**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

**ННК «ІПСА»**

**Кафедра системного проектування**

**ЗВІТ**

**про виконання курсової роботи**

**з дисципліни**

**«Паралельні обчислення»**

**на тему: "Створення інвертованого індексу"**

**Студентка IІI курсу групи ДА-82**

**Хоміч Ліна Ігорівна**

**Керівник: Яременко В. С.**

**Київ – 2021**

**Теоретичні відомості**

Інвертований індекс являє собою структуру даних, в якій для кожного слова колекції документів співставленні всі документи, в яких це слово зустрілось. Зазвичай, інвертований індекс використовують як структуру для пошуку документів за тестом/текстами (близький до аналогу пошукової системи).

Як і було сказано раніше, кожне слово з колекції документів має відповідний масив, який зберігає назви файлів, в яких було знайдено це слово. Якщо у звичайних текстах, наприклад, книги, ми йдемо від сторінки до слів, у інвертованому індексу – як можна здогадатись з назви – все відбувається навпаки.

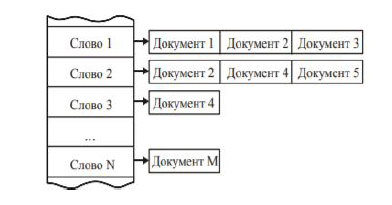


Рис. 1 – Приклад будови інвертованого індексу

Будувати інвертований індекс можна у декілька способів, основними з них є:

1. Однопрохідна індексація
2. Двопрохідна індексація
3. Блочна індексація

У своїй реалізації я використала однопрохідну індексацію

**Практична частина**

Для того щоб вирішити завдання я діяла з такими кроками:  
1) Створила функцію createListWithFilesName, для того, щоб зробити список з усіх текстів, які необхідні для побудови індексу

