

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління  
(повна назва)

Кафедра Автоматизації проектування обчислювальної техніки  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)  
(рівень вищої освіти)

Моделі впливу характеристик сайта на його позицію  
у видачі пошукових систем  
(тема)

Виконав: студент 2 курсу, групи СКСм-22-1

Оленич П. С.  
(прізвище, ініціали)

Спеціальність 123 Комп'ютерна інженерія  
(код і повна назва спеціальності)


Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

Освітня програма \_\_\_\_\_  
Спеціалізовані комп'ютерні системи  
(повна назва освітньої програми)

Керівник проф. Свірь І. Б.  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

  
(підпис)

Чумаченко С. В.  
(прізвище, ініціали)

2023 р.

Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ Комп'ютерної інженерії та управління \_\_\_\_\_  
Кафедра \_\_\_\_\_ Автоматизації проектування обчислювальної техніки \_\_\_\_\_  
Рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_  
Спеціальність \_\_\_\_\_ 123 \_\_\_\_\_ Комп'ютерна інженерія \_\_\_\_\_  
(шифр і назва)  
Тип програми \_\_\_\_\_ Освітньо-професійна \_\_\_\_\_  
(освітньо-професійна або освітньо-наукова)  
Освітня програма \_\_\_\_\_ Спеціалізовані комп'ютерні системи \_\_\_\_\_  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

«03» вересня 2023 р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові \_\_\_\_\_ Оленич Павлу Сергійовичу \_\_\_\_\_

(прізвище, ім'я, по батькові)

1. Тема роботи \_\_\_\_\_ Моделі впливу характеристик сайта на його позицію у видачі пошукових систем \_\_\_\_\_

затверджена наказом по університету від \_\_\_\_\_ 03.11.2023 р. № \_\_\_\_\_ 1288СТ \_\_\_\_\_

2. Термін подання студентом роботи до екзаменаційної комісії \_\_\_\_\_ 25.01.2024 р. \_\_\_\_\_

3. Вихідні дані до роботи \_\_\_\_\_  
\_\_\_\_\_ Мова розмітки гіпертексту HTML \_\_\_\_\_  
\_\_\_\_\_ Таблиці стилей CSS \_\_\_\_\_  
\_\_\_\_\_ Мова програмування JavaScript \_\_\_\_\_  
\_\_\_\_\_ Методи SEO-оптимізації \_\_\_\_\_

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_  
\_\_\_\_\_ Особливі характеристики пошукових систем \_\_\_\_\_  
\_\_\_\_\_ Засоби побудови семантично коректних сторінок \_\_\_\_\_  
\_\_\_\_\_ Фактори ранжування у видачі результатів пошуку \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 16 слайдів

---

---

---

---

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1 )

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата


7. Дата видачі завдання 02.09.2023

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	02.09.2023-06.09.2023	
2	Аналіз літератури за темою	06.09.2023-30.09.2023	
3	Суть технічної проблеми	30.09.2023-08.10.2023	
4	Постановка задачі	08.10.2023-12.10.2023	
5	Існуючі методи для вирішення задачі	12.10.2023-20.10.2023	
6	Обрання підходів SEO-оптимізації	20.10.2023-25.10.2023	
7	Створення експериментальних сайтів	25.10.2023-20.11.2023	
8	Порівняння результатів індексації	20.11.2023-30.11.2023	
9	Оформлення пояснювальної записки	30.11.2023-17.12.2023	
10	Оформлення графічної частини	17.12.2023-12.01.2024	

Студент 

(підпис)

Керівник роботи   
(підпис)

проф. Свір'я І. Б.  
(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи містить 66 сторінок, 10 рисунків, 4 таблиці, 10 формул, 8 джерел за переліком посилань.

### ІНТЕРНЕТ, SEO, ПОШУК ІНФОРМАЦІЇ, GOOGLE, ОПТИМІЗАЦІЯ САЙТУ, СЕМАНТИКА WEB СТОРІНКИ

Метою роботи є дослідження залежностей між характеристиками сайту та його позицією у видачі результатів пошуку пошуковими системами. Необхідно встановити вплив розміру сайту та якості його структури на швидкість та ефективність індексування.

Проведений аналіз існуючих засобів створення семантично вірних сторінок, що містять функції розширеного доступу, адаптування до пристрою користувача та обробку очікуваних помилок та проблем. Визначено чіткі залежності між якістю згаданих вище характеристик Web сторінки та її позицією у пошукових системах. Досліджені способи підвищення кількості показів сайту у видачі із застосуванням можливостей мови розмітки гіпертексту HTML та правильної організації структурних елементів. Запропоновано методи оптимізації сайту для максимально ефективного індексування за мінімальної витрати ресурсів. Визначена залежність між складністю ключового слова та кількістю позицій на яку підіймається сайт за тих самих застосованих методів оптимізації.

## ABSTRACT

The explanatory note to the qualification work contains 67 pages, 10 figures, 4 tables, 10 formulas, 8 sources according to the list of references.

INTERNET, SEO, SEARCH DATA, GOOGLE, WEB SITE OPTIMIZATION, WEB PAGE SEMANTIC

The purpose of the work is to study the relationship between the site's characteristics and its position in search engine results. It is necessary to establish the influence of the size of the site and the quality of its structure on the speed and efficiency of indexing.

An analysis of existing means of creating semantically correct pages, containing the functions of extended access, adaptation to the user's device and processing of expected errors and problems, was carried out. Clear dependencies between the quality of the above-mentioned characteristics of the Web page and its position in search engines have been determined. Researched ways to increase the number of site views in output using the capabilities of the HTML hypertext markup language and the correct organization of structural elements. Methods of site optimization for maximum effective indexing with minimal resource consumption are proposed. The relationship between the complexity of the keyword and the number of positions the site climbs according to the same applied optimization methods is determined.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	7
ВСТУП.....	8
1 АНАЛІЗ ЗАГАЛЬНОЇ ІНФОРМАЦІЇ ПРО ПОШУКОВІ СИСТЕМИ.....	10
1.1 Історія розвитку пошукових систем.....	10
1.2 Загальні принципи роботи пошукових систем.....	11
1.3 Сутність пошукових машин.....	13
1.4 Глобальні пошукові системи.....	15
1.4.1 Пошукова система Google.....	15
1.4.2 Пошукова система Yahoo.....	18
1.5 Постановка задачі.....	20
2 СТРАТЕГІЇ І МЕТОДИКА ПРОФЕСІЙНОГО ІНФОРМАЦІЙНОГО ПОШУКУ.....	21
2.1 Помилки просування сайта. Обман пошукових машин.....	24
2.2 Текстове оформлення Web-сторінок.....	28
2.3 Структура сайта.....	31
3 ЗОВНІШНІ ФАКТОРИ РАНЖУВАННЯ.....	32
3.1 Облік зовнішніх посилань на сайт.....	32
3.2 Важливість посилань (індекс цитування).....	32
3.3 Google PageRank.....	34
3.4 Підбір ключових слів.....	37
4 ХАРАКТЕРИСТИКИ ПОШУКОВИХ СИСТЕМ.....	40
4.1 Google SandBox.....	40
4.2 Google LocalRank.....	42
4.3 Підготовка до створення ефективного сайту.....	45
5 КОНКУРЕНТОСПРОМОЖНІСТЬ КЛЮЧОВИХ СЛІВ.....	48
5.1 Статистичний аналіз параметрів сайту.....	48
5.2 Вплив параметрів сайту на його позицію у видачі.....	52
ВИСНОВКИ.....	64
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	66
ДОДАТОК А.....	67

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

SEO (Search engine optimization) – оптимізація у пошукових системах

GOOGLE – пошукова система

GSC – Google Search Console

Web server – веб-сервер

QoS (Quality Of Service) – якість обслуговування

RSVP (Resource Reservation Protocol) – протокол резервування ресурсів

TE (Traffic Engineering) – перерозподіл потоків

ТІЦ – тематичний індекс цитування

ІЦ – індекс цитування

ЗІЦ – зважений індекс цитування

ВЧ – високочастотний запит

## ВСТУП

Як заявити про свій сайт? Можна вказувати його адресу у всіх видах реклами, можна надрукувати на візитівках, можна вставити його в підпис на форумах. Є ще безліч способів проінформувати про свій ресурс потенційних клієнтів, але всі вони навіть разом узяті не дадуть того ефекту, що здатна забезпечити та просування сайту в пошукових системах. Саме провідні пошукові системи – Google, Bing, Yahoo та деякі інші дають левову частку трафіку. Саме вони наводять на сайт найвмотивованіших, зацікавлених у співпраці відвідувачів. Тому одне з найголовніших завдань розкручування сайту – це його оптимізація для пошукових систем.

Що дає оптимізація сайту?

Основна мета, яку переслідує оптимізація сайту – це підготовка його вмісту для якнайшвидшої та детальнішої індексації пошуковими системами. Справа в тому, що пошуковим системам аж ніяк не байдуже, як влаштовані сайти, які вони показують у результатах своєї видачі. Пошукові роботи – спеціальні програми, які досліджують сайти та збирають інформацію про них, мають особливі вимоги до програмного коду ресурсу, структури сайту, текстів, розміщених на його сторінках. Приведення всіх цих складових у відповідність до вимог пошукових роботів, тобто оптимізація сайту, зробить його видимим для пошукових систем, які справно індексуватимуть якісний ресурс та підніматимуть його у своїх рейтингах. А це не що інше, як широко відчинені двері для цільових відвідувачів.

Що включає оптимізація сайту?

Як правило, оптимізація сайту – це одноразове завдання, успішно виконавши яке, надалі достатньо лише дотримуватись основних напрямів розвитку.

Якісна оптимізація сайту включає наступні етапи:

– всебічний аналіз ресурсу, який дозволить виявити слабкі місця та

розробити план оптимізації з урахуванням тематики сайту, конкурентності основних запитів, цілей розкручування;

- удосконалення HTML-коду;

- покращення структури сайту, наприклад, додавання необхідних розділів, створення карти сайту, впровадження коректної перелінковки сторінок тощо;

- написання грамотних оптимізованих текстів, наявність яких на сайті є найважливішою умовою успішного розкручування ресурсу.

Всі ці заходи можна поєднати одним коротким терміном – SEO (Search engine optimization), оптимізація у пошукових системах. Сьогодні правильне SEO є запорукою ефективності ресурсу.

Необхідно розкрити питання про те, для яких видів ключових слів оптимізація зіграє найбільшу роль? Ретельно вивчити зв'язок між складністю ключового слова та кількістю позицій, на яку сайт підніметься у видачі за однакових підходів та методів просування.

# 1 АНАЛІЗ ЗАГАЛЬНОЇ ІНФОРМАЦІЇ ПРО ПОШУКОВІ СИСТЕМИ

## 1.1 Історія розвитку пошукових систем

У початковий період розвитку Інтернету кількість його користувачів була невелика, а обсяг доступної інформації відносно невеликим. У більшості випадків доступ до Інтернету мали співробітники різних університетів та лабораторій, а загалом Мережа використовувалася в наукових цілях. У цей час задача пошуку інформації в мережі Інтернет була не так актуальною, як у даний час.

Одним із перших способів організації доступу до інформаційних ресурсів мережі стало створення каталогів сайтів, у яких посилання на ресурси групувалися згідно з тематикою. Першим таким проектом став сайт Yahoo, що відкрився у квітні 1994 року. Після того, як кількість сайтів у каталозі Yahoo значно збільшилася, було додано можливість пошуку інформації з каталогу. Це, звичайно ж, не було пошуковою системою в повному розумінні, оскільки область пошуку була обмежена лише ресурсами, присутніми в каталозі, а не всіма ресурсами Інтернету.

