

ДОДАТОК А
Графічний матеріал атестаційної роботи

Харківський національний університет радіоелектроніки
кафедра ЕОМ

МЕТОДИ ПРИСКОРЕНОЇ ОБРОБКИ ТЕКСТОВОЇ
ІНФОРМАЦІЇ
ДЛЯ ОРГАНІЗАЦІЇ СХОВИЩ ІНФОРМАЦІЙНИХ
ОБ'ЄКТІВ

Атестаційна робота
Другий рівень (магістр)

Автор

Зайцева С.Г.

ст. гр. КСМм-19-1

Керівник

Барковська О.Ю.

доц. каф. ЕОМ

МЕТА АТЕСТАЦІЙНОЇ РОБОТИ

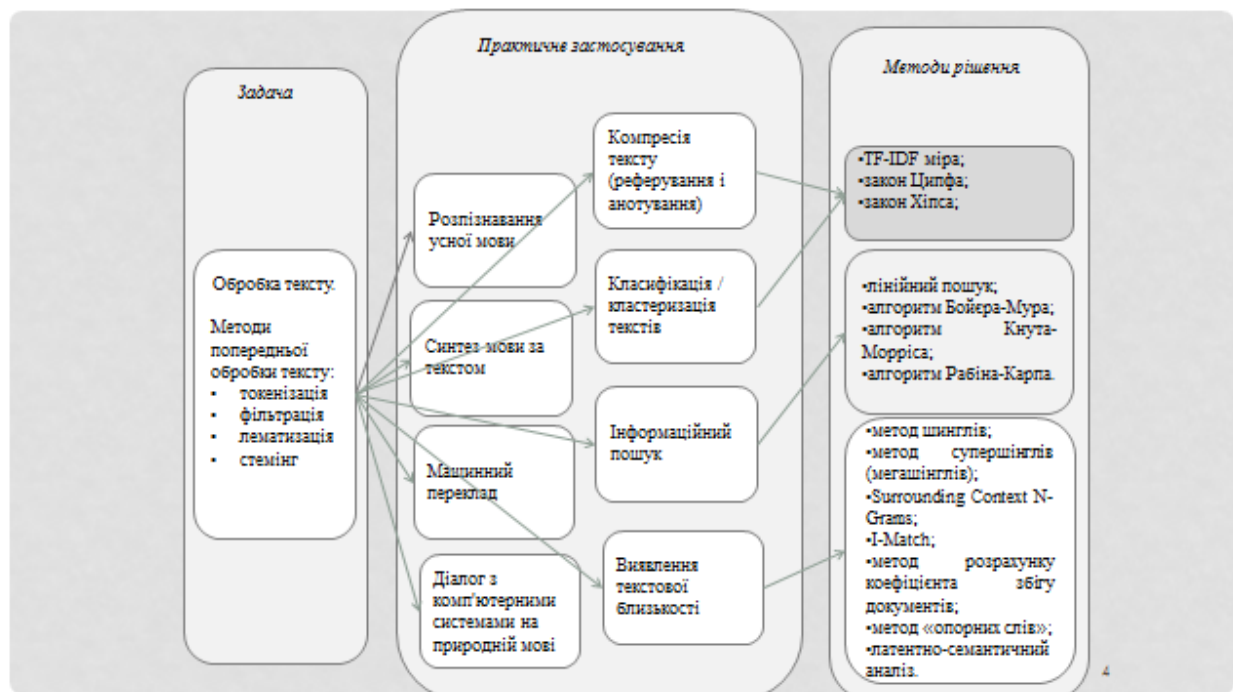
- створення моделі сховищ інформаційних об'єктів із модифікованими та вдосконаленими методами прискореної обробки текстової інформації.



ЗАДАЧІ АТЕСТАЦІЙНОЇ РОБОТИ

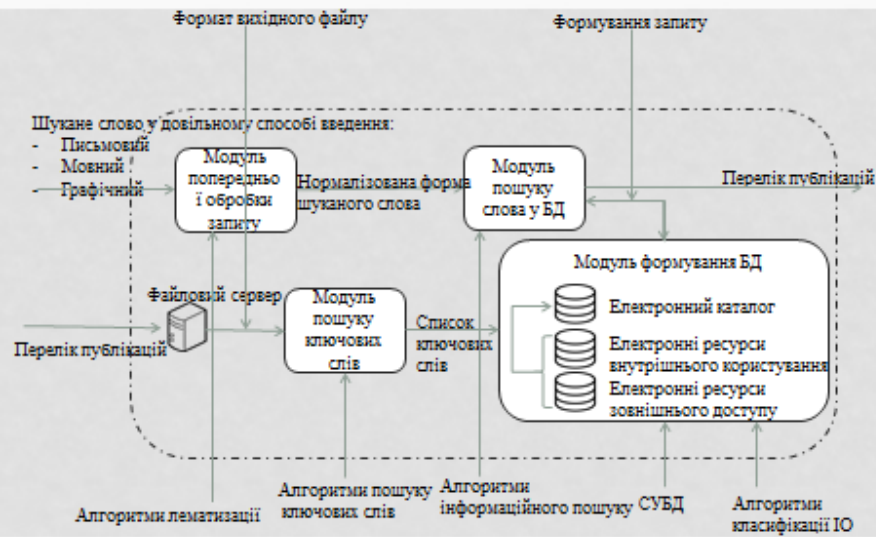
- розробка моделі сховища інформаційних об'єктів;
- аналіз методів попередньої обробки тексту на виявлення можливості реалізації на системах із масовим паралелізмом;
- дослідження впливу характеристик обчислювальної системи на час реалізації модифікованого методу визначення ваги слів у корпусі тексту на основі методу BoW;
- аналіз отриманих результатів.

