Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра автоматики та управління в технічних системах

**Лабораторна робота №7**

З дисципліни «Алгоритми та структури даних-2»

Тема:  «Дослідження операцій на тексті**»**

Виконали: Перевірив:

студенти групи ІТ-03 Смолій В. В.

Митєв А. Ю.

Яремчук Д. В.

Шевчук Д. Д.

Дата здачі: 19.05.21

Захищено з балом \_\_\_\_\_\_\_\_

Київ 2021

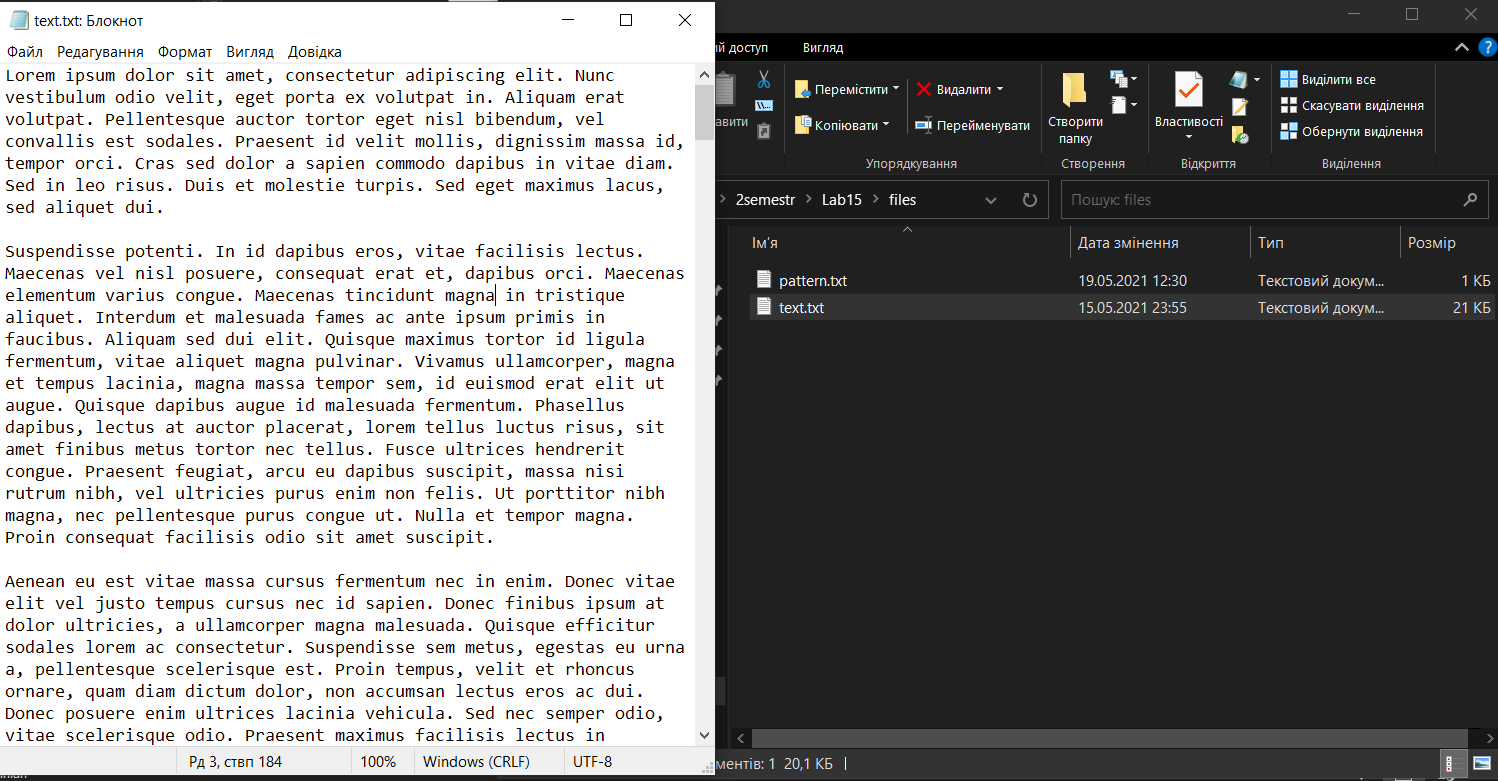
**Тема**: «Дослідження операцій на тексті»

