є розділення задачі на менші підзадачі, які можна виконувати паралельно. Основними компонентами будуть масиви об’єктів, що будуть розділятись та зливатись, а також індекси, що вказують на границі підмасивів. При розробці паралельного алгоритму з використанням ForkJoin, стратегія паралельного виконання грає ключову роль в подальшій ефективності алгоритму.

Основні кроки для визначення стратегії виконання:

* Початковий масив даних буде розділятись на менші підмасиви, кожен з яких буде оброблятись як окрема задача. Для цього використаємо рекурсивний підхід, де кожна задача розділяється на дві половини до досягнення порогового розміру.
* Після розділення задач на менші підзадачі, підзадачі мають виконуватись паралельно. За паралельне виконання підзадач відповідає ForkJoinPool.
* Після завершення виконання підзадач необхідно злити їх результати в результуючий масив.

Також необхідно розглянути обробку базового випадку, коли розмір підмасиву стає досить малим, а його подальше ділення зменшує ефективність. У такому випадку до цього підмасиву застосовується послідовний алгоритм сортування.

## 4.2 Реалізація

# 5 ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ АЛГОРИТМУ

…представлені у таблиці 5.1…..

Таблиця 5.1. – Назва таблиці

|  |  |  |
| --- | --- | --- |
| Кількість елементів | Час послідовного алгоритму, мікросекунд | Час паралельного алгоритму, мікросекунд |
| 1000 | 2477 | 475 |
| 3000 | 2338 | 837 |

На рисунку 5.1….

…..(рис.5.3).

Рисунок 5.1. – Назва рисунку

Рисунок 5.3. – Результат виконання програми у консолі

# ВИСНОВКИ

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. …...
2. …..
3. …..
4. ….
5. …..

# ДОДАТКИ

## Додаток А. Назва додатку

## Додаток Б. Назва додатку

## Додаток В. Назва додатку