

Інформаційна технологія оптимізації пошукового просунення сайтів

Виконав: ст.гр. 2КН-17м

Горобець Ю.В.

Науковий керівник:

PhD, професор Савчук Т.О.

Об'єкт, предмет та мета дослідження

- ▶ Об'єктом дослідження є процес пошукового просування сайтів з використанням сучасних методів SEO-оптимізації сайтів.
- ▶ Предметом дослідження є програмні засоби, для оптимізації пошукового просування сайтів.
- ▶ Метою дослідження є оптимізація пошукового просування сайтів за рахунок застосування сучасних методів SEO і оптимізації створення семантичного ядра та метатегів сайту.

Актуальність

- ▶ SEO-оптимізація є важливим інструментом для просування бренду в мережі Інтернет, але цей процес є рутинним і займає багато часу, в частині створення семантичного ядра сайту та створення метатегів. Тому, є актуальною задачею розробка інформаційної технології для оптимізації пошукового просування сайтів SEO, а також автоматизації створення семантичного ядра сайту і генерації метатегів.

Математична модель створення метатегів

- ▶ Для генерації змісту тегу *title*, можна застосувати таку формулу: $Title = sitename + documentname$,
- ▶ Для генерації тегу *description*, застосовується формула: $Description = sentence \cup page_{keywords}$,
- ▶ Для генерації тегу *keywords* застосовується формула: $keywords = cluster_{name} \cup page + page_{keywords}$,

Математична модель генерації семантичного ядра сайту

▶
$$P_{x,y} = \frac{\sum_{i=1}^n (x_i - M_x)(y_i - M_y)}{(N-1)\sigma_x\sigma_y},$$

(2.1)

- ▶ де N - це потужність кластерів;
- ▶ x_i, y_i - значення парних змінних; частота слів i в документі x і в документі y ;
- ▶ M_x, M_y - це значення для x і y , яке показує середню частотність цих всіх слів у документі x і y , $\sigma_x \sigma_y$ - стандартне відхилення для документів x і y .

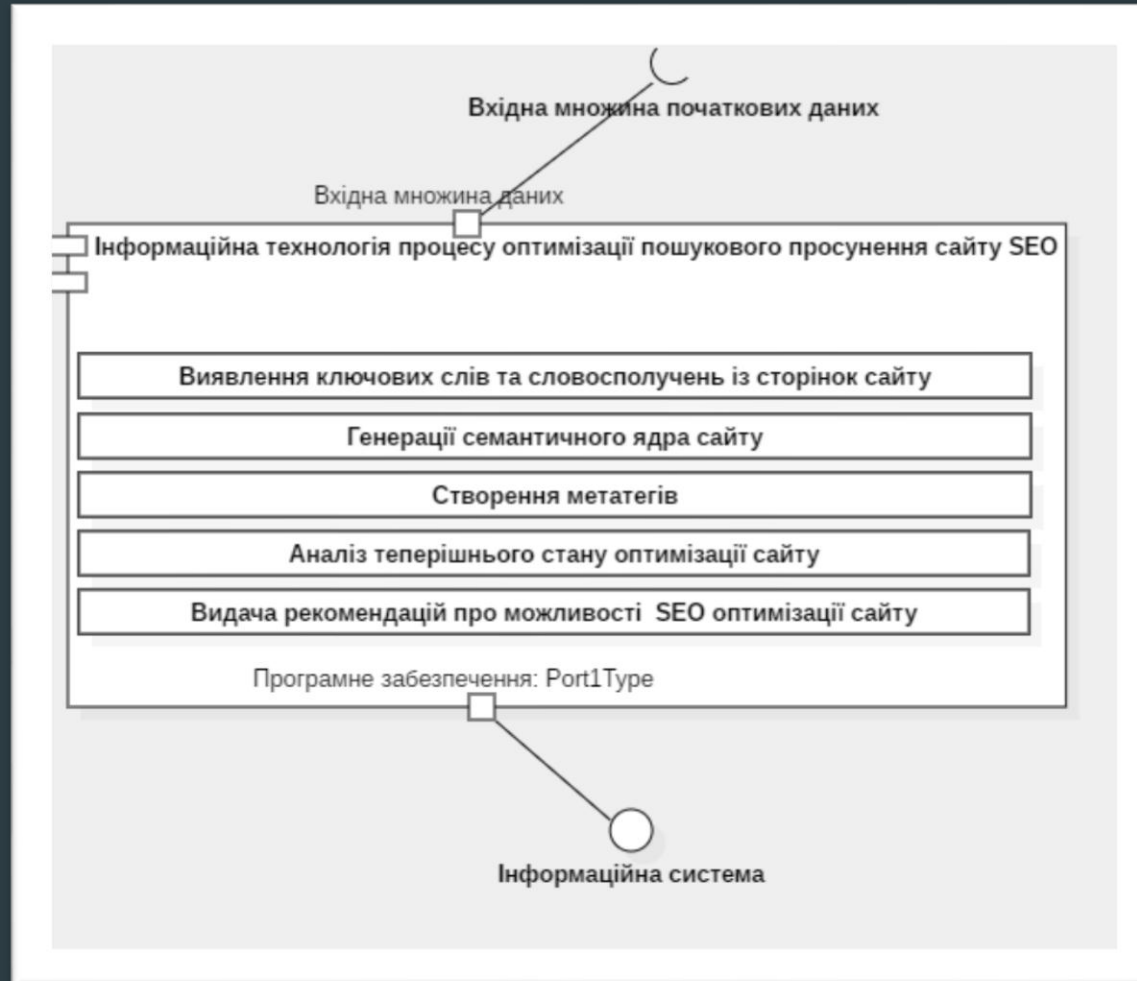
Математична модель виявлення ключових слів