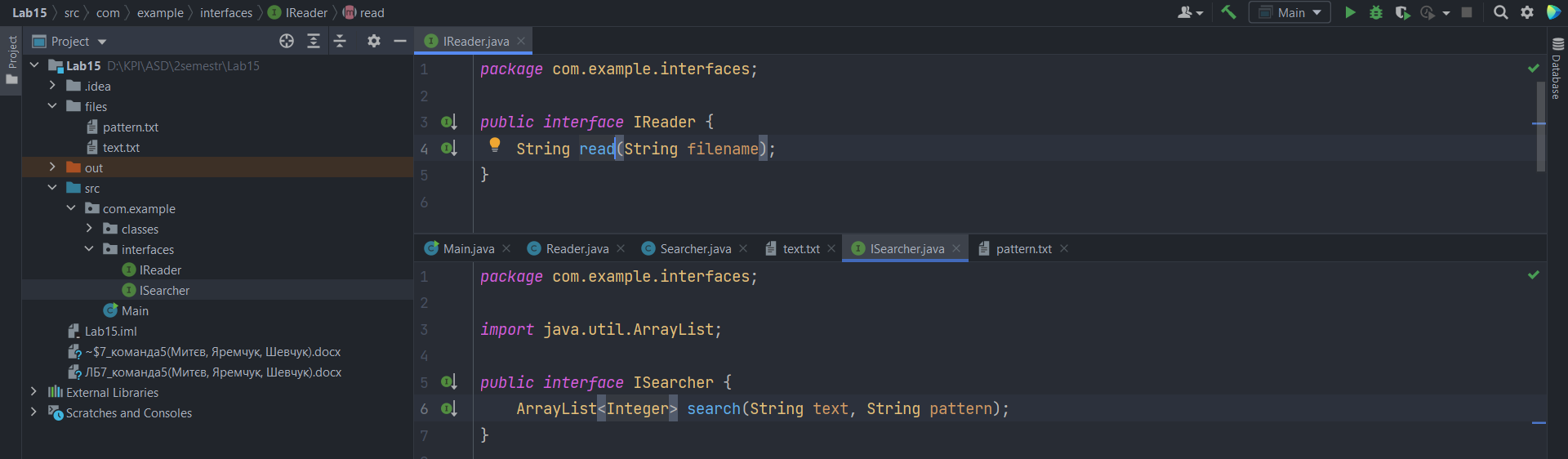
**Мета**: Дослідити особливості алгоритмів пошуку у текстових масивах даних та здійснити їх реалізацію.

