

WEB-ОРІЄНТОВАНА СИСТЕМА ПЕРЕВІРКИ ТЕКСТІВ НА ПЛАГІАТ

Фриз П.П.

Тернопільський національний економічний університет, магістрант

I. Постановка проблеми

Сучасний розвиток інформаційних технологій, а саме розвиток мережі Інтернет, значно спростили процес обміну інформацією. Важливого значення набуває проблема визначення унікальності даних в інформаційному просторі з метою виключення явища плагіату у процесі ідентифікаційного аналізу наявних ресурсів баз даних [1].

II. Мета роботи

Метою дослідження є розробка принципів, методів і алгоритмів пошуку плагіату в інформаційних ресурсах (текстових файлах та базах даних) та їх програмна реалізація.

III. Особливості програмної реалізації системи

Узагальнена модель автоматизованої системи перевірки текстів на плагіат (рисунок 1) презентує структурну взаємодію базових блоків, що формують програмне середовище системи. Блок введення вхідних даних реалізує інтерфейсну людино-машинну взаємодію. Блок ієрархічної структуризації даних формує набір елементарних об'єктів дослідження. Ієрархічна структуризація вхідної інформації зумовлює підтримку єдиного підходу до формування інформаційного забезпечення досліджуваного ресурсу, що передбачає можливість оперування даними нетекстового формату.

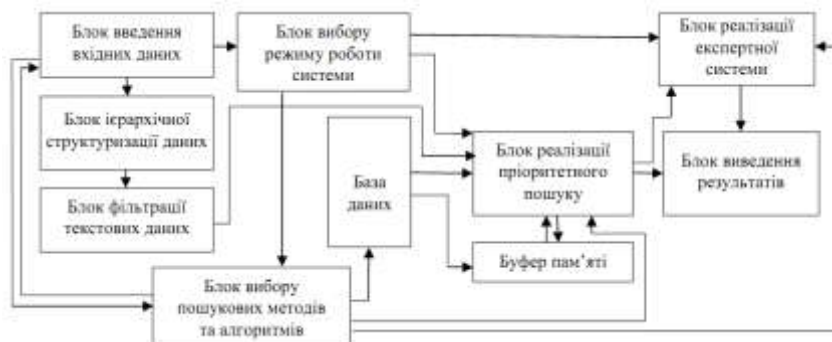


Рисунок 1 - Структура системи

Блок фільтрації дозволяє оптимізувати набір вхідних текстових даних, шляхом мінімізації інформативних ресурсів засобами ієрархічної фільтрації. Блок вибору режиму роботи призначений для встановлення пріоритетності пошукових умов, які впливають на кінцевий вибір методів і алгоритмів перевірки текстів на плагіат у середовищі автоматизованої системи. Блок вибору пошукових методів та алгоритмів забезпечує вибір набору робочих алгоритмів з урахуванням вимог користувача до умов проведення пошукових процесів. Блок реалізації пріоритетного пошуку забезпечує підтримку обраних режимів та реалізує виконання вказаних методів з ідентифікованим набором робочих алгоритмів. Блок виведення результатів презентує результати перевірки у зручному для користувача вигляді.

Коефіцієнт плагіату визначається відношенням:

$$P = 100 \frac{\beta_p}{\beta_s}, \quad (1)$$

де β_s - загальна кількість структурних елементів у відфільтрованих вхідних даних; β_p - кількість структурних елементів у яких було виявлено плагіат.

Висновок

Результатом роботи системи є визначення коефіцієнту наявності плагіату в досліджуваному документі та ілюстративне виведення ідентифікованого збігу семантичних конструкцій.

Список використаних джерел

1. Мокін, В. Б. Автоматизована система перевірки текстів на плагіат [Текст] / В. Б. Мокін, С. В. Бевз, О. В. Говенко, І. А. Білоус // Вісник Вінницького політехнічного інституту. – 2010. – № 5. – с. 12 – 17.