- ▶ Ймовірність того, що фраза є ключовою вираховується за теоремою Баєса.
$$Pr[key | T, D] = \frac{Pr[T|key] \times Pr[D|key] \times Pr[key]}{Pr[T,D]}, \quad (2.6)$$
- ▶ де T - це TD-IDF значення, обчислене раніше за формулою (2.5);
- ▶ D - відстань від початку документу до першої появи фрази (кількість фраз перед даною фразою);
- ▶ $Pr [T | key]$ - це ймовірність того, що фраза, яка розглядається має TD-IDF значення T ;
- ▶ $Pr [D | key]$ - це ймовірність того, що фраза з'являється в документі на відстані D ;
- ▶ $Pr [key]$ - це ймовірність того, що із усіх фраз в документі, ця фраза являється ключовою;
- ▶ $Pr [T, D]$ використовується для нормалізації результатів у діапазоні від $[0, 1]$. Після чого фрази сортуються по ймовірності бути ключовими з урахуванням T і D , і k ключових фраз вилучаються вершини цього відсортованого списку.

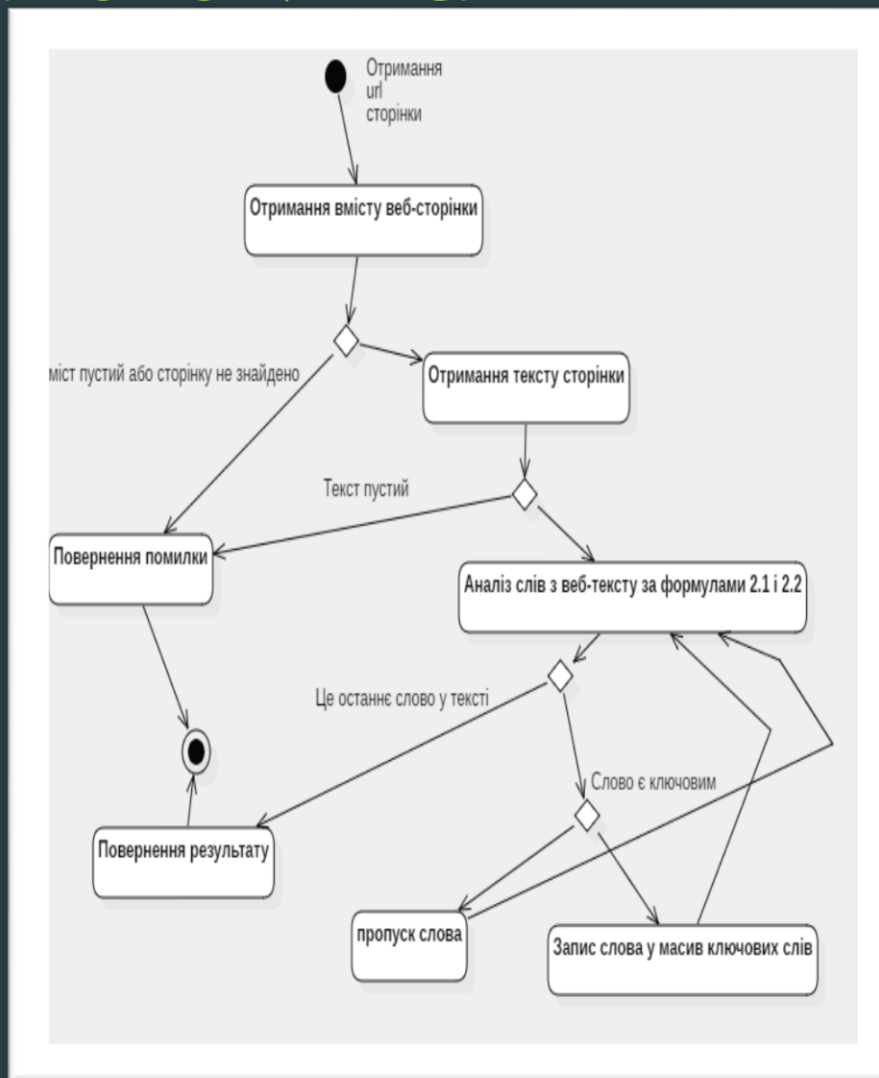
Аналізу теперішнього стану оптимізації

- ▶ $H = \{Cache, Assets, Scache, Saturation, Position, External\}$
- ▶ де H - це множина всіх параметрів, які впливають на SEO оптимізацію,
- ▶ *Cache* - кеш активований,
- ▶ *Assets* - файли скриптів, стилів, зображень,
- ▶ *Scache* - серверне кешування,
- ▶ *Saturation* - відсоток насиченості ключових слів,
- ▶ *Position* - позиції в пошукових системах,
- ▶ *External* - кількість зовнішніх посилань.

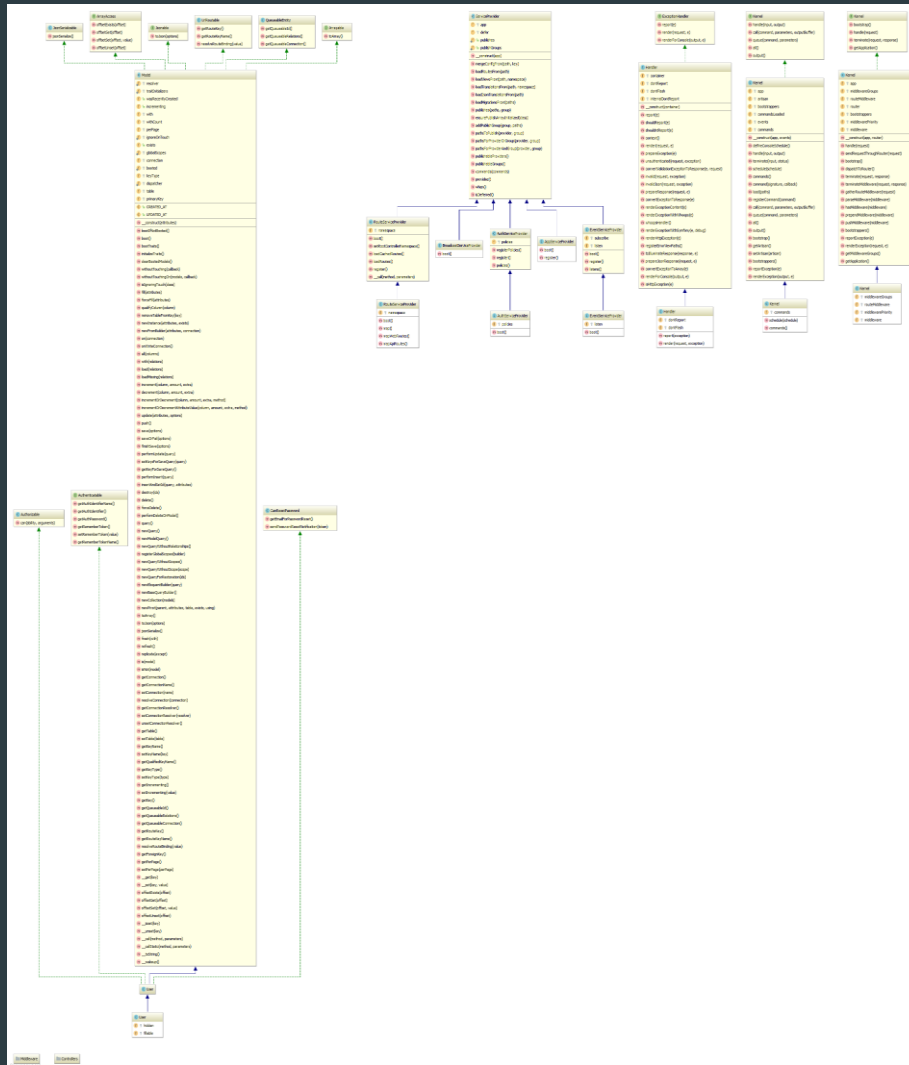
Структура інформаційної технології оптимізації пошукового просування сайтів