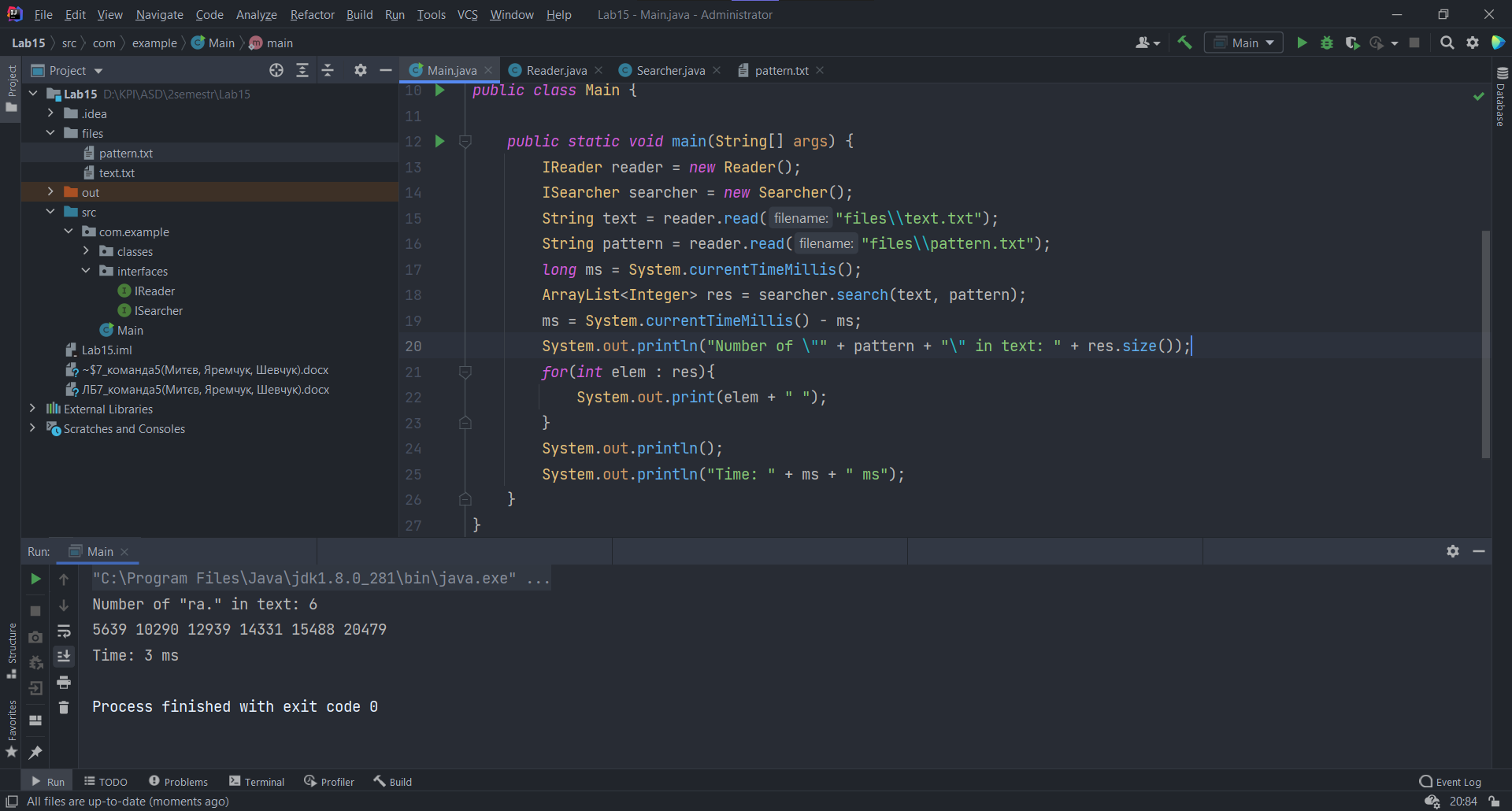
**Завдання:**

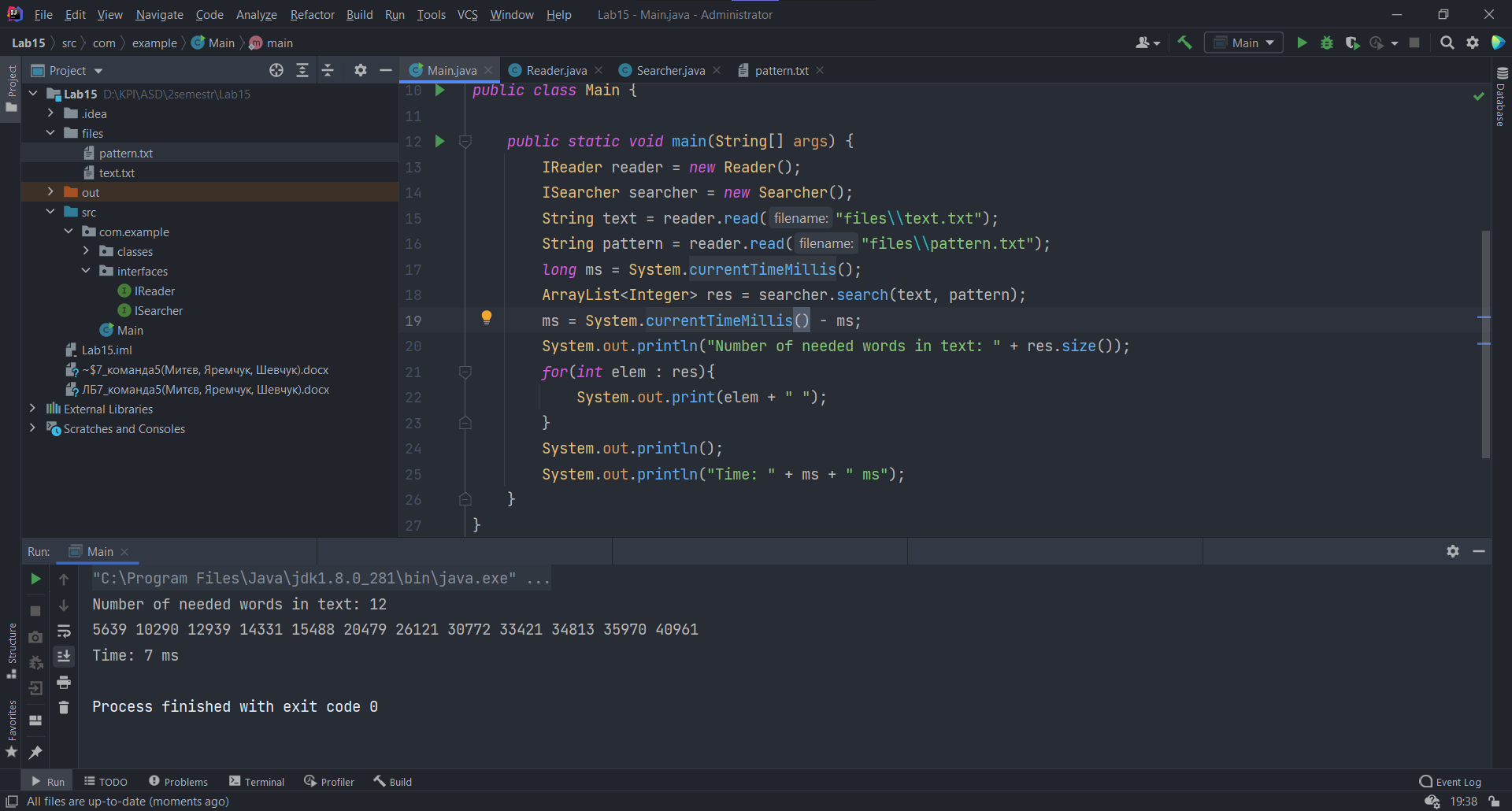
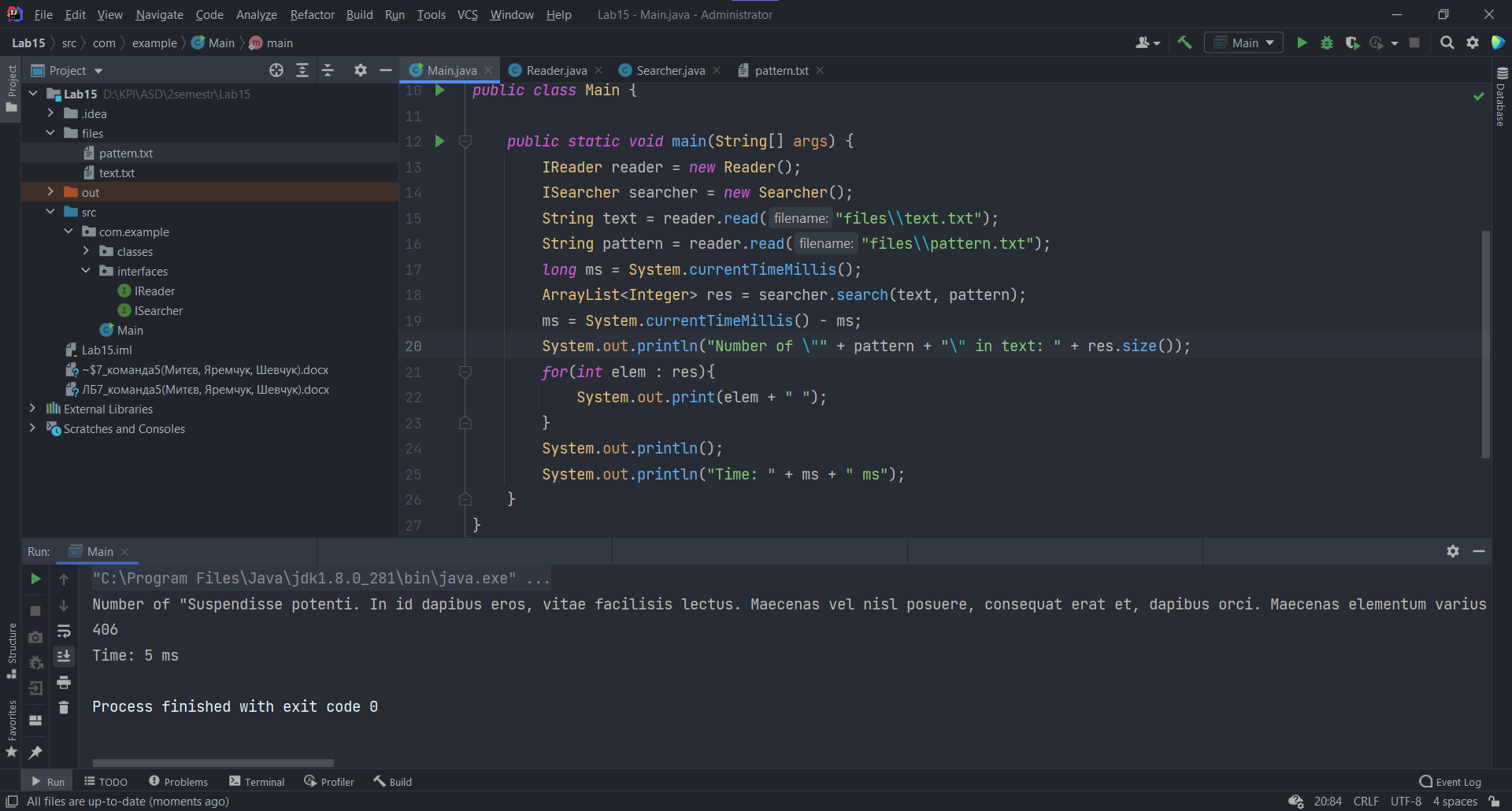
1. Ознайомитися з особливостями алгоритмів пошуку;
2. Реалізувати один з алгоритмів (на вибір, з обгрунтуванням) пошуку для фрази.
3. Визначити залежність трудомісткості від об'єму тексту (не менш ніж 20 кб plain-текст) та довжини шаблону, навести отримані оцінки з поясненнями результатів.