Каталоги посилань широко використовувалися раніше, але практично втратили свою популярність нині. Причина цього дуже проста – навіть сучасні каталоги, що містять величезну кількість ресурсів, подають інформацію лише про дуже малу частину мережі Інтернет. Найбільший каталог мережі DMOZ (або Open Directory Project) містить інформацію про 5 мільйонів ресурсів, у той час як база пошукової системи Google складається з понад 8 мільярдів документів.

Першою повноцінною пошуковою системою став проект WebCrawler, що з'явився в 1994 році.

У 1995 році з'явилися пошукові системи Lycos та AltaVista. Остання

довгі роки була лідером у сфері пошуку інформації в Інтернеті.

У 1997 році Сергій Брін та Ларрі Пейдж створили Google у рамках дослідницького проекту у Стенфордському університеті. Зараз Google найпопулярніша пошукова система у світі.

В даний час існує 3 основних міжнародних пошукових системи – Google, Yahoo і Bing, які мають власні бази та алгоритми пошуку. Більшість інших пошукових систем (яких можна нарахувати дуже багато) використовує в тому чи іншому вигляді результати трьох перерахованих. Наприклад, пошук AOL (search.aol.com) використовують базу Google, а AltaVista, Lycos та AllTheWeb – базу Yahoo.

## 1.2 Загальні принципи роботи пошукових систем

Пошукова система складається з таких основних компонентів:

Spider (павук) – браузероподібна програма, яка завантажує веб-сторінки;

Crawler (краулер, «подорожуючий» павук) – програма, яка автоматично проходить за всіма посиланнями, знайденими на сторінці;

Indexer (індексатор) – програма, яка аналізує веб-сторінки, завантажені павуками;

Database (база даних) – сховище скачаних та оброблених сторінок;

Search engine results engine (система видачі результатів) – отримує результати пошуку з бази даних;

Web server (веб-сервер) – веб-сервер, який здійснює взаємодію між користувачем та іншими компонентами пошукової системи.

Детальна реалізація пошукових механізмів може відрізнитися один від одного (наприклад, зв'язка Spider+Crawler+Indexer може бути виконана у вигляді єдиної програми, яка завантажує відомі веб-сторінки, аналізує їх та шукає за посиланнями нові ресурси), проте всім пошуковим системам притаманні описані загальні риси.

Павук (Spider) – це програма, яка завантажує веб-сторінки тим самим способом, як і браузер користувача. Відмінність полягає в тому, що браузер відображає інформацію, що міститься на сторінці (текстову, графічну і т.д.), павук не має жодних візуальних компонентів і працює безпосередньо з HTML-текстом сторінки (ви можете зробити «перегляд HTML-коду» в вашому браузері, щоб побачити «сирий» HTML-текст).

Crawler – виділяє всі посилання на сторінці. Його завдання – визначити, куди далі має йти павук, спираючись на посилання або виходячи із заздалегідь заданого списку адрес. Краулер, слідуючи за знайденими посиланнями, здійснює пошук нових документів, ще невідомих пошуковій системі.

Індексатор (Indexer) розбирає сторінку на складові та аналізує їх. Виділяються та аналізуються різні елементи сторінки, такі як текст, заголовки, структурні та стильові особливості, спеціальні службові html-теги тощо. Database. База даних – це сховище всіх даних, які пошукова система завантажує та аналізує. Іноді базу даних називають індексом пошукової системи.

Система видачі результатів (Search Engine Results Engine) займається ранжуванням сторінок. Вона вирішує, які сторінки задовольняють запит користувача, і в якому порядку вони повинні бути відсортовані. Це відбувається згідно з алгоритмами ранжування пошукової системи. Ця інформація є найціннішою та найцікавішою для нас – саме з цим компонентом пошукової системи взаємодіє оптимізатор, намагаючись покращити позиції сайту у видачі, тому надалі ми докладно розглянемо всі фактори, що впливають на ранжування результатів.

Як правило, на сервері присутня HTML-сторінка з полем введення, в якому користувач може задати пошуковий термін, що його цікавить. Веб-сервер (Web server) також відповідає за видачу результатів користувачеві як HTML-сторінки.

### 1.3 Сутність пошукових машин

Завдання пошукових машин – забезпечувати детальне розшук інформації в електронному всесвіті, що може бути досягнуто тільки за рахунок обліку (індексування) всього змісту максимально можливої кількості web-сторінок. На відміну від довідників, усі вони функціонують в автоматизованому режимі та мають однаковий принцип дії. Пошукові системи складаються із двох базових компонентів. Перший компонент є програмою – робот, завдання якого подорожувати з сервера на сервер, знаходити там нові або змінилися документи і завантажувати їх на головний комп'ютер системи. Робот, переглядаючи вміст документа, знаходить нові посилання, як на інші документи даного сервера, так і на зовнішні сайти. Програма самостійно направляє за вказаними посиланнями, знаходить нові документи та посилання в них, після чого процес повторюється знову, нагадуючи добре відомий у бібліографії “метод снігової грудки”.

Виявлені документи обробляються (індексуються) другим компонентом пошукової системи. При цьому, як правило, враховується весь зміст сторінки, включаючи текст, ілюстрації, аудіо та відео файли та ін. Індксації піддаються всі слова в документі, що дає можливість використовувати пошукові системи для детального пошуку за вузькою тематикою. Утворені величезні індексні файли, що зберігають інформацію про те, яке слово, скільки разів, в якому документі і на якому сервері використовується, і становлять базу даних, до якої відбувається звернення користувачів, що вводять у рядок запиту поєднання ключових слів.

Видача результатів здійснюється за допомогою спеціального модуля, який здійснює інтелектуальне ранжування результатів. При цьому береться до уваги місце розташування терміна в документі (назва, заголовок, основний текст), частота його повторення, відсоткове співвідношення шуканого терміна до іншого тексту сторінки, а також число та авторитетність зовнішніх посилань на цю сторінку з інших сайтів.

До основних параметрів пошукових систем належать:

- обсяг індексних файлів (число проіндексованих серверів та окремих документів);
- ступінь оперативності оновлення бази даних за рахунок включення відомостей про нові матеріали та видалення застарілих;
- можливості для складання запиту;
- інтелектуальність системи ранжування результатів пошуку;
- наявність додаткових сервісних функцій, що полегшують роботу користувача.

Перша величина, що є ключовою, встановлює широту охоплення матеріалу та визначається кількістю проіндексованих документів. Зараз ця цифра для лідерів світового мережевого пошуку коливається в межах від 1 до 3 з гаком мільярдів.

Враховуючи той факт, що в середньому інтернет-адреса зберігає актуальність до півроку, після чого документ або змінює місцезнаходження або забирається з сервера, велике значення має рівень оперативності оновлення даних, що характеризує ступінь відповідності індексного файлу пошукової системи реальному розташуванню документів на сайтах. В даний час цей параметр коливається від двох тижнів до півтора місяця.

Можливості пошукового механізму висловлювати запит максимально точно значною мірою визначають частку релевантних документів у переліку отриманих результатів. Кожна машина має власну лексику, яка по-різному дозволяє деталізувати пошукове розпорядження.

Усі пошукові машини мають модуль ранжування результатів пошуку. Створення таких модулів – ціла сфера програмування, у якій конкурують найскладніші алгоритми, створені різними компаніями. Перелік факторів, що приймаються до уваги при визначенні місця документа в переліку посилань, надзвичайно широкий: від розташування слова на сторінці до рейтингу (авторитету) сторінок, які мають посилання на знайдений документ.

Не останню роль грає і простота інтерфейсу, наявність додаткових

сервісних функцій, як-от, можливість перекладу тексту документа іноземною мовою, здатність виділяти всі документи з певного сайту, звуження критеріїв під час пошуку, знаходження документів «за зразком» і т.д.

За цими параметрами серед великої кількості пошукових систем виділяються кілька найбільш визнаних, що дозволяють виявляти інформацію з високим ступенем повноти та надійності. До найавторитетніших пошукових систем всесвітнього масштабу нині відносяться Google ([www.google.com](http://www.google.com)), AlltheWeb ([www.alltheweb.com](http://www.alltheweb.com)) та Alta Vista ([www.altavista.com](http://www.altavista.com)).

Практично всі всесвітньо відомі довідники та пошукові системи в даний час перетворилися на значні інформаційні корпорації з багатомільйонними доходами. Заробивши авторитет найбільш відвідуваних місць у Мережі, вони надають свої сторінки для розміщення рекламної інформації, доходи від якої складають основу їх бюджету. Поступово пошукові сервери перетворюються на багатофункціональні портали, в яких пошуковий сервіс залишається головною принадою для користувачів, але далеко не єдиною і навіть не основною послугою. Крім розшуку інформації, такі сервери зазвичай надають користувачам безкоштовну електронну пошту, можливість безкоштовно розміщувати власні сторінки, інформацію про погоду, поточні новини, біржові котирування, карти місцевості тощо.

## 1.4 Глобальні пошукові системи

### 1.4.1 Пошукова система Google

Пошукова система, запущена в 1998 році і є нині одноосібним лідером серед глобальних пошукових систем за всіма важливими параметрами. Головна перевага Google – обсяг його індексного файлу, який становить на сьогодні понад 3 мільярди web-сторінок та статей із груп новин за інтересами. За добу програми-роботи системи індексують близько трьох

мільйонів нових та оновлених сторінок, при тому, що актуалізація бази проводиться кожні 28 днів.

Друга безперечна перевага Google – його здатність індексувати документи не тільки у вигляді HTML-файлів, але також документи у форматах PDF, RTF, PS, DOC, XLS, PPT, WP5 та інших. При цьому Google дозволяє моментально конвертувати сторінки у зазначених форматах у звичайний HTML-файл, що звільняє користувача від необхідності мати спеціальне програмне забезпечення для доступу до файлу.

Наступною найважливішою перевагою є спеціально розроблений модуль ранжування результатів – PageRank. Він заснований на алгоритмі, згідно з яким спочатку встановлюється структура посилань по всій Мережі, а потім кожна окрема сторінка ранжується відповідно до числа та значущості посилань на неї з інших сторінок. При цьому авторитетність зовнішніх посилань важливіша, ніж їх кількість. Подібний алгоритм дозволяє суттєво підвищити релевантність посилань, внаслідок чого Google відрізняє високий рівень відповідності знайденої інформації інтересам користувача. Цей результат досягається, зокрема, ще й за рахунок спеціальної підсистеми захисту користувача від сайтів, що просуваються за допомогою різних недобросовісних методів.

Google відрізняється високим ступенем комфорту для користувача. Незважаючи на те, що це глобальна пошукова система, користувачі з неангломовних країн автоматично переадресовуються на інтерфейс їхньою рідною мовою. Російськомовний інтерфейс зокрема знаходиться за адресою [www.google.com](http://www.google.com). Тривалість процесу здебільшого вбирається у однієї секунди, попри величезний обсяг індексного файлу системи.