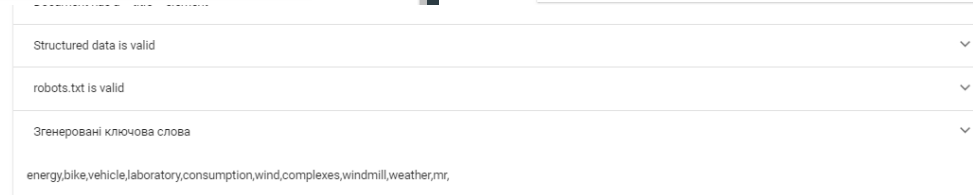
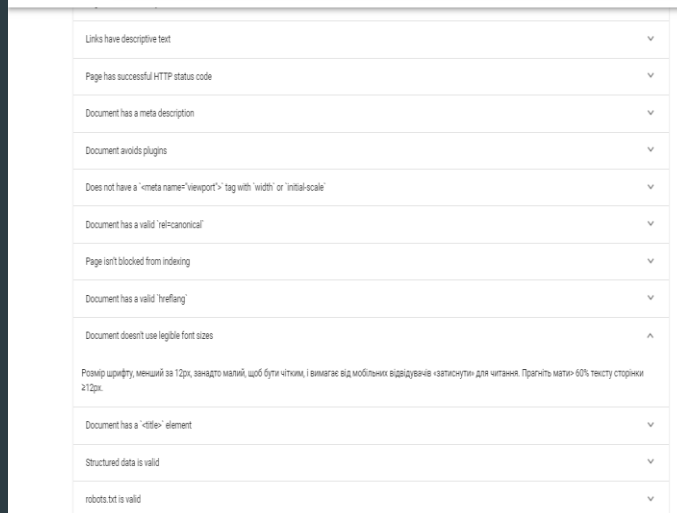
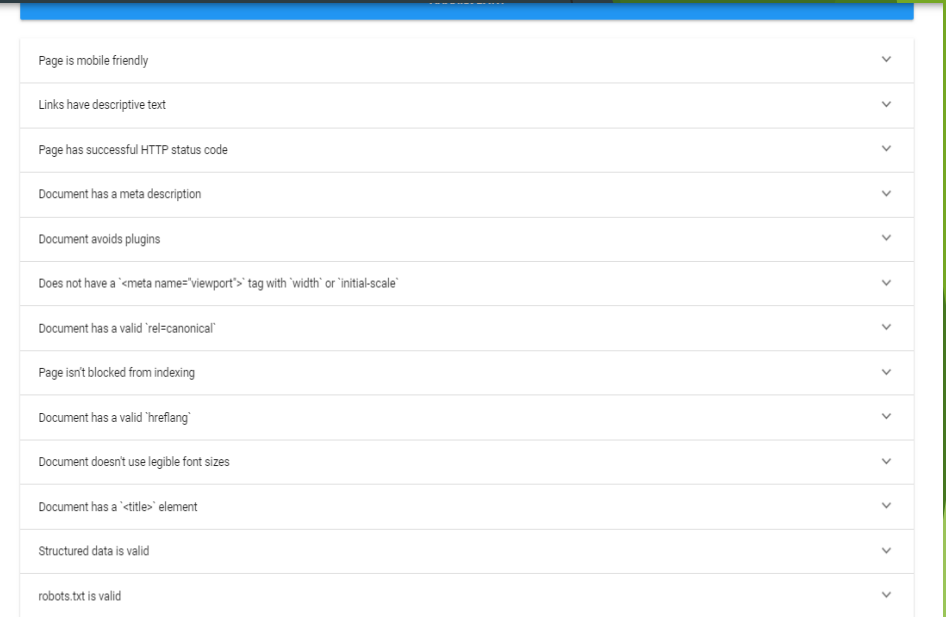
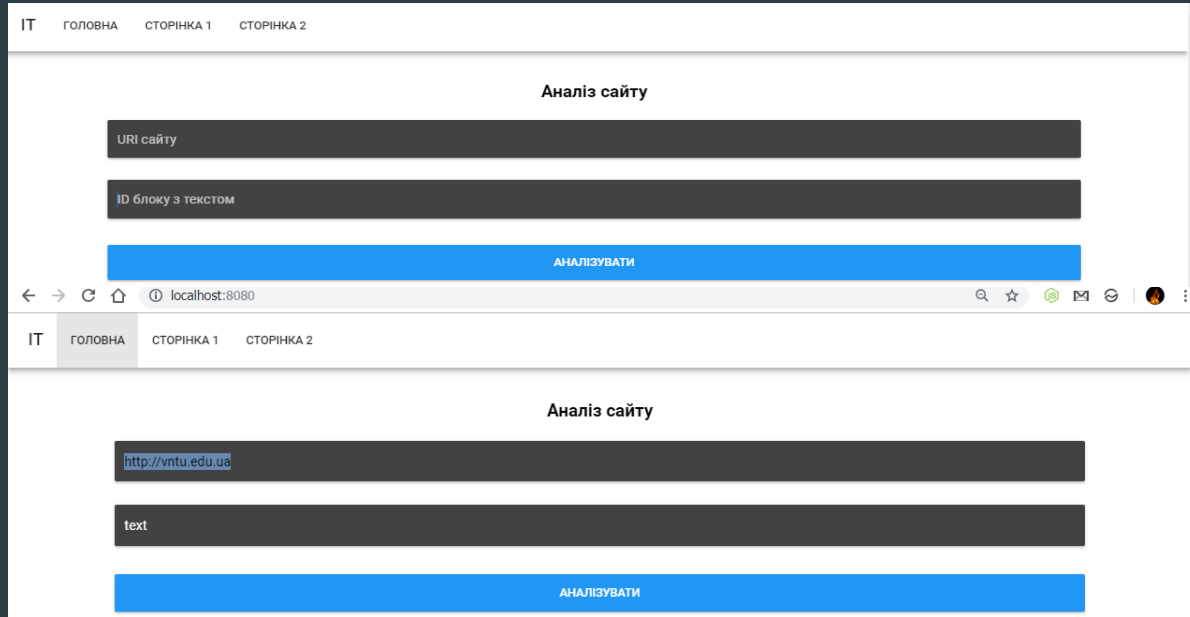
Діаграма активності виявлення ключових слів