3



4

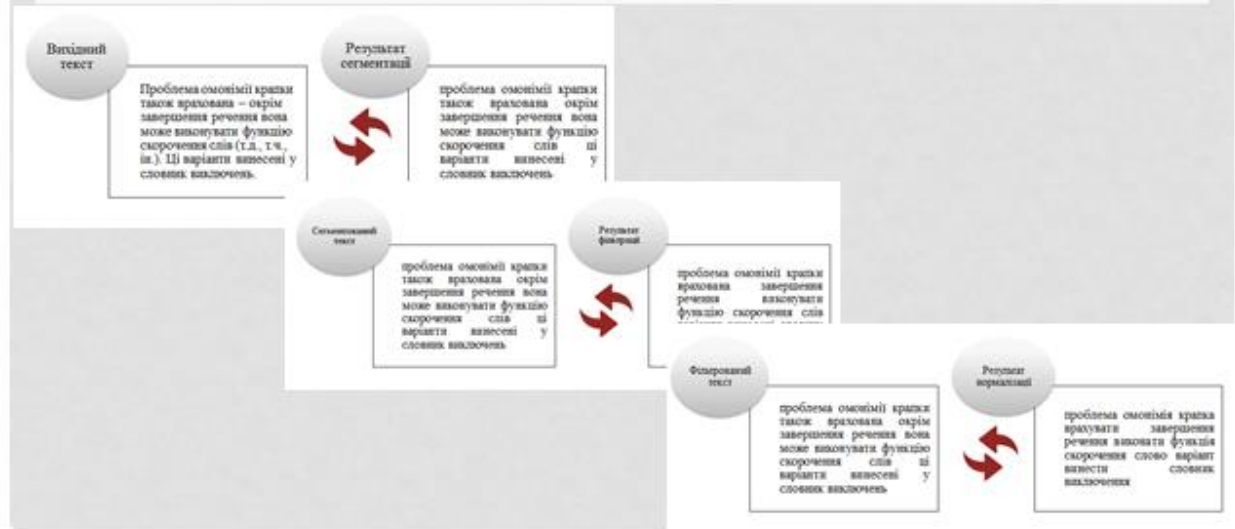
РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. МОДЕЛЬ АВТОМАТИЗОВАНОЇ БІБЛІОТЕЧНОЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ ВНЗ



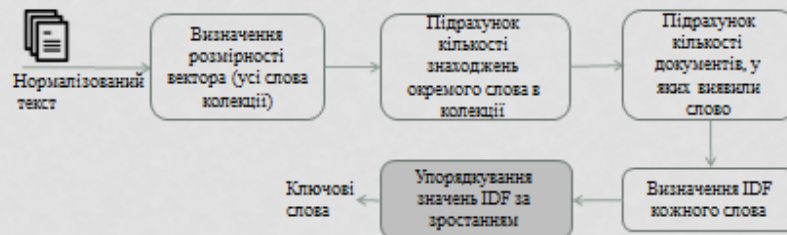
РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. ОРГАНІЗАЦІЯ ЕТАПУ НАКОПИЧЕННЯ ІНФОРМАЦІЇ



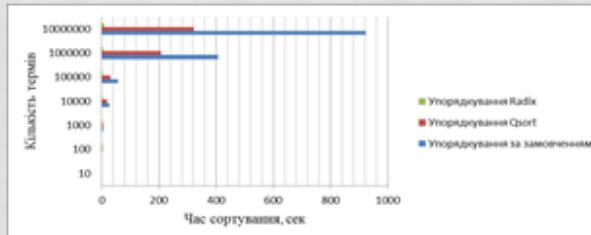
РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. ТЕСТУВАННЯ МЕТОДІВ ПОПЕРЕДНЬОЇ ОБРОБКИ ТЕКСТУ НА ГРАФЕМНОМУ ТА МОРФОЛОГІЧНОМУ РІВНЯХ



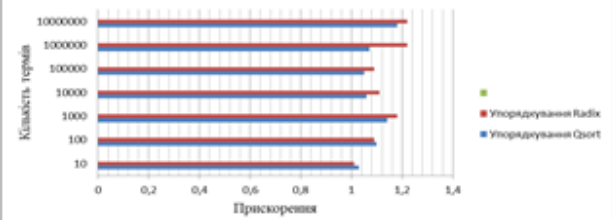
РІШЕННЯ ПОСТАВЛЕНОЇ ЗАДАЧІ. ДЕТАЛІЗАЦІЯ РОБОТИ СИНТАКСИЧНОГО РІВНЯ



АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ



Час упорядкування термів



Отримане прискорення для блоку синтаксичної обробки

ВИСНОВКИ

- запропонована модель сховища інформаційних об'єктів із модифікованими та вдосконаленими методами прискореної обробки текстової інформації.
- проведено аналіз методів попередньої обробки тексту на виявлення можливості реалізації на системах із масовим паралелізмом, який показав можливість та необхідність реалізації методів інформаційного пошуку та побудови частотного словнику на високопродуктивних обчислювальних системах, оскільки мають явний паралелізм даних;
- виконано дослідження впливу характеристик обчислювальної системи на час реалізації модифікованого методу визначення ваги слів у корпусі тексту завдяки використанню алгоритмів прискореного упорядкування елементів.

Аналіз отриманих результатів показав, що запропоноване впровадження сортування ваг термів на системах із масовим паралелізмом у побудованому частотному словнику скорочує час роботи блоку синтаксичного рівня запропонованої моделі майже на 18%. При цьому видно, що для малих обсягів даних прискорення майже відсутнє. Для великих обсягів даних прискорення складає майже 100 разів у порівнянні із послідовним сортуванням, що використовується за замовчуванням.