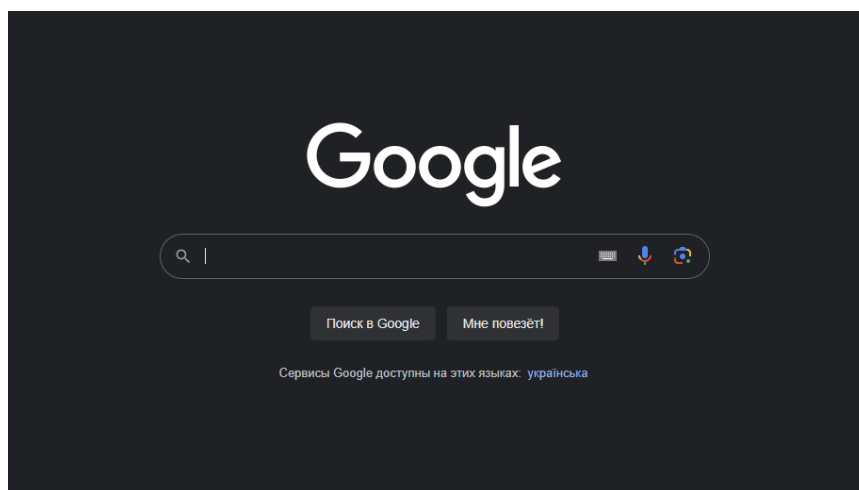


Рисунок 1.1 – Інтерфейс першої сторінки Google

Методика пошуку за допомогою Google дуже проста. У пошуковий рядок водиться запит природною мовою – неважливо російською, англійською чи будь-якою іншою. Мова запитів не допускає усічення термінів знаком «\*», тому всі можливі варіанти слів (library, libraries, librarians) користувачеві слід вводити самостійно. Усі терміни запиту за умовчанням об'єднуються умовою AND (I) – їх немає потреби ставити знак «+». Таким чином, до списку результатів потрапляють лише сторінки, які містять усі введені ключові слова. Для пошуку за точною фразою традиційно використовуються лапки: так запит "Хто вбив кішку у мадам Полосухер?" Прямо призведе до посилання на повний текст «Собачого серця» Михайла Булгакова.

Пошуковий механізм ігнорує стоп-слова (прийменники, спілки, артиклі), проте якщо якийсь із таких слів суттєво перед ним необхідно поставити «+», даючи зрозуміти системі, що в даному випадку термін навіть з однієї літери є значущим (наприклад: Олександр +I).

Google має у своєму арсеналі багато опцій для максимальної конкретизації запиту. Всі вони доступні через меню Advanced Search «Розширений пошук». Крім вже описаних можливостей додаються фільтри, що обмежують мову документа, його формат (наприклад, «тільки документи

у PDF»), час опублікування («останні три місяці»), місце терміна в самому документі («в заголовку сторінки») або розташування сторінки у певному домені чи навіть сайті.

Посилання на документи, що видаються в результаті пошуку, крім традиційних відомостей про назву документа, контекст шуканих слів і даних про розмір містять функцію Cached «Збережено», що дозволяє повністю відновити весь документ, якщо з будь-яких причин сайт на якому він розташований недоступний. Ще одна функція Similar pages «Схожі сторінки» дозволяє отримати перелік сторінок, зміст яких схожий із зазначеним джерелом. Ця функція, втім, поки що виконана без особливого успіху.

Крім пошуку текстових матеріалів Google має найкращі на сьогодні можливості пошуку ілюстрацій за допомогою режиму «Пошук зображень» (Images, «Зображення»). У його базі даних враховано понад 390 тисяч ілюстрацій, розшук яких ведеться аналогічно до пошуку текстових документів з можливістю обмеження певним розміром, форматом або кольоровістю графічних файлів – все через опцію «Розширений пошук зображень».

При використанні Google слід, однак, пам'ятати, що при роботі з файлами великого обсягу він індексує не все джерело, а лише перші 101 Кб. (для PDF-файлів ліміт обмежений 120 Кб.) тому індексація документа поки не завжди гарантує можливість його знаходження за будь-яким фрагментом тексту.

Як власний довідник ресурсів Інтернет Google використовує вдосконалений масив Open Directory Project, що часом дозволяє поєднувати переваги обох пошукових інструментів.

#### 1.4.2 Пошукова система Yahoo

Пошукова система, яка існує з 1997 року, розташована в Європі (Норвегія) і спочатку була орієнтована переважно на європейські сайти. Нині

відображає понад 2.1 мільярд документів, серед яких вагоме місце займають російськомовні матеріали. Оновлення бази проводиться раз на два тижні, середній час індексації – 5 тижнів. Yahoo здатний індексувати PDF-файли, які обробляються без обмеження їхнього розміру.

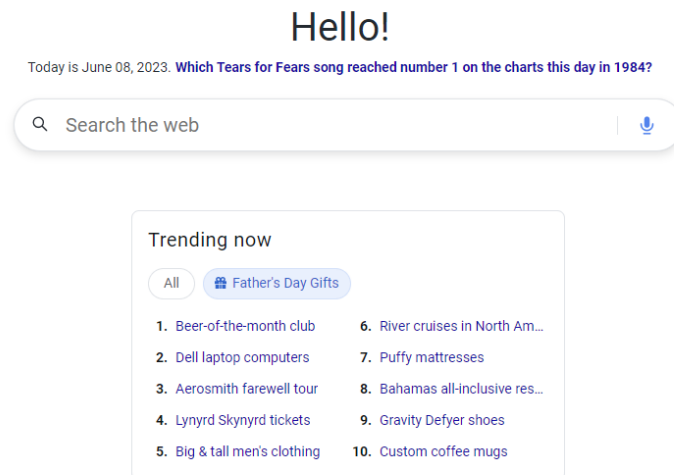


Рисунок 1.2 – Web інтерфейс головної сторінки

Система оперує мовою запитів, що включає знаки «+», «-», лапки для пошуку цитат, круглі дужки для пошуку за умовою АБО. Більш складні обмеження задаються специфічними операторами системи. Однак для формування багаторівневих запитів краще звернутися до Advanced Search («Розширеного пошуку»), який дозволяє легко застосувати фільтри, серед яких обмеження за місцезнаходженням терміна в документі, певному домені або сайті, часі опублікування, мові і навіть обсягу вихідної сторінки.

Серед сервісних функцій Yahoo є можливість автоматично об'єднувати результати в «тематичні» папки. Спеціальна функція дозволяє кожному користувачеві створити свій власний інтерфейс системи із зазначенням пріоритетів при сортуванні результатів та додатковими послугами, як, наприклад, відправлення знайдених посилань електронною поштою.

## 1.5 Постановка задачі

Об'єкт дослідження у роботі – просування сайту та покращення його позиції у видачі пошукових систем.

Предмет дослідження – методи SEO-оптимізації сайтів; алгоритми, що використовуються пошуковими машинами для ранжування сайтів.

Мета роботи – визначення впливу параметрів сайту на його позицію у видачі пошукових систем, визначення залежності між складністю ключового слова та ефективністю певних методів просування, надати рекомендації щодо просування сайтів для ключових слів різної складності.

Задачі, які необхідно вирішити для поставленої мети:

- аналіз алгоритмів ранжування сайтів пошуковими системами;
- аналіз методів та стратегій просування сайтів;
- створення експериментальних сайтів та дослідження їх позицій.

Основні вимоги до сайтів, що використовуватимуться у експерименті:

- зразки мають мати однакову тематику та зовнішній вигляд для користувача, дослідження має проводитись лише для параметрів, що мають значення тільки для пошукових систем;
- для коректності досліджень ІЦ та кількість ключових слів у текстовому контенті мають бути однаковими;
- зразок із вищою якістю просування має бути створений із використання семантично вірних тегів, мати додаткову мета-інформацію у заголовках сторінок та додаткових файлах (robots.txt, sitemap.xml);
- другий зразок має містити лише необхідний мінімум оптимізації для того, щоб було можливо оцінити його позицію у видачі;
- обидвом зразкам достатньо лише бути проіндексованими та потрапити у видачу для обраних ключових слів, реальний функціонал для можливих користувачів не передбачений.

## 2 СТРАТЕГІЇ І МЕТОДИКА ПРОФЕСІЙНОГО ІНФОРМАЦІЙНОГО ПОШУКУ

Починаючи інформаційний пошук в Інтернеті, слід завжди пам'ятати кілька основних моментів. Насамперед, жодні засоби навігації – довідники чи пошукові машини не охоплюють всього поточного інформаційного масиву Інтернет. За деякими оцінками, навіть такі визнані лідери мережевого пошуку як Google або AlltheWeb відображають не більше третини сукупного змісту Мережі. Причина цього – постійний колосальний приріст обсягів інформації в Інтернеті, який, незважаючи на всі зусилля навігаційних служб, містить величезну кількість білих плям.

Крім швидкого зростання та зміни розташування документів, більшість пошукових систем мають внутрішні обмеження на відображення матеріалів одного сайту та на обсяг індексованої частини сторінки. Програми-роботи часто не йдуть углиб сервера далі певної директорії, що також скорочує кількість відображених матеріалів.

У той же час деякі сервери мають власну систему пошуку, яка відображає весь їхній інформаційний масив. Виявивши такі сервери за допомогою довідників, можна провести більш детальне їхнє обстеження, використовуючи локальний пошуковий механізм. Наприклад, при пошуку відомостей про конкретний вид креветки, що штучно розводиться людиною, дуже раціональним буде знайти і переглянути сервери, присвячені в цілому аквакультури, галузі, що займається вирощуванням морепродуктів у штучних теплих водоймах, а при виявленні даних про конкретне захворювання – сервера установ, що ведуть дослідження у цій галузі.

Таким чином, для досягнення найбільш повних результатів слід застосовувати довідники та пошукові системи у поєднанні один з одним.

Існує також низка загальних вимог до пошукової діяльності,

дотримання яких підвищує ефективність та економить час, що витрачається на розшук даних. Розглянемо деякі з них.

1. Для пошуку матеріалів за вкрай вузькою специфічною тематикою варто починати з мета-машин, щоб одразу отримати уявлення про те, наскільки багато дана проблематика представлена в Інтернет.

2. Для отримання більш повних результатів за складним запитом (наприклад, там, де є обмеження не лише за змістом документа, але й за датою оновлення або розташування документів) пошук рекомендується проводити окремо в кожній пошуковій машині. Пошукові системи мають сильне рознесення у відображенні документів, та їх послідовне використання значною мірою розширює охоплення матеріалу.

3. При розшуванні документів про окрему країну або конкретною мовою слід віддати перевагу національним/регіональним пошуковим засобам. Так, наприклад, при пошуку матеріалів іспанською мовою, варто звертатися не до глобальних, а до іспанських пошукових систем, наприклад Trovator (<http://www.trovator.com>).

4. Формувати запит треба максимально точно, використовуючи всі можливості механізму складання запиту. Витрати часу на детальне складання пошукового розпорядження окупаються під час аналізу результатів пошуку. При точному формуванні запиту відсоток інформаційного шуму буде набагато нижчим.

При складанні запиту на текстовий матеріал завжди варто подумки уявити, яким може бути передбачуваний зміст документа. Під час пошуку джерел про Т.Г. Шевченка недостатньо просто запровадити його прізвище (список результатів буде рясніти адресами різних установ, розташованих на вулицях Шевченко різних містах). Набагато більший ефект принесе пошук на прізвище разом із назвами творів поета. Для розшуку ж повних текстів творів, навпаки, варто вводити в пошукові рядки з твору, які не є розхожими фразами (наприклад, для пошуку повного тексту поеми «Катерина» – не варто використовувати «Мій дядько найчесніших правил», набагато кращий

результат принесе «Коли на блідому небосхилі зірок зникає хоровод») або інша характерна, але не така поширена точна фраза.

При постійному занятті пошуковою діяльністю необхідно стежити за новинами, які стосуються пошукового сервісу. У навколишньому світі взагалі, а в цифровому світі тим більше немає нічого вічного. Кожна пошукова система переживає періоди зародження, розквіту та занепаду. Не бійтеся відмовитися від використання улюбленої, але застарілої пошукової системи та перейти на використання нових пошукових інструментів, які мають більшу ефективність.

Причина складнощів, що виникають при інформаційному пошуку в Інтернеті, визначається двома головними факторами. По-перше, кількість джерел у Мережі надзвичайно велика. Наприкінці 2001 року приблизні підрахунки вказували орієнтовну цифру в 7,5 мільярдів документів, розміщених на серверах по всьому світу. По-друге, масив інформації в Мережі не тільки колосальний за обсягом, а й вкрай динамічний. За ті півхвилини, що ви витратили на читання перших рядків цього розділу, у віртуальному всесвіті з'явилося близько сотні нових або змінених документів, десятки були переміщені на нові адреси, а одиниці назавжди припинили своє існування.