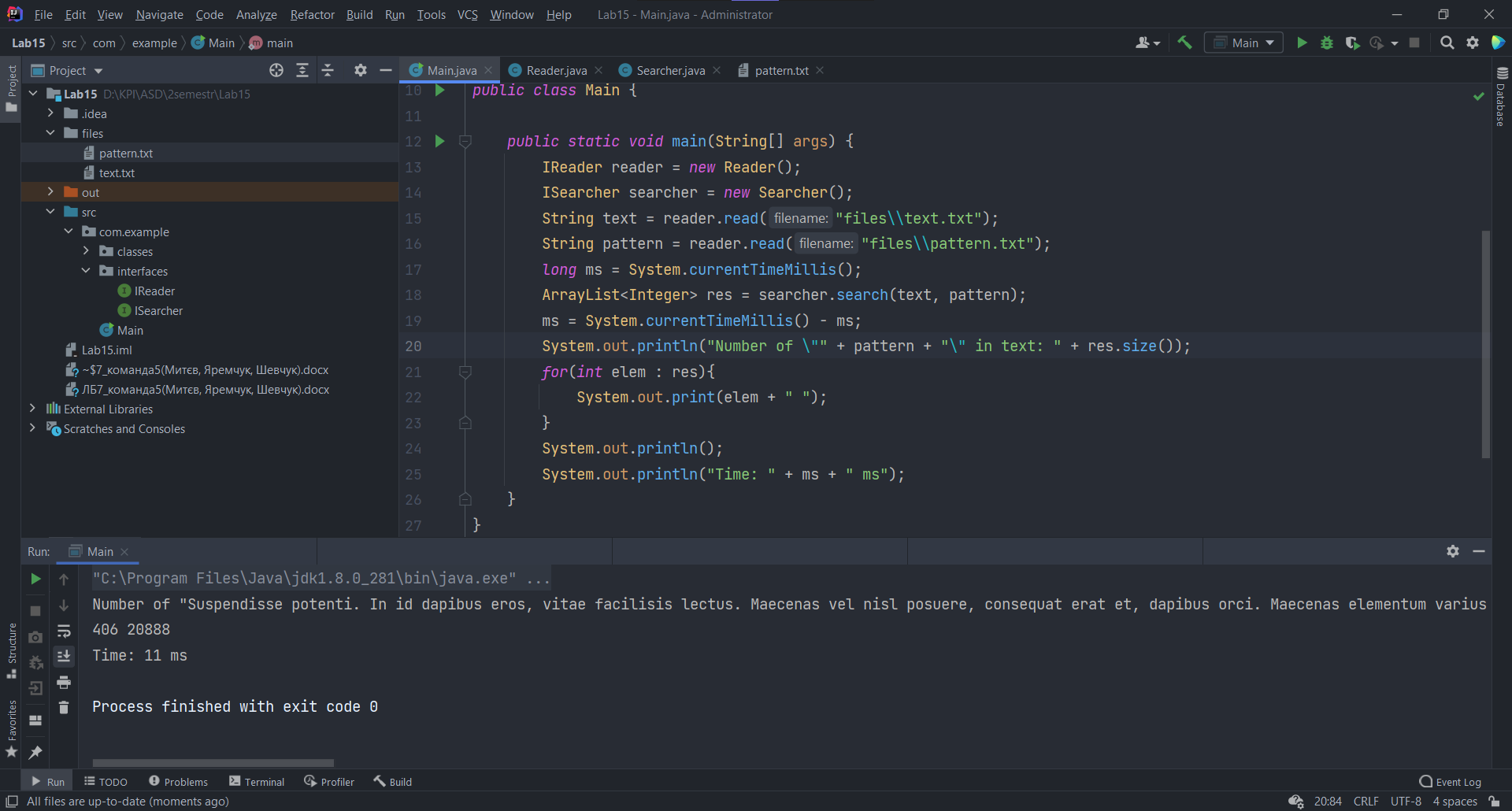
**Хід роботи**

1. Створили файл з текстом на 20 кб та файл з словом/фразою, яку треба знайти:
2. Створили інтерфейси для роботи з файлами: IReader (метод для зчитування з файлу), ISearcher (метод для пошуку в тексті):



1. Ми обрали спосіб пошуку Рабіна-Карпа, в основі якого лежить порівняння хешкодів підрядків тексту і шаблону. Він доволі простий в реалізації та має теоретичну складність в найкращому випадку O(n + m), де m – довжина слова, фрази для пошук, n – довжина тексту
2. Результат пошуку короткого рядку в 20кб і 40кб файлі з текстом та довгого рядку в 20кб і 40кб файлі з текстом:



******

Ми обрали звичайний спосіб хешування реалізований в мові Java. Він виявився доволі ефективним і при тестах ні разу не помилявся.

Тобто по часу видно, що на час виконання впливають і довжина тексту і довжина рядка, який треба знайти.

Складність нашого алгоритму складає O((n – m + 1) + cm) = O(n + m(c-1)):

c – кількість зсувів, тобто коли хеш-коди рівні

m – довжина рядка, який треба знайти

n – довжина тексту, в якому шукаємо

Цей алгоритм буде набагато швидшим за звичайний пошук в лоб, адже в будь-якому випадку порівняння хеш-кодів буде швидшим за порівняння рядків.

***Висновки***

Ми ознайомилися з такою структурою даних як хештаблиця, дізналися її основні операції та час їхнього виконання.

Як і раніше, ми все виконували разом, тобто кожен брав участь у розробці кожного підзавдання, головне, що ми показали бажання та старанність в дослідженні алгоритмів пошуку підрядків.