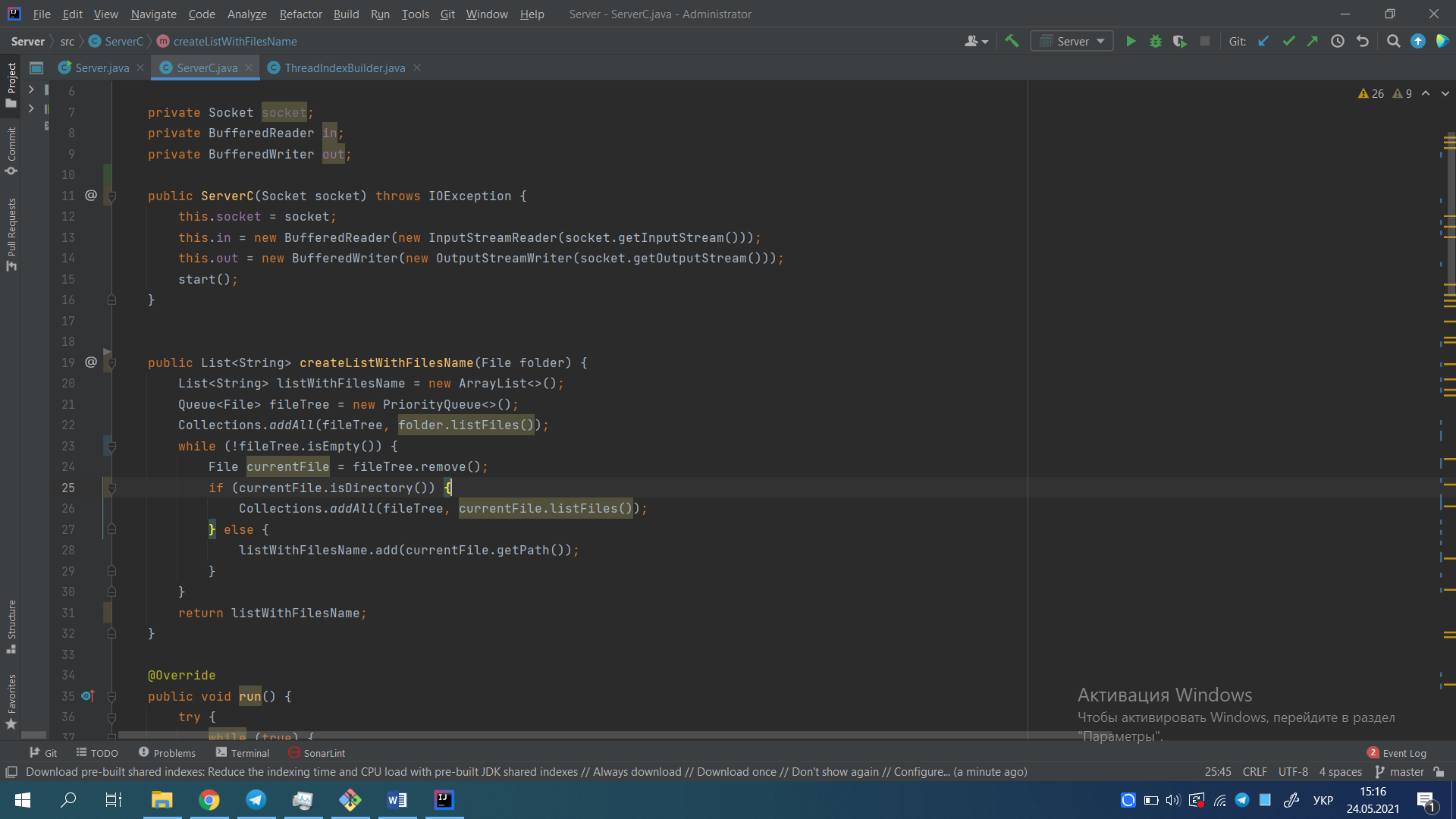


Рис. 1

2) Потім реалізувала побудову індексу з розпаралелюванням даним. Тобто на вхід потоку подається список усіх текстів, та початковий і кінцевий індекс, який обраховується в залежності від кількості потоків. У даній реалізації для кожного потоку створюється Map<String, HashSet<String>> indexTemporary = new HashMap();, які потім будуть об’єднуватися в одну. Побудова індексу реалізується таким чином: циклом проходимося по sublist та прочитуємо дані файли, розділяємо рядок на слова, проходимося циклом по кожному слову та записуємо у хешмапу за принципом, що якщо слово нове, то додаємо його як ключ, а значення це назва файлу, у якому воно було знайдене, якщо ж такий ключ уже є, то значення записуємо у Set.

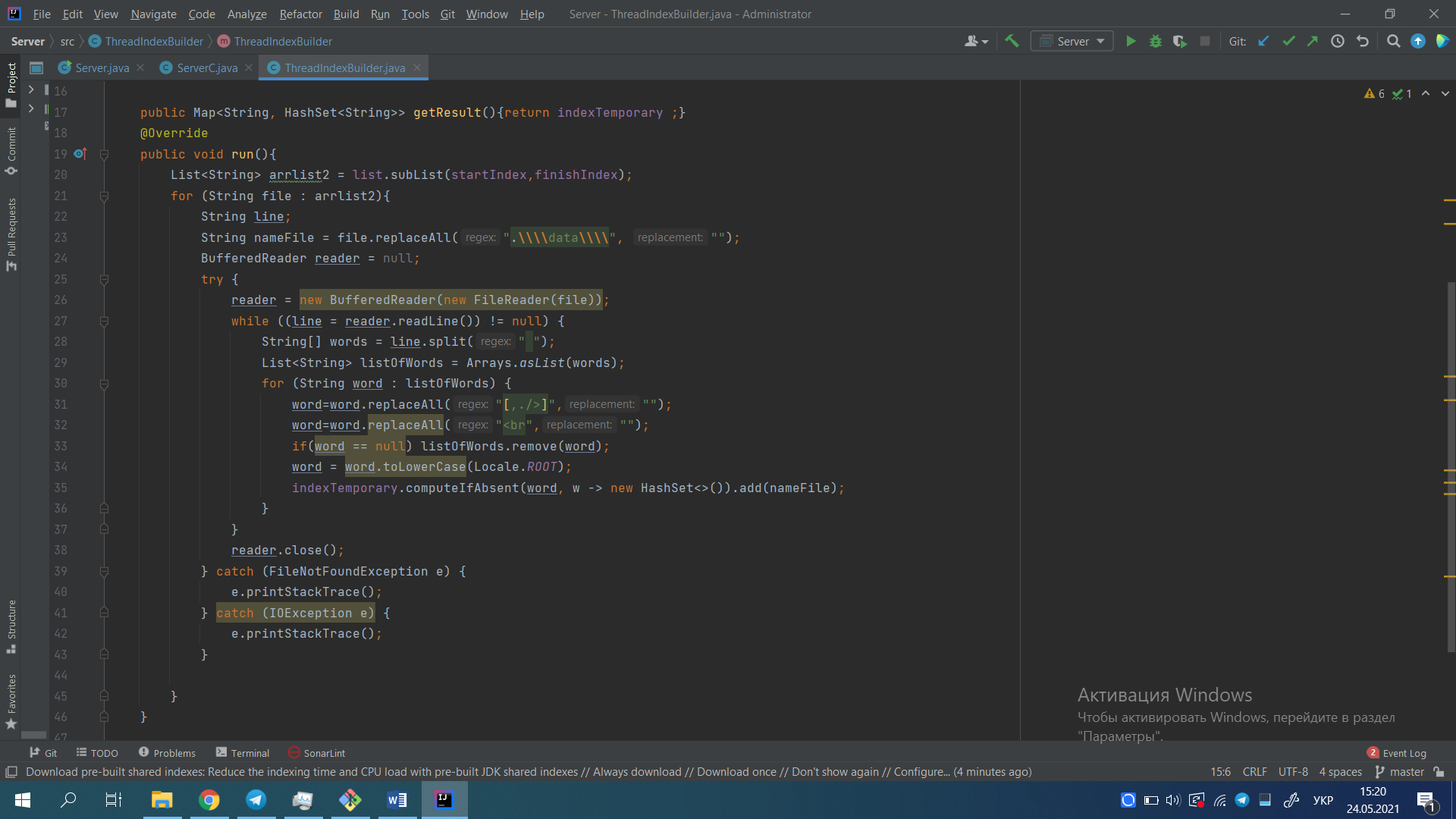


Рис. 2

3) Зливаємо потоки в один створюючи один загальний індекс.