На відміну від стабільного та контрольованого фонду документів у бібліотеці, в Мережі ми маємо справу з гігантським та безперервно змінним інформаційним масивом, пошук даних у якому є вельми складним процесом.

Навички інформаційних розвідок, в тій чи іншій мірі, мають більшість користувачів глобальних комп'ютерних мереж. І дилетанти і професіонали часто користуються одними й тими самими інструментами. Проте результати розвідок і витрачений ними час різняться дуже значною мірою.

Пошукові системи (search engines) поширені у набагато більшій кількості, ніж електронні довідники та кількість їх, що становить сьогодні кілька десятків, продовжує неухильно збільшуватися. Професійна робота з ними вимагає спеціальних навичок, оскільки просте введення шуканого

терміна в пошуковий рядок швидше за все призведе до отримання списку з сотень тисяч документів, що містять дане поняття, що практично рівнозначне нульовому результату.

## 2.1 Помилки просування сайту. Обман пошукових машин

Помилки просування сайтів – дуже поширене явище. Помилки просування роблять як початківці вебмайстра, так і досвідчені через деякі обставини. Новачки роблять помилки з незнання, досвідчені з недогляду, наголошуючи на одних методах просування і тим самим, упускаючи інші. Що мається на увазі під помилками просування сайту – це те, що краще взагалі не робити в процесі просування сайту, або не варто робити, так як ефект від цього буде не передбачуваний.

Розглянемо основні помилки просування, які краще не допускати відразу, інакше вони можуть призвести в гіршому випадку до зниження рейтингу ресурсу або викиду його з бази пошукових систем, у кращому випадку сайт не просунеться до перших позицій пошукових систем, застрягши на місці.

Помилка просування перша – обман пошукових машин. Найпоширеніша помилка просування – це обман пошукових машин, яку допускають вебмайстри з незнання, начитавшись різного роду чтива: «Як швидко потрапити на перші позиції пошукових систем», або, свідомо сподіваючись на «може» пронесе. Обдурити пошуковик не просто, і навіть якщо такий трюк пройде, обман буде розкритий, і Ви отримаєте бан.

Якби пошукові машини жили самі по собі, то такий варіант розкрутки сайту як, обман пошукових машин, був би дуже доречним. Але вся проблема в тому, що пошукові роботи працюють за алгоритмами, які зробили люди, отже, вони знають всі способи обману пошукових машин. Не думаєте ж, Ви, що ті люди, які створили алгоритм роботи пошукових роботів, не знають, які способи обману існують. Все це постійно відстежується, і власники

пошукових систем постійно змінюють і оновлюють алгоритми їх роботи, однією з основних цілей яких є виявлення ошуканців, спамерів і несумлінних власників сайтів. Розглянемо докладніше, які методи обману пошукових машин використовують вебмайстра, виконуючи процес просування веб-сайту.

1. Використання на сторінках сайту великої кількості ключових слів. Є однією з найпоширеніших помилок просування, скоріше навіть оптимізації сайту. Суть її в тому, що при складанні контенту сайту, вебмайстри дуже захоплюються, і в текстовому вмісті сторінок використовують багато ключових слів і фраз, що є грубою помилкою просування. Оптимальною кількістю ключових слів буде 3-7% загальної кількості слів на сторінці. Чим загрожує перебір ключових слів? Тим, що пошукові системи можуть сприйняти це як спам і викинути сайт зі своєї бази, переставши індексувати його сторінки.

2. Написання великої кількості ключових слів під колір фону сторінки або дуже дрібним шрифтом, який не бачать відвідувачі, але добре читається пошуковими роботами, так само буде грубою помилкою просування. Застосовується для швидкого підняття сайту у списках видачі пошукових систем за запитами. Не, що інше, як обман пошукових машин, який припиняється суворо і сайт буде видалено з їхньої бази назавжди, без можливості відновлення.

3. Застосування ключових слів, що не мають відношення до тематики змісту сторінок, наприклад, які використовують ресурси з високою відвідуваністю (розважальні та еротичного змісту). Також є помилкою просування та оптимізації, спробою обману пошуковика. Пошукові роботи перевіряють текст і порівнюють його за релевантністю з ключовикам, якщо вони виявлять, що ключові слова не мають відношення до вмісту сторінки, то припинять її індексувати. Ключові слова повинні бути релевантними вмісту сайту і нести в собі короткий зміст тематичного вмісту сторінки сайту, а не виконувати функцію обману пошукових систем.

4. Дорвеї чи переадресатори. Явна спроба обману пошукача. Дорвей – це сторінка, яка не має свого логічного вмісту, просто напхана ключовими словами і при спробі зайти на неї відбувається перекидання користувача на сайт, для розкручування якого вона призначена. До теми «помилки просування», використання дорвеїв підходить дуже доречно. Пошукові роботи щодня відстежують і видаляють зі своїх баз сотні дорвеїв. Якщо така спроба обману пошуковика буде помічена, сайту також відмовлять в індексації.

Помилка просування друга – обмін посиланнями. Помилки просування сайту виникають в основному через відсутність досвіду та застосування дій з оптимізації з чуток. Обмін посиланнями методом ти мені, я тобі, не зовсім буде корисним для сайту, тим більше, якщо ресурс іншої тематики, що посилається на Вас. Не влаштовуйте на сайті лінкопомийку. Куди буде корисніше, якщо на Ваш сайт посилатимуться ресурси схожої тематики та високими показниками ТІЦ та PR, без встановлення зворотного посилання. Іноді трапляються сайти, які реєструються в сотнях безкоштовних каталогів, які вимагають встановлення зворотного посилання. Їхні сторінки буквально, засипані кнопками посилань таких каталогів. Вихід один намагатися потрапити в білі каталоги, вони не вимагають встановлення зворотного посилання.

Помилка просування третя – використання контекстної реклами для підвищення позицій сайту. Деякі вебмайстри вважають, що застосовуючи такий вид пошукового просування сайту, вони потраплять у «топ» 10 пошукової системи. Може вони десь мають рацію, і вийдуть на першу сторінку на запит, але ефект від такого просування короткочасний і не розрахований на довгострокове перебування на лідируючих позиціях пошукових систем, а це і є помилка просування. Контекстну рекламу як таку краще застосовувати для залучення на сайт покупця або замовника послуг, але не використовувати її для просування сайту, тільки витратите гроші. Контекстна реклама у принципі спрямовано залучення споживачів певного

товару, а чи не на стабільне відвідування ресурсів.

Помилка просування четверта – використання систем активного розкручування для просування сайту. Взагалі вступати до САР та залучати відвідувачів на сайт, таким чином, не забороняється, і шкоди для репутації сайту цей спосіб не завдасть. Його також можна віднести до розділу помилки просування. Дає ефект короточасного підняття сайту в рейтингах та тимчасового збільшення відвідуваності сайту. Але якщо у випадку з контекстною рекламою до Вас приходять цільові та унікальні відвідувачі, яких цікавить інформація на Вашому сайті, то із систем активної реклами заходять користувачі, яким Ваш сайт може бути зовсім не цікавий. Як тільки Ви припините користуватися послугами активної реклами, рейтинг сайту піде вниз. Свого роду це теж обман, і не чесне підвищення рейтингу по відношенню до інших учасників.

Якщо після прочитання цієї статті Ви зрозумієте, що помилки просування сайту були вже допущені, то постарайтеся їх виправити поки що не надто пізно, якщо пошукові системи Вас вже не забанили і не викинули зі своїх баз. І не допускайте помилок надалі, виконуючи просування сайту. Усуваючи помилки просування, застосовуйте антидії до способів обману пошукових машин, викладених раніше.

Алгоритми роботи пошукових машин цілком і повністю спрямовані на надання інформації користувачу, яка релевантна запиту, зробленому ним через форму пошуку. І припускаючись тих чи інших помилок просування, Ви ризикуєте бути відловленим пошуковими роботами і бути викинутими з їх обителі. Тому не думайте про пошукове просування, як про припасування сайту під пошукові системи, а робіть сайт та його вміст для користувачів, тоді і пошукові системи полюблять Вас і самі плавно піднімуть Ваш сайт на верхні позиції видачі. Врахуйте, швидко високо злетівши, за допомогою обману доведеться високо падати.

## 2.2 Текстове оформлення Web-сторінок

Пошуковики цінують сайти, багаті на інформаційний зміст. Загалом слід прагнути до збільшення текстового наповнення сайту.

Оптимальними слід вважати сторінки, що містять 500-3000 слів або 2-20 кВ. тексту (від 2 до 20 тис. символів).

Сторінка, що складається лише з кількох пропозицій, має менше шансів потрапити до «топу» пошукових систем.

Крім того, більше тексту на сторінці збільшує видимість сторінки в пошукових системах за рахунок рідкісних або випадкових пошукових фраз, що в деяких випадках може давати непоганий приплив відвідувачів.

Ключові слова (фрази) мають зустрічатися в тексті щонайменше 3-4 рази. Верхня межа залежить від загального обсягу сторінки – що більше загальний обсяг, то більше повторень можна зробити.

Окремо слід розглянути ситуацію з пошуковими фразами, тобто словосполученнями з кількох ключових слів. Найкращі результати спостерігаються, якщо фраза зустрічається в тексті кілька разів саме як фраза (тобто всі слова разом у потрібному порядку), а крім того, слова фрази трапляються в тексті кілька разів поодиночі. Також має існувати певна відмінність (розбалансування) між числом входжень кожного зі слів, що становлять фразу.

Розглянемо ситуацію з прикладу. Припустимо, ми оптимізуємо сторінку під фразу DVD програвач. Хороший варіант – фраза «dvd програвач» зустрічається в тексті 10 разів, крім того, слово «dvd» зустрічається окремо ще 7 разів, слово «програвач» ще 5 разів. Усі цифри на прикладі умовні, але добре показують загальну ідею.

Щільність ключового слова на сторінці вказує на відносну частоту змісту слова в тексті. Щільність вимірюється у відсотках. Наприклад, якщо задане слово зустрілося 5 разів на сторінці зі 100 слів, то густина цього слова дорівнює 5%. Надто низька щільність призведе до того, що пошукова

система не надасть належного значення цьому слову. Занадто висока щільність здатна включити спам-фільтр пошукової системи (тобто сторінку буде штучно знижено в результатах пошуку через надмірно частого вживання ключової фрази).

Оптимальною вважається густина ключового тексту 5-7%. У разі фраз, які з кількох слів, слід порахувати сумарну щільність всіх ключових слів, складових фразу і переконатися, що вона вкладається у зазначені межі.

Практика показує, що щільність ключового тексту більше 7-8%, хоч і не веде до якихось негативних наслідків, але й особливого сенсу в більшості випадків також не має.

Пошукові системи надають особливого значення тексту, тим чи іншим способом, виділеним на сторінці. Можна розглянути такі випадки

1. Використовуйте ключові слова в заголовках (текст, виділений тегами «H», особливо «h1» і «h2»). В даний час використання css дозволяє перевизначити вид тексту, виділеного цими тегами, тому використання тегів «H» має менше значення, ніж раніше, проте нехтувати ними ні в якому разі не варто.

2. Виділяйте ключові слова жирним шрифтом (не у всьому тексті, звичайно ж, але зробити таке виділення 2-3 рази на сторінці не завадить). Для цього рекомендується використовувати тег «strong» замість більш традиційного тега «B» (bold).

Один з найважливіших тегів, якому пошукові системи надають величезного значення. Обов'язково потрібно використовувати ключові слова у тезі TITLE.

Крім того, посилання на ваш сайт у видачі пошукової системи буде містити текст з тега TITLE, так що це певною мірою візитна картка сторінки.

Саме за цим посиланням здійснюється перехід відвідувача пошукової системи на ваш сайт, тому тег TITLE має не тільки містити ключові слова, але бути інформативним та привабливим.

