

## ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

Харківський національний університет радіоелектроніки  
кафедра ЕОМ

## ТЕМА

Кваліфікаційна робота  
Другий рівень (магістр)

### Автор

Бас Є.О.

ст. гр. КСМм-20-1

### Керівник

Барковська О.Ю.

доц. каф. ЕОМ

## МЕТА КВАЛІФІКАЦІЙНОЇ РОБОТИ

- є дослідження впливу використання різних методів сегментації та шумопригнічення на точність та швидкість розпізнавання сканованих документів.

## ЗАДАЧІ КВАЛІФІКАЦІЙНОЇ РОБОТИ

- дослідження загальних параметрів систем розпізнавання сканованих документів;
- дослідження існуючих методів шумопригнічення та сегментації зображень;
- розробка системи впорядкованого зберігання сканованих документів
- адаптація методів шумопригнічення сканованих документів під різні типи обчислювачів;
- оцінка таймінгу роботи;
- аналіз отриманих результатів

3

## АКТУАЛЬНІСТЬ КВАЛІФІКАЦІЙНОЇ РОБОТИ

- збільшення обсягів інформації у цифровому вигляді;
- зростання вимог щодо швидкості обробки та доступу до інформації;
- підвищення продуктивності в пошуку інформації;
- вдосконалення якості розпізнаного тексту;
- широкий спектр областей застосування.

4

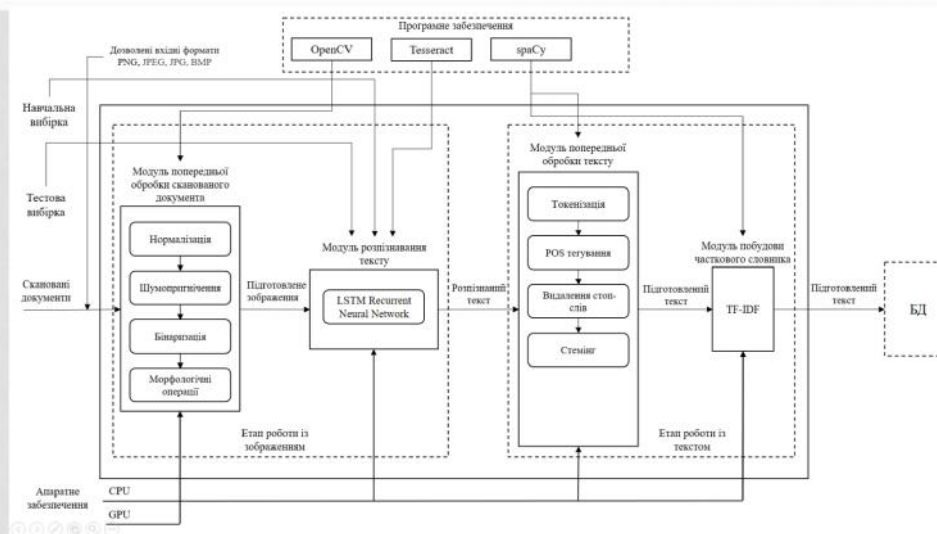
## ОГЛЯД ІСНУЮЧИХ АНАЛОГІВ У СФЕРІ ТЕСТОВОГО ПРОЦЕСІНГУ

Існуючі аналоги:

- ABBYY Finereader
- ABBYY Flexicapture
- Adobe Acrobat DC

5

## РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. МОДЕЛЬ СИСТЕМИ ВПОРЯДКОВАНОГО ЗБЕРІГАННЯ ДОКУМЕНТІВ



6

## РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. ПОПЕРЕДНЯ ОБРОБКА ЗОБРАЖЕННЯ

In 1830 there were but twenty-three miles of railroad in operation in the United States, and in that year Kentucky took the initial step in the work west of the Alleghanies. An Act to incorporate the Lexington & Ohio Railway Company was approved by Gov. Metcalf, January 27, 1830. It provided for the construction and re-

Оригінал

In 1830 there were but twenty-three miles of railroad in operation in the United States, and in that year Kentucky took the initial step in the work west of the Alleghanies. An Act to incorporate the Lexington & Ohio Railway Company was approved by Gov. Metcalf, January 27, 1830. It provided for the construction and re-

Фільтрація

In 1830 there were but twenty-three miles of railroad in operation in the United States, and in that year Kentucky took the initial step in the work west of the Alleghanies. An Act to incorporate the Lexington & Ohio Railway Company was approved by Gov. Metcalf, January 27, 1830. It provided for the construction and re-