Діаграма класів



Приклад роботи програми



energy,bike,vehicle,laboratory,consumption,wind,complexes,windmill,weather,mr,

Основні результати роботи

- ▶ Удосконалено метод виявлення ключових слів TFxIDF до якого додано Баєсових класифікатор, що дозволить виявляти ключові слова з тексту в незалежності до його тематики
- ▶ Математично описано задачу генерації семантичного ядра сайту та метатегів, що дозволить спроектувати відповідну інформаційну технологію.
- ▶ Удосконалено метод ієрархічної кластеризації за рахунок використання критерію Пірсона, що дозволить автоматично згенерувати семантичне ядро веб-сайту та метатеги для сторінок у відповідно до загальної тематики сайту.
- ▶ Основні теоретичні та практичні результати роботи представлені на 3 конференціях, написана одна стаття у фаховому виданні, отримано 2 свідоцтва про реєстрацію авторського права на твір.
- ▶ Результати дослідження отримані під час виконання магістерської кваліфікаційної роботи підтверджують підвищення ефективності створення семантичного ядра сайту за рахунок автоматизації цього процесу, результати показали, що для веб-сайтів, які знаходяться у середині пошукової видачі генерується семантичне ядро з новими ключовими слова, які підвищують релевантність сайту в пошуковій видачі на 25-35%. Також, автоматично створені рекомендації щодо пошукового просування сайтів є актуальними і дадуть відповідний ефект після наступного ранжування сайту, а саме підвищення SEO оптимізації на 20-22%.

Дякую за увагу!