Як правило, у видачу пошукової системи потрапляє 50-80 символів із

тега TITLE, тому розмір заголовка бажано обмежити цією довжиною.

Також дуже просте правило – використовуйте ключові слова в тексті вихідних посилань з ваших сторінок (як на інші внутрішні сторінки вашого сайту, так і інші ресурси мережі), це може додати вам невелику перевагу при ранжируванні.

Будь-яке зображення на сторінці має спеціальний атрибут альтернативний текст, який задається в тезі ALT. Цей текст буде відображено на екрані, якщо завантажити зображення не вдалося або показ зображень заблокований у браузері.

Пошукові системи запам'ятовують значення тега ALT при розборі (індексації) сторінки, проте не використовують його під час ранжування результатів пошуку.

На даний момент достовірно відомо, що пошукова система Google враховує текст у тезі ALT тих зображень, які є посиланнями на інші сторінки, решта тегів ALT ігнорується. За іншими пошуковими системами точних даних немає, але можна припускати щось подібне.

Загалом варто дати таку пораду – використовувати ключові слова у тегах ALT можна і потрібно, хоча принципового значення це не має.

Мета-тег Description призначений для завдання опису сторінки. Цей тег ніяк не впливає на ранжування, проте дуже важливий. Багато пошукових систем (і, зокрема, найбільша Google) відображають інформацію з цього тега в результатах пошуку, якщо цей тег присутній на сторінці та його вміст відповідає вмісту сторінки та пошуковому запиту.

Можна з упевненістю сказати, що високе місце у результатах пошуку не завжди забезпечує велику кількість відвідувачів. Якщо опис ваших конкурентів у результатах видачі буде більш привабливим, ніж вашого сайту, відвідувачі пошукової системи оберуть саме їх, а не ваш ресурс.

Тому грамотне складання мета-тегу Description має велике значення. Опис має бути коротким, але інформативним та привабливим, містити ключові слова, характерні для цієї сторінки.

Мета-тег Keywords – тег спочатку призначався для вказівки ключових слів цієї сторінки. Проте нині майже не використовується пошуковими системами.

Проте, варто заповнити цей тег «про всяк випадок». При заповненні слід дотримуватись наступного правила: додавати лише ті ключові слова, які реально присутні на сторінці.

### 2.3 Структура сайту

Загальне правило – що більше, то краще. Збільшення кількості сторінок сайту покращує його видимість у пошукових системах.

Крім того, поступове додавання нових інформаційних матеріалів до сайту сприймається пошуковими системами як розвиток сайту, що може дати додаткові переваги при ранжуванні.

Таким чином, намагайтеся розміщувати на сайті більше інформації – новини, прес-релізи, статті, корисні поради тощо.

Як правило, будь-який сайт має меню навігації. Використовуйте ключові слова у посиланнях меню, це дозволить надати додаткову вагу тим сторінкам, на які веде посилання.

Існує думка, що використання ключових слів у назві HTML-файлу сторінки може позитивно позначитися на її місці в результатах пошуку. Звісно, це стосується лише англомовних запитів.

Намагайтеся оптимізувати кожну сторінку під власну ключову фразу. Іноді можна вибрати 2-3 споріднені фрази, але не слід оптимізувати одну сторінку під 5-10 фраз відразу, швидше за все результату не буде ніякого.

## 3 ЗОВНІШНІ ФАКТОРИ РАНЖУВАННЯ

### 3.1 Облік зовнішніх посилань на сайт

Як можна побачити з попереднього розділу, майже всі фактори, що впливають на ранжування, знаходяться під контролем автора сторінки. Таким чином, для пошукової системи стає неможливим відрізнити дійсно якісний документ, від сторінки, створеної спеціально під задану пошукову фразу або навіть сторінки, згенерованої роботом і корисної інформації, що взагалі не несе. Тому одним із ключових факторів при ранжуванні сторінок є аналіз зовнішніх посилань на кожну сторінку, що оцінюється. Це єдиний фактор, який не підконтрольний автору сайту.

Логічно припустити, що чим більше зовнішніх посилань є на сайт, тим більший інтерес цей сайт для відвідувачів. Якщо власники інших сайтів у мережі поставили посилання на ресурс, що оцінюється, значить, вони вважають цей ресурс досить якісним. Дотримуючись цього критерію, пошукова система також може вирішити, яку вагу надати тому чи іншому документу.

Таким чином, існує два основних фактори, за якими сторінки, наявні в базі пошуковика, будуть відсортовані під час видачі. Це релевантність (тобто, наскільки розглянута сторінка пов'язані з темою запиту – чинники, описані у попередньому розділі) і кількість і якість зовнішніх посилань. Останній фактор також отримав назви посилання цитування, популярність посилання або індекс цитування.

### 3.2 Важливість посилань (індекс цитування)

Легко побачити, що простий підрахунок числа зовнішніх посилань не дає достатньої інформації для оцінки сайту. Очевидно, що посилання з сайту

www.microsoft.com має означати набагато більше, ніж посилання з домашньої сторінки [www.hostingcompany.com/~myhomepage.html](http://www.hostingcompany.com/~myhomepage.html), тому порівнювати популярність сайтів лише за кількістю зовнішніх посилань не можна – необхідно враховувати також і важливість посилань.

Для оцінки кількості та якості зовнішніх посилань на сайт пошукові системи вводять поняття індексу цитування.

Індекс цитування або ІЦ – це загальне позначення чисельних показників, що оцінюють популярність того чи іншого ресурсу, тобто абсолютне значення важливості сторінки. Кожна пошукова система використовує свої алгоритми для підрахунку власного індексу цитування, зазвичай ці значення ніде не публікуються.

Крім звичайного індексу цитування, який є абсолютний показник (тобто деяке конкретне число), вводиться термін зважений індекс цитування, який є відносним значенням, тобто показує популярність цієї сторінки щодо популярності інших сторінок в Інтернет.

Детальний опис індексів цитування та алгоритмів їх підрахунку буде наведено в наступних розділах.

Величезне значення під час ранжування результатів пошуку надається тексту зовнішніх посилань на сайт. Текст посилання (або інакше якірний чи посилальний текст) – це текст, що стоїть між тегами «А» і «/А», тобто той текст, яким можна «клікнути» вказівником миші у браузері для переходу на нову сторінку.

Якщо текст посилання містить потрібні ключові слова, пошукова система сприймає це як додаткову і дуже важливу рекомендацію, підтвердження того, що сайт дійсно містить цінну інформацію, що відповідає темі пошукового запиту.

### 3.4 Google PageRank

Першими, хто запатентував систему обліку зовнішніх посилань, стала

компанія Google. Алгоритм отримав назву PageRank.

PageRank розраховується для кожної веб-сторінки окремо, і визначається PageRank'ом (цитованістю) сторінок, що посилаються на неї. Свого роду замкнене коло.

Головне завдання полягає в тому, щоб знайти критерій, що виражає важливість сторінки. У випадку з PageRank таким критерієм було обрано теоретичну відвідуваність сторінки.

Розглянемо модель подорожі користувача через мережу шляхом переходу за посиланнями. Передбачається, що користувач починає переглядати сайти з певної випадково вибраної сторінки. Потім, за посиланнями, він переходить на інші ресурси. При цьому є ймовірність того, що відвідувач залишить сайт і знову почне перегляд документів з випадкової сторінки (в алгоритмі PageRank ймовірність такої дії прийнята на 0.15 на кожному кроці). Відповідно, з ймовірністю 0.85 він продовжить подорож, перейшовши однією з доступних на поточній сторінці посилань (усі посилання при цьому рівноправні). Продовжуючи подорож до нескінченності, він побуває на популярних сторінках багато разів, а маловідомих – менше.

Таким чином, PageRank веб-сторінки визначається як ймовірність знаходження користувача на даній веб-сторінці; сума ймовірностей по всіх веб-сторінках мережі дорівнює одиниці, тому що користувач обов'язково знаходиться на будь-якій сторінці.

Оскільки оперувати ймовірностями не завжди зручно, після ряду перетворень з PageRank можна працювати у вигляді конкретних чисел (як, наприклад, ми звикли бачити його в Google ToolBar, де кожна сторінка має PageRank від 0 до 10).

Згідно з описаною вище моделлю отримуємо, декілька тверджень.

1. Кожна сторінка в мережі (навіть якщо на неї немає зовнішніх посилань) спочатку має ненульовий PageRank (хоч і дуже маленький).
2. Кожна сторінка, яка має вихідні посилання, передає частину свого

PageRank сторінкам, на які посилається. При цьому, переданий PageRank обернено пропорційний числу посилань на сторінці – чим більше посилань, тим менший PageRank передається по кожній.

3. PageRank передається не повністю, на кожному кроці відбувається згасання (та сама ймовірність 15%, коли користувач починає перегляд з нової, випадково вибраної сторінки).

Розглянемо тепер, як PageRank може впливати на ранжування результатів пошуку (говоримо «може», оскільки у чистому вигляді PageRank вже давно не бере участі в алгоритмі Google, як це було раніше, але про це нижче). З впливом PageRank все дуже просто – після того як пошукова система знайшла ряд релевантних документів (використовуючи текстові критерії), відсортувати їх можна згідно PageRank – оскільки логічно буде припустити, що документ, що має більшу кількість якісних зовнішніх посилань, містить найбільш цінну інформацію.

Таким чином, алгоритм PageRank «витісняє» нагору у пошуку ті документи, які і без пошукової системи найбільш популярні.

Наразі PageRank не використовується безпосередньо в алгоритмі Google. Це і зрозуміло – адже PageRank характеризує лише кількість і якість зовнішніх посилань на сайт, але абсолютно не враховує посилальний текст і інформаційний вміст сторінок, що посилаються – а саме цим факторам доведеться максимальне значення при ранжуванні. Передбачається, що для ранжування Google використовує так званий тематичний PageRank (тобто той, що враховує лише посилання з тематично пов'язаних сторінок), проте деталі цього алгоритму відомі лише розробникам Google.

Дізнатися значення PageRank для будь-якої веб-сторінки можна за допомогою Google ToolBar, який показує значення PageRank в діапазоні від 0 до 10. Слід враховувати, що Google ToolBar показує не точне значення PageRank, а лише діапазон PageRank, в який потрапляє сайт, номер діапазону (від 0 до 10) визначається за логарифмічною шкалою.

Пояснимо на прикладі: кожна сторінка має точне значення PageRank,

відоме лише Google. Для визначення потрібного діапазону та виведення інформації на ToolBar використовується логарифмічна шкала (приклад показаний у таблиці)

Всі цифри умовні, проте наочно демонструють, що діапазони PageRank, що відображаються в Google ToolBar, не є рівнозначними один одному. Наприклад, підняти PageRank з 1 до 2 легко, а з 6 до 7 набагато складніше.

На практиці PageRank використовується в основному у двох цілях.

1. Швидка оцінка рівня розкрученості сайту. PageRank не дає точної інформації про сторінки, що посилаються, але дозволяє швидко і просто «прикинути» рівень розвитку сайту. Для англomовних сайтів можна дотримуватись наступної градації: PR 4-5 – найбільш типовий PR для більшості сайтів середньої розкрученості. PR 6 – дуже добре розкручений сайт. PR 7 – величина, практично недосяжна для звичайного вебмайстра, але іноді трапляється. PR 8, 9, 10 – зустрічаються лише у сайтів великих компаній (Microsoft, Google і т.п.). Знання PageRank можна використовувати під час обміну посиланнями, щоб оцінити якість запропонованої до обміну сторінки та інших подібних ситуаціях.

2. Оцінка рівня конкуренції за пошуковим запитом. Хоча PageRank і не використовується безпосередньо в алгоритмах ранжування, проте дозволяє опосередковано оцінити конкурентність заданого запиту. Наприклад, якщо у видачі пошукової системи стоять сайти з PageRank 6-7, сайт з PageRank 4 має дуже мало шансів піднятися в «топ».