Бінаризація

In 1830 there were but twenty-three miles of railroad in operation in the United States, and in that year Kentucky took the initial step in the work west of the Alleghanies. An Act to incorporate the Lexington & Ohio Railway Company was approved by Gov. Metcalf, January 27, 1830. It provided for the construction and re-

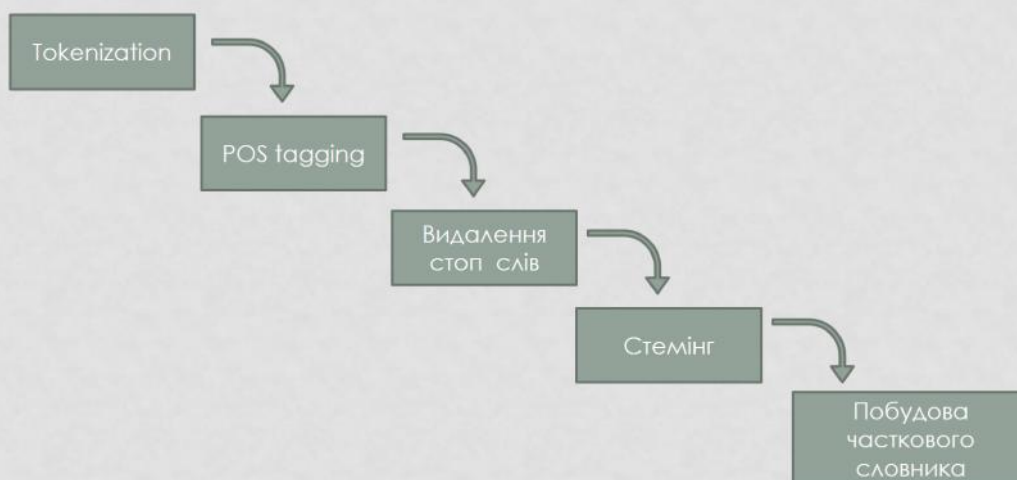
Застосування  
морфологічних  
операцій

Сегментація

In 1830 there were but twenty-three miles of railroad in operation in the United States, and in that year Kentucky took the initial step in the work west of the Alleghanies. An Act to incorporate the Lexington & Ohio Railway Company was approved by Gov. Metcalf, January 27, 1830. It provided for the construction and re-

7

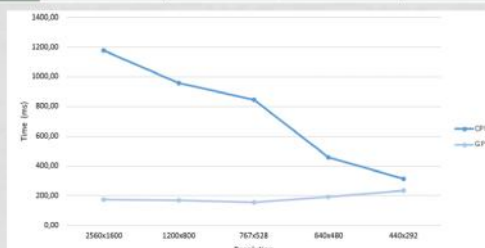
## РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. АНАЛІЗ ТЕКСТУ



8

## АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ ПРИСКОРЕННЯ ОБРОБКИ ЗОБРАЖЕННЯ

| Розмір зображення | Час виконання |        | Прискорення |
|-------------------|---------------|--------|-------------|
|                   | CPU           | GPU    |             |
| 2560x1600         | 1176,00       | 173,45 | 6,780051888 |
| 1200x800          | 955,33        | 167,67 | 5,697813121 |
| 767x528           | 844,00        | 153,76 | 5,489073881 |
| 640x480           | 456,67        | 190,43 | 2,398081535 |
| 440x292           | 315,33        | 234,67 | 1,34375     |



Графік прискорення

9

## АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ ПРИСКОРЕННЯ АНАЛІЗУ ТЕКСТУ

| Метод           | Кількість документів |       |       |
|-----------------|----------------------|-------|-------|
|                 | 1                    | 10    | 25    |
| Без прискорення | 3.5 с                | 39 с  | 107 с |
| nlr.pipe        | 3.6 с                | 28 с  | 72 с  |
| Розпаралелення  | 2 потоки             | 4.5 с | 21 с  |
|                 | 4 потоки             | 5.2 с | 11 с  |



Графік прискорення

10

## ВИСНОВКИ

- В процесі роботи над дипломним проектом було проведено дослідження параметрів, які повинна задовольняти система розпізнавання сканованих документів. Також були розглянуті існуючі методи обробки зображення для шумопригнічення та сегментації.
- На підставі проведеного дослідження була запропонована та розроблена система впорядкованого зберігання сканованих документів з виділенням ключових слів. В процесі розробки системи була проведена адаптація методів шумопригнічення та сегментації сканованих документів під різні типи обчислювачів. Також для вирішення завдання аналізу тексту було запропоновано методи прискорення обробки тексту.
- В ході дослідження були проведені оцінка таймінгу роботи розробленої системи впорядкованого зберігання сканованих документів та аналіз отриманих результатів.