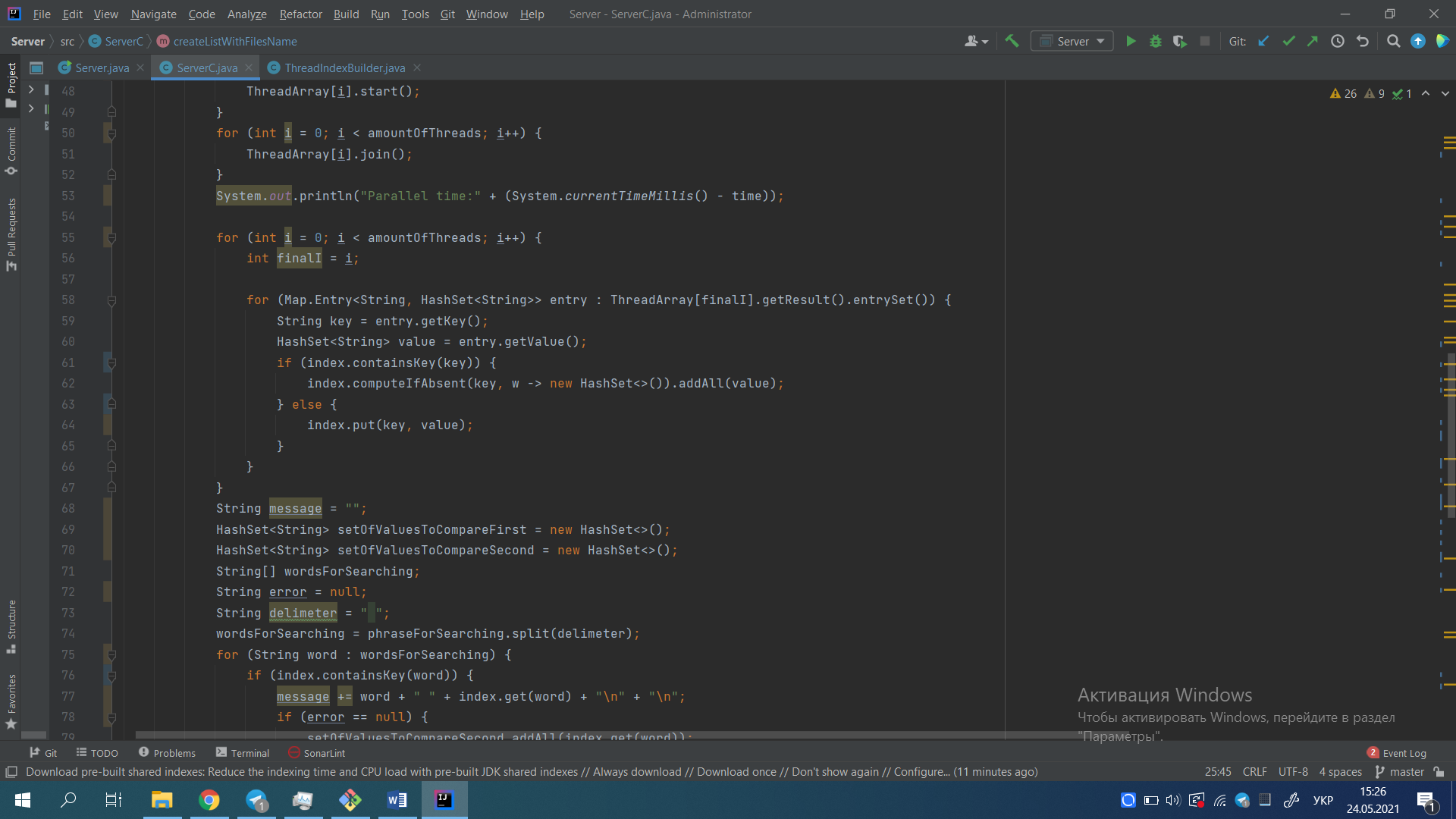
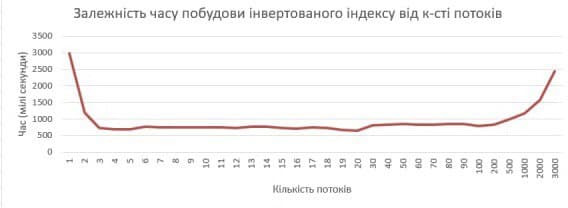


Рис.3

Щоб проаналізувати роботу алгоритму, провела дослідження по часу, який затрачений на створення індексу та побудувала графік залежності кільксті потоків від часу



**Висновки**

У ході виконання курсової роботи, було зроблено власну реалізацію паралельної структури даних під назвою інвертований індекс.

Було реалізовано кліент-серверний підхід, де клієнтом виступав користувач, який робить запит до індексу, який знаходится на сервері.

У ході написання проекту було використано систему контролю версій Git.