Ще одне важливе зауваження – значення PageRank, що показуються в Google ToolBar, перераховуються досить рідко (раз на кілька місяців), тому ToolBar показує певною мірою застарілу інформацію. Тобто, сама пошукова система Google враховує зміни у зовнішніх посиланнях набагато швидше, ніж ці зміни відображаються в Google ToolBar.

### 3.5 Підбір ключових слів

При оптимізації сайту можна виділити дві стратегії – оптимізація під мале число високопопулярних ключових слів, або під велику кількість малопопулярних. Насправді зазвичай поєднуються обидві.

Недолік високочастотних запитів – зазвичай, високий рівень конкуренції із них. Для молодого сайту не завжди можливо піднятися в «топ» за цими запитами.

Для низькочастотних запитів, часто буває достатнім згадування потрібного словосполучення на сторінці, або мінімальної текстової оптимізації. За певних умов низькочастотні запити можуть давати непоганий пошуковий трафік.

Мета більшості комерційних сайтів – продати той чи інший товар чи послугу, або якимось іншим способом заробити на своїх відвідувачах. Це потрібно враховувати при пошуковій оптимізації та підборі ключових слів. Потрібно прагнути отримувати цільових відвідувачів на сайт (тобто готових до купівлі пропонованого товару чи послуги), ніж просто до великої кількості відвідувачів.

Приклад. Запит «монітор» набагато популярніший і в рази конкурентніший, ніж запит «монітор samsung 710N» (точна назва моделі). Однак для продавця моніторів другий відвідувач набагато цінніший, а отримати його набагато легше, оскільки рівень конкуренції за другим запитом невеликий. Це ще одна можлива різниця між високочастотними та низькочастотними запитами, яку слід враховувати.

Після того, як набір ключових слів приблизно відомий, слід визначити основне ядро слів, під які проводитиметься оптимізація.

Низькочастотні запити зі зрозумілих причин відкидаються одразу (тимчасово). У попередньому розділі ми описали користь від низькочастотних запитів, проте вони й низькочастотні, що не вимагають особливої оптимізації. Тому у цьому розділі ми їх не розглядаємо.

За дуже популярними фразами рівень конкуренції зазвичай дуже високий, тому потрібно реально оцінити можливості свого сайту. Для оцінки рівня конкуренції слід підрахувати низку показників для першої десятки сайтів у видачі пошуковика:

- середній PageRank сторінок у видачі;
- середній ТІЦ сайтів, чії сторінки потрапили у видачу;
- середня кількість зовнішніх посилань на сайти у видачі за версією різних пошукових систем.

Додаткові параметри:

- число сторінок в Інтернеті, що містять заданий пошуковий термін (тобто число результатів пошуку);
- число сторінок в Інтернеті, що містять точний збіг заданої фрази (як при пошуку в лапках).

Ці додаткові параметри допоможуть опосередковано оцінити складність виведення сайту в топ за заданою фразою.

Крім описаних параметрів, можна також перевірити скільки сайтів з видачі присутні в основних каталогах, таких як каталоги DMOZ, Yahoo.

Аналіз усіх зазначених вище параметрів та порівняння їх із параметрами власного сайту дозволять досить чітко передбачити перспективи виведення вашого сайту у «топ» за вказаною фразою.

Оцінивши рівень конкуренції за всіма підібраними фразами, ви зможете вибрати низку досить популярних фраз з прийнятним рівнем конкуренції, на які і буде зроблено основну ставку при розкручуванні та оптимізації.

Як говорилося вище, послуги пошукових систем часто дають дуже неточну інформацію. Тому визначити ідеальний на вашому сайті набір ключових слів з першого разу вдається досить рідко.

Після того, як ваш сайт зверстаний і зроблено певні кроки по його розкрутці, у ваших руках виявляється додаткова статистика за ключовими словами: ви знаєте рейтинг вашого сайту у видачі пошукових систем за тією

чи іншою фразою, і знаєте також кількість заходів на ваш сайт за цією фразі.

Володіючи цією інформацією, можна досить чітко визначити вдалі та невдалі фрази. Найчастіше навіть не потрібно чекати, щоб сайт вийшов у «топ» за фразами, що оцінюються у всіх пошукових системах - достатньо однієї або двох.

Приклад. Допустимо, ваш сайт зайняв перше місце в пошуковій системі Rambler за цією фразою. При цьому в Google його ще немає у видачі за цією фразою. Однак, знаючи відсоткове співвідношення заходів на ваш сайт із різних пошукових систем (наприклад, Yahoo – 70%, Google – 20%, Rambler – 10%), ви вже можете передбачити зразковий трафік за цією фразою та вирішити, підходить вона для вашого сайту або ні.

Крім виділення невдалих фраз ви можете знайти нові вдалі варіанти. Наприклад, побачити, що деяка фраза, під яку не робилося ніякої розкрутки, приносить непоганий трафік, навіть незважаючи на те, що ваш сайт за цією фразою знаходиться на 2 або 3 сторінці у видачі.

Таким чином, у ваших руках виявляється новий, уточнений набір ключових слів. Після цього слід розпочати перебудову сайту – зміну текстів під вдалі словосполучення, створення нових сторінок під нові знайдені фрази тощо. Таким чином, через деякий час ви зможете знайти найкращий набір ключових слів для вашого сайту та суттєво збільшити пошуковий трафік.

Ще деякі поради. За статистикою, на головну сторінку сайту припадає до 30% -50% всього пошукового трафіку. Вона найкраще видно у пошукових системах і має найбільше зовнішніх посилань. Тому головну сторінку сайту слід оптимізувати під найбільш популярні та конкурентні запити. Кожну сторінку сайту слід оптимізувати під 1-2 основні словосполучення (і можливо під ряд низькочастотних запитів). Це збільшить шанси виходу до «топу» пошукових систем за заданими фразами.

## 4 ХАРАКТЕРИСТИКИ ПОШУКОВИХ СИСТЕМ

### 4.1 Google SandBox

У початку 2004 року серед оптимізаторів з'явилася нова таємнича концепція – Google SandBox або пісочниця Google. Таке позначення отримав новий спам-фільтр Google, спрямований на виключення з видачі молодих, недавно створених сайтів.

Фільтр SandBox проявляється в тому, що новостворені сайти відсутні в видачі пошукової системи практично за всіма фразами. Це відбувається незважаючи на наявність якісного та унікального інформаційного вмісту і правильно проведений просування (без використання спам-методів).

На даний момент SandBox стосується лише англomовного сегменту, сайти на українській та інших мовах не піддаються цьому фільтру. Однак, цілком можливо, що цей фільтр може розширити свій вплив.

Можна припустити, що ціль фільтра SandBox полягає в виключенні з видачі спам-сайтів – насправді жоден пошуковий спамер не зможе чекати місяці до появи результатів. Однак разом з цим страждає велика кількість нормальних, новостворених сайтів.

Точної інформації про те, що ж конкретно представляє собою фільтр SandBox, досі немає. Існує ряд припущень, отриманих на підставі досвіду, які ми тут наведемо.

SandBox – це фільтр для молодих сайтів. Новостворений сайт потрапляє в «пісочницю» і перебуває в ній невизначений час, доки пошукова система не переведе його в розряд «звичайних».

SandBox – це фільтр на нові посилання, які ставляться на новостворені сайти. Спробуйте помітити принципову відмінність від попереднього припущення – фільтр накладається не на вік сайту, а на вік посилань на сайт.

Іншими словами, Google не має претензій до сайту, але відмовляється враховувати зовнішні посилання на нього, якщо з моменту їх появи минуло менше X місяців. Оскільки зовнішні посилання є одним з основних чинників ранжування, ігнорування зовнішніх посилань еквівалентно відсутності сайту в видачі пошукової системи. Яке з двох наведених припущень більш вірно сказати складно, цілком можливо, що обидва вони вірні.

Сайт може перебувати в пісочниці від 3-х місяців до року і більше. Є також спостереження, що сайти виходять з пісочниці масово. Тобто термін пісочниці визначається не індивідуально для кожного сайту, а для великих груп сайтів (сайти, створені в певний часовий діапазон, потрапляють в одну групу). Фільтр потім знімається одразу для всієї групи, таким чином, сайти з однієї групи пробудуть в «піску» різний час.

Типові ознаки того, що ваш сайт перебуває в пісочниці:

- ваш сайт нормально проіндексований Google, регулярно відвідується пошуковим роботом;
- ваш сайт має PageRank, пошукова система знає і коректно відображає зовнішні посилання на ваш сайт;
- пошук за адресою сайту ([www.site.com](http://www.site.com)) видає правильні результати, з правильним заголовком, сніпетом (описом ресурсу) і т.д.;
- ваш сайт нормально видається по рідкісних та унікальних словосполученнях, що містяться в тексті сторінки;
- ваш сайт не видно в першій тисячі результатів по будь-яким іншим запитам, навіть по тим, під які він спочатку створювався.

Іноді зустрічаються винятки, і сайт за деякими запитам з'являється на 500-600 позиціях, що, звісно, не змінює суті.

Методів обігу фільтра практично не існує. Є ряд припущень, як це можна зробити, але це не більше, ніж припущення, і до того ж малоприйнятні для звичайного вебмастера. Основний метод – це працювати над сайтом і чекати закінчення фільтра. Після того, як фільтр знімається, відбувається стрімкий ріст рейтингів на 400-500 і більше позицій.

## 4.2 Google LocalRank

25 лютого 2003 року компанія Google отримала патент на новий алгоритм ранжування сторінок, який отримав назву LocalRank. В основі цього алгоритму лежить ідея ранжування сторінок не за їх глобальною кількістю посилань, а за кількістю посилань в середовищі сторінок, які тематично пов'язані з запитом користувача.

Алгоритм LocalRank на практиці не використовується (принаймні в тому вигляді, як описано в патенті), але патент містить цікаві ідеї, з якими, на нашу думку, кожен оптимізатор повинен бути ознайомлений. Враховуючи, що цей розділ містить теоретичну інформацію, а не практичний посібник, давайте розглянемо основні ідеї алгоритму LocalRank:

Використовуючи певний алгоритм, вибирається певна кількість документів, які відповідають запиту користувача (позначимо це число як  $N$ ). Ці документи спочатку впорядковані за певним критерієм (це може бути PageRank, оцінка релевантності або інший критерій або їх комбінація). Ми позначимо числове вираження цього критерію як OldScore.

Кожна з  $N$  сторінок проходить нову процедуру ранжування, в результаті якої кожна сторінка отримує новий ранг. Ми позначимо його як LocalScore.

На цьому етапі значення OldScore і LocalScore множаться, у результаті чого отримується нове значення NewScore, за яким і відбувається остаточне ранжування сторінок.

Ключовим у цьому алгоритмі є новий процес ранжування, в результаті якого кожній сторінці надається новий ранг, LocalScore. Давайте більш детально опишемо цей процес.

За допомогою певного алгоритму ранжування відбираються  $N$  сторінок, які відповідають запиту користувача. Новий алгоритм ранжування буде працювати тільки з цими  $N$  сторінками. Кожна сторінка в цій групі має

певний ранг *OldScore*.

Під час розрахунку *LocalScore* для даної сторінки виділяються всі сторінки з *N*, які мають зовнішні посилання на дану сторінку. Ми позначимо цю множину як *M*. При цьому в множину *M* не потраплять сторінки з того ж хоста (*host*, фільтрація відбувається за IP-адресою), а також сторінки, які є дзеркалами даної сторінки.

Множина *M* розбивається на підмножини *L<sub>i</sub>*. В ці підмножини потрапляють сторінки, об'єднані наступними ознаками.

1. Належність до одного (або схожого) хоста. Таким чином, в одну групу потраплять сторінки, у яких перші три октети IP-адреси співпадають. Іншими словами, сторінки, IP-адреса яких належать до діапазону xxx.xxx.xxx.0 – xxx.xxx.xxx.255, вважаються належними до однієї групи.

2. Сторінки, які мають однаковий або схожий вміст (дзеркала, mirrors).

3. Сторінки одного сайту (домену).

Кожна сторінка в кожній підмножині *L<sub>i</sub>* має певний ранг (*OldScore*). З кожної підмножини вибирається одна сторінка з найвищим значенням *OldScore*, інші вилучаються з розгляду. Таким чином, ми отримуємо певну множину *K* сторінок, які посилаються на дану сторінку.

Сторінки в множині *K* сортуються за параметром *OldScore*, а потім відбираються лише перші *k* сторінок (де *k* – певне задане число), інші сторінки вилучаються з розгляду.

На цьому етапі розраховується *LocalScore*. Для залишених *k* сторінок обчислюється сума їх значень *OldScore*. Це можна виразити наступною формулою (4.1).

$$LocalScore = \sum_{i=1}^k OldScore(i)^m \quad (4.1)$$

Тут *m* – це деякий заданий параметр, який може змінюватися від 1 до 3.

Нажаль, інформація, яка міститься в патенті на описуваний алгоритм, не надає докладного опису цього параметра.

Після розрахунку LocalScore для кожної сторінки з множини N завершується розрахунок значень NewScore і пересортування сторінок відповідно до нового критерію. Для розрахунку NewScore використовується формула (4.2).

$$\text{NewScore}(i) = (a + \text{LocalScore}(i) / \text{MaxLS}) * (b + \text{OldScore}(i) / \text{MaxOS}) \quad (4.2)$$

$i$  – це сторінка, для якої розраховується новий рейтинг;

$a$  і  $b$  – деякі числа (патент не надає більш докладної інформації щодо цих параметрів);

MaxLS – максимальне з розрахованих значень LocalScore;

MaxOS – максимальне зі значень OldScore.

Тепер спробуємо відірватися від математики і повторимо все сказане простою мовою.

На першому етапі відбувається відбір певної кількості сторінок, які відповідають запиту. Це робиться за алгоритмами, які не враховують тематику посилань (наприклад, за релевантністю та загальною популярністю посилань).

Після того, як група сторінок визначена, буде підрахована локальна популярність кожної зі сторінок. Усі сторінки, однаково, пов'язані з темою пошукового запиту  $i$ , отже, мають частково схожу тематику. Проаналізувавши посилання одна на одну в обраній групі сторінок (ігноруючи всі інші сторінки в Інтернеті), отримаємо локальну (тематичну) популярність посилань.

Після цього кроку у нас є значення OldScore (рейтинг сторінки на основі релевантності, загальної популярності посилань та інших факторів) та LocalScore (рейтинг сторінки серед тематично пов'язаних сторінок). Загальний рейтинг і ранжування сторінок проводиться на основі поєднання

цих двох факторів.

Всі ідеї щодо оптимізації тексту та підвищення популярності посилань, які були висловлені вище, застосовні до всіх пошукових систем у рівній мірі. Більш докладний опис Google пояснюється великою кількістю інформації про цю пошукову систему, що доступна вільно, проте ідеї, висловлені щодо Google, в значній мірі застосовні і до інших пошукових систем.

Загалом, я не є прихильником пошуку «секретного знання» щодо того, як саме працюють алгоритми різних пошукових систем. Усі вони, в певному вигляді, підкоряються загальним правилам, і правильна робота над сайтом (без врахування якихось особливостей) призводить до хороших позицій практично в усіх пошукових системах.

Тем не менше, ось деякі особливості різних пошукових систем:

Google – дуже швидка індексація, велике значення приділяється зовнішнім посиланням. Базу Google використовує дуже велика кількість інших пошукових систем і порталів.

MSN – більший, ніж у інших пошукових систем, акцент на інформаційному змісті сайту.

Rambler – найбільш загадкова пошукова система. За наявними спостереженнями, знижує в рейтингу сайти, які активно займаються просуванням (швидке збільшення кількості зовнішніх посилань). Віддає перевагу наявності пошукових термінів у простому тексті сторінки (без виділення різними стилістичними тегами).

#### 4.3 Підготовка до створення ефективного сайту

У цьому розділі представлена інформація, яка виникла в результаті аналізу різних статей, спілкування оптимізаторів, практичних спостережень і т. п. Ця інформація не є точною і надійною – це лише припущення і ідеї, але ідеї цікаві. Дані, які представлені в цьому розділі, слід сприймати не як точне керівництво, а як інформацію для роздумів.

Вихідні посилання. Посилайтеся на авторитетні ресурси в вашій області, використовуючи потрібні ключові слова. Пошукові системи цінують посилання на інші ресурси тієї ж тематики.

Вихідні посилання. Не посилайтеся на FFA-сайти та інші сайти, які виключені з індексу пошукової системи. Це може призвести до зниження рейтингу вашого власного сайту.

Вихідні посилання. Сторінка не повинна містити більше 50-100 вихідних посилань. Це не призводить до зниження рейтингу сторінки, але посилання понад цю кількість не будуть враховані пошуковою системою.

Зовнішні site-wide посилання, тобто посилання, які розміщені на кожній сторінці сайту. Вважається, що пошукові системи негативно ставляться до таких посилань і не враховують їх при ранжуванні. Існує також інше думку, що це стосується лише великих сайтів з тисячами сторінок.

Ідеальна щільність ключових слів. Дуже часто доводиться чути подібне запитання. Відповідь полягає в тому, що ідеальної щільності ключових слів не існує, або, точніше, вона різна для кожного запиту, тобто обчислюється пошуковою системою динамічно, залежно від пошукового терміна. Наша порада – проаналізувати перші сайти в результатах пошукової системи, що дозволить приблизно оцінити ситуацію;

Вік сайту. Пошукові системи віддають перевагу старим сайтам, як більш стабільним.

Оновлення сайту. Пошукові системи віддають перевагу розвиваються сайтам, тобто тим, на яких періодично додається нова інформація, нові сторінки;

Доменна зона (стосується західних пошукових систем). Віддається перевага сайтам, розташованим в зонах .edu, .mil, .gov і т.п. Такі домени можуть зареєструвати лише відповідні організації, тому довіри до таких сайтів більше;

Пошукові системи відслідковують, який відсоток відвідувачів

повертається до пошуку після відвідування того чи іншого сайту з результатів пошуку. Великий відсоток повернень свідчить про не тематичний зміст, і така сторінка понижується в пошуку;

Пошукові системи відслідковують, наскільки часто вибирається та чи інша посилання в результатах пошуку. Якщо посилання вибирається рідко, це означає, що сторінка не представляє цікавості, і така сторінка понижується в рейтингу;

Використовуйте синоніми та споріднені форми ключових слів, це буде оцінено пошуковими системами;

Занадто швидкий зріст кількості зовнішніх посилань сприймається пошуковими системами як штучне просування і призводить до зниження рейтингу. Дуже спірне твердження, передусім через те, що такий спосіб може використовуватися для зниження рейтингу конкурентів;

Google не враховує зовнішні посилання, якщо вони розміщені на одному (або схожому) хості, тобто сторінках, IP-адреса яких належать діапазону xxx.xxx.xxx.0 xxx.xxx.xxx.255. Така думка, швидше за все, виникла через те, що Google висловлював цю ідею в своїх патентах. Проте співробітники Google заявляють, що немає жодних обмежень за IP-адресою на зовнішні посилання і немає підстав не довіряти їм;

Пошукові системи перевіряють інформацію про власника домену. Відповідно посилання з сайтів, які належать одному власнику, мають менший вагомий, ніж звичайні посилання. Інформація представлена в патенті;

Термін, на який зареєстрований домен. Чим довший термін, тим більша перевага надається сайту.

## 5 КОНКУРЕНТОСПРОМОЖНІСТЬ КЛЮЧОВИХ СЛІВ

### 5.1 Статистичний аналіз параметрів сайту

Особливістю цього дослідження є те, що для аналізу параметрів веб-сайтів використовувалися методи математичної статистики, а саме – кореляційний та регресійний аналіз.

Коефіцієнт кореляції – двовимірна описова статистика, кількісний показник взаємозв'язку (спільної змінності) двох змінних.

Історія розробки та застосування коефіцієнтів кореляції для вивчення взаємозв'язків фактично почалася одночасно з виникненням вимірювального підходу до дослідження індивідуальних відмінностей – у 1870-1880 роках. Піонером у вимірюванні здібностей людини, а також автором самого терміна «коефіцієнт кореляції» був Френсіс Гальтон, і найпопулярніші коефіцієнти кореляції були розроблені його послідовником Карлом Пірсоном. З тих пір вивчення взаємозв'язків за допомогою коефіцієнтів кореляції є однією з найпопулярніших справ в психології.

На сьогоднішній день існує велика кількість різних коефіцієнтів кореляції, і сотні книг присвячені вимірюванню взаємозв'язку за їх допомогою. Тому, не претендуючи на повноту викладу, ми розглянемо лише найважливіші, дійсно незамінні в дослідженнях міри зв'язку –  $r$ -Пірсона,  $r$ -Спірмена та  $\tau$ -Кендалла. Загальною особливістю їх є те, що вони відображають взаємозв'язок двох ознак, виміряних у кількісній шкалі – ранжованій або метричній.

Загалом, будь-яке емпіричне дослідження спрямоване на вивчення взаємозв'язків двох або більше змінних.

Будь-яке дослідження можна звести до вивчення кореляцій, оскільки були розроблені найрізноманітніші коефіцієнти кореляції для практично

будь-якої дослідницької ситуації. Але в подальшому викладі ми будемо розрізняти два класи завдань:

- дослідження кореляцій – коли дві змінні представлені в числовій шкалі;
- дослідження відмінностей – коли принаймні одна з двох змінних представлена в номінативній шкалі.

Взаємозв'язки в математиці зазвичай описуються за допомогою функцій, які графічно відображаються у вигляді ліній. Якщо зміна однієї змінної на одиницю завжди призводить до зміни іншої змінної на однакову величину, функція є лінійною (її графік представляє собою пряму лінію); будь-який інший зв'язок – нелінійний. Якщо збільшення однієї змінної пов'язане зі збільшенням іншої, то зв'язок – позитивний (прямий); якщо збільшення однієї змінної пов'язане зі зменшенням іншої, то зв'язок – від'ємний (зворотний). Якщо напрямок зміни однієї змінної не змінюється зі зростанням (зменшенням) іншої змінної, то така функція – монотонна; в іншому випадку функцію називають немонотонною.

Коефіцієнт кореляції – це кількісна міра сили і напрямку ймовірного зв'язку двох змінних; приймає значення в діапазоні від  $-1$  до  $+1$ . Сила зв'язку досягає максимуму при взаємній однозначній відповідності: коли кожному значенню однієї змінної відповідає лише одне значення іншої змінної (і навпаки), емпіричний зв'язок при цьому співпадає з функціональним лінійним зв'язком. Показником сили зв'язку є абсолютна (без врахування знаку) величина коефіцієнта кореляції. Напрямок зв'язку визначається прямим або зворотним відношенням значень двох змінних: якщо зростанню значень однієї змінної відповідає зростання значень іншої змінної, то взаємозв'язок називається прямим (позитивним); якщо зростанню значень однієї змінної відповідає зменшення значень іншої змінної, то взаємозв'язок є зворотнім (від'ємним). Показником напрямку зв'язку є знак коефіцієнта кореляції.

Коефіцієнт кореляції  $r$ -Пірсона (Pearson  $r$ ) застосовується для вивчення взаємозв'язку двох метричних змінних, які вимірюються на одній і тій же

вибірці. Існує багато ситуацій, в яких доцільно його використовувати. Чи впливає інтелект на успішність на старших курсах університету? Чи пов'язано розмір заробітної плати працівника з його доброзичливістю до колег. Чи впливає настрої школяра на успішність розв'язання складної арифметичної задачі? Для відповіді на подібні питання дослідник повинен виміряти два його цікавлячих показники для кожного члена вибірки. Дані для вивчення взаємозв'язку потім об'єднуються в таблицю, як у прикладі нижче.

Перед тим як навести формулу коефіцієнта кореляції, давайте спробуємо простежити логіку її виникнення. Положення кожної  $i$ -точки (досліджуваної особи з номером  $i$ ) на діаграмі розсіювання відносно інших точок може бути задане величинами та знаками відхилень відповідних значень змінних від їх середніх значень:  $(x_i - M_x)$  та  $(y_i - M_y)$ . Якщо знаки цих відхилень співпадають, то це свідчить на користь позитивного зв'язку (великі значення за  $x$  відповідають великим значенням за  $y$  або малим значенням за  $x$  відповідають малим значенням за  $y$ ).

Отже, якщо добуток відхилень  $(x_i - M_x) * (y_i - M_y)$  є позитивним, то дані  $i$ -го досліджуваного свідчать про прямий (позитивний) зв'язок, а якщо він є від'ємним, то про зворотний (від'ємний) зв'язок. Відповідно, якщо  $x$  і  $y$  переважно прямо пропорційні, то більшість добутоків відхилень буде позитивними, і якщо вони пов'язані зворотним співвідношенням, то більшість добутоків буде від'ємними. Отже, загальним показником для сили і напрямку зв'язку може слугувати сума всіх добутоків відхилень для даної вибірки:

У випадку прямої пропорційної зв'язку між змінними ця величина є великою і позитивною – для більшості досліджуваних відхилень збігаються за знаком (великим значенням однієї змінної відповідають великі значення іншої змінної і навпаки). Якщо  $x$  і  $y$  мають зворотну залежність, то для більшості досліджуваних більшим значенням однієї змінної будуть відповідати менші значення іншої змінної, тобто знаки добутоків будуть від'ємними, і сума добутоків в цілому також буде великою за абсолютною

величиною, але від'ємною за знаком. Якщо систематичного зв'язку між змінними не спостерігається, то позитивні доданки (добутки відхилень) вирівнюються від'ємними доданками, і сума всіх добутків відхилень буде близькою до нуля.

Щоб сума добутків не залежала від обсягу вибірки, досить її усереднити. Проте міра зв'язку нас цікавить як обчислювана оцінка, а не як генеральний параметр – статистика. Тому, так само як і для формули дисперсії, в цьому випадку ми ділимо суму добутків відхилень не на  $N$ , а на  $N-1$ .

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^N (x_i - M_x)^2 * \sum_{i=1}^N (y_i - M_y)^2}} \quad (5.1)$$

Рівняння (5.1) є основною формулою коефіцієнта кореляції Пірсона. Ця формула цілком змістовна, але не дуже зручна для обчислень «вручну» або на калькуляторі. Тому існують похідні формули – більш складні за виглядом, менш зрозумілі, але спрощують обчислення.

Знову зауважимо: розмір коефіцієнта кореляції не залежить від одиниць вимірювання ознак. Отже, будь-які лінійні перетворення ознак (множення на константу, додавання константи:  $(y_i = x_i b + a)$ ) не змінюють значення коефіцієнта кореляції. Винятком є множення однієї з ознак на від'ємну константу: коефіцієнт кореляції змінює свій знак на протилежний. Отже, коефіцієнт кореляції Пірсона є мірою прямолінійного зв'язку; він не чутливий до криволінійних зв'язків.

## 5.2 Вплив параметрів сайту на його позицію у видачі

Для проведення експерименту щодо впливу параметрів сайту на його позицію у видачу пошуковика необхідно:

- створити два сайти-зразки із розбіжністю у параметрах, що впливають на успіх просування ресурсу;
- завантажити обидва сайти на сервери хостингу в однаковий час;
- після успішного індексування та появи у видачі пошукової системи Google проаналізувати позиції зразків за декількома обраними ключовими словами;
- обчислити коефіцієнт кореляції Пірсона для складності обраних ключових слів та різниці позицій зразків у видачах за ними ж;
- визначити рівняння регресії для встановлення залежності між згаданими вище величинами.

Нижче розглянемо результати отримані на кожному з перелічених етапів. Загальною метою експерименту є саме визначення рівняння регресії, що дозволить надати рекомендації щодо виділення ресурсів та часу на SEO-оптимізацію під ключові слова із різною складністю.

В першу чергу було створено сайт-зразок із максимальною можливою кількістю прийомів оптимізації (далі – перший зразок). Розглянемо деякі з них. Під час написання HTML-розмітки, окрім наповнення текстового контексту достатньою кількістю ключових слів відповідної тематики, було виконано наступні дії:

- за рахунок використання семантично вірних тегів досягнуто високої якості семантики web-сторінок;
- до заголовків сторінок додано мета-теги із додатковою інформацією, особливо для коректного зовнішнього вигляду ресурсу у видачі;
- застосовано усі можливі відносно невеликі методи покращення сторінки (наприклад, alt-атрибути для всіх зображень).

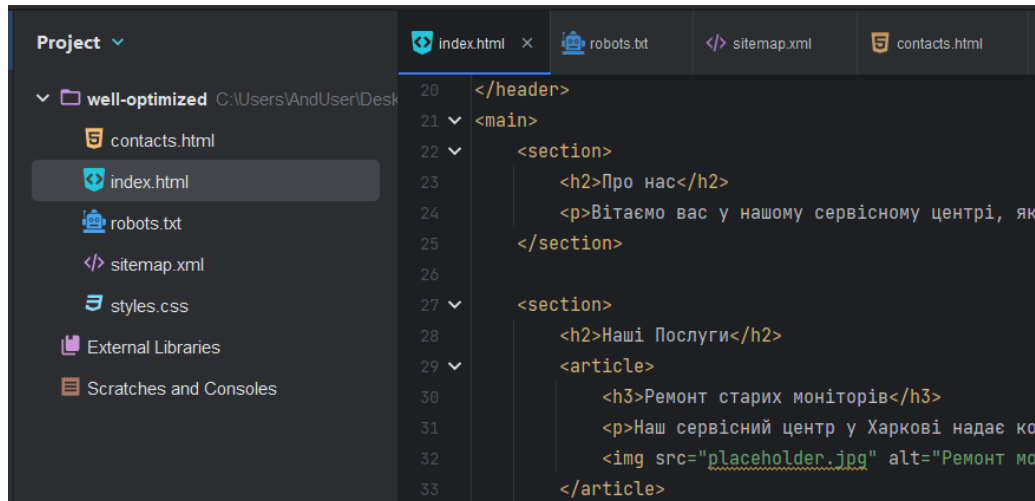


Рисунок 5.1 – Приклад розмітки та структури першого зразка із семантичним підсвічуванням середовища розробки WebStorm

Також перший зразок отримав файли robots.txt та sitemap.xml, що також має позитивно вплинути на позицію у видачі пошукових систем. Варто відмітити, що загалом вихід на перші позиції не є головною ціллю експерименту. Загалом абсолютні значення позицій не становлять основний інтерес для даного дослідження – найголовнішим вимірюваним показником буде саме позиція зразків відносно один одного. Тож, розглянемо процес створення другого зразка.

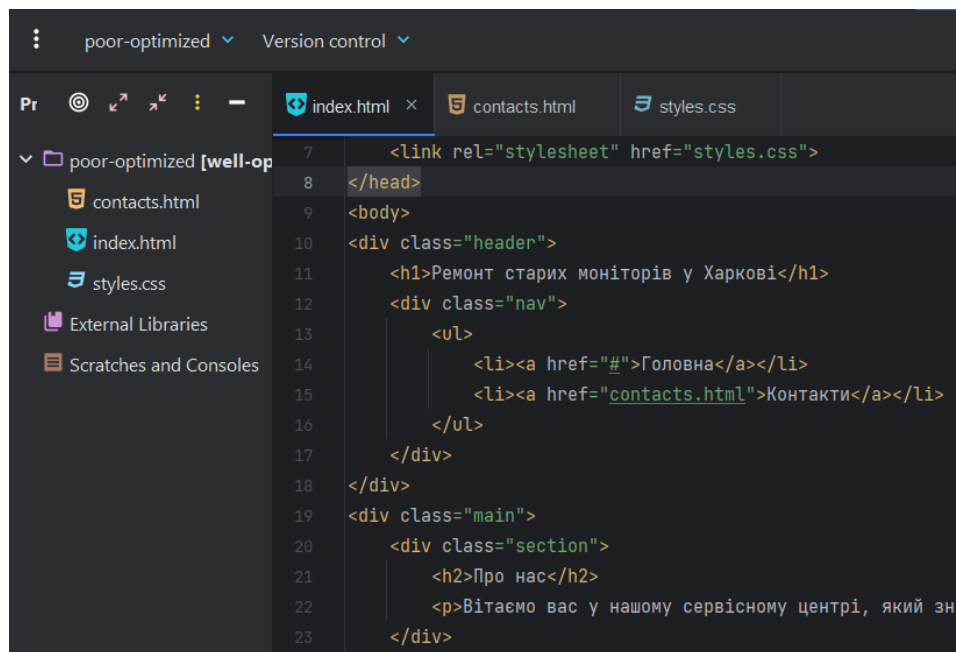
Фактично, другий зразок є відредагованою копією першого, що було зроблено для досягнення мінімальної залежності результату експерименту від зовнішніх факторів. Розглянемо задачі, що були виконані під час редагування розмітки та структури сторінок.

1. Файли robots.txt та sitemap.xml були видалені. Наявність цих файлів полегшую процес індексування сторінок сайту пошуковими роботами, але не є обов'язковою.

2. В усіх можливих випадках семантично вірні теги були замінені на звичайні контейнери div. Під можливими випадками мається на увазі усі елементи розмітки тіла сторінки, що не впливають на зовнішній вигляд у

видачі пошуковика (фактично, залишилися лише заголовки).

3. Також була зменшена кількість мета-інформації у заголовку сторінки та деяких тегах тіла. Як і у ситуації з семантикою, повне видалення метаданих не було можливим, тож для збереження однакового вигляду у видачі із першим зразком деякі теги не були видалені (наприклад, мета-тег з ім'ям description).



```

7 <link rel="stylesheet" href="styles.css">
8 </head>
9 <body>
10 <div class="header">
11 <h1>Ремонт старих моніторів у Харкові</h1>
12 <div class="nav">
13 <ul>
14 <li><a href="#">Головна</a></li>
15 <li><a href="contacts.html">Контакти</a></li>
16 </ul>
17 </div>
18 </div>
19 <div class="main">
20 <div class="section">
21 <h2>Про нас</h2>
22 <p>Вітаємо вас у нашому сервісному центрі, який зна
23 </div>

```

Рисунок 5.2 – Приклад розмітки та структури другого зразка із семантичним підсвічуванням середовища розробки WebStorm

Після успішного створення зразків, їх було завантажено на підготовлені заздалегідь домени за допомогою інструментів хостингу UKRAINE-hosting. Загалом, даний продукт також надає велику кількість додаткових інструментів для хостингу повноцінних web-додатків, проте у даному експерименті він був використаний лише як Apache-сервер для зберігання статичних файлів зразків, тож приділяти йому багато уваги не є доцільним.

Під час очікування завершення процесу індексації та появи зразків у

видачі пошуковика були вибрані та вивчені ключові слова, що буду виступати основними точками для дослідження залежності між складністю ключового слова та різниці позицій створених зразків. На рисунку 5.3 наведено процес визначення кількості проіндексованих сторінок за певними ключовими словами, що будуть використані для подальшого обчислення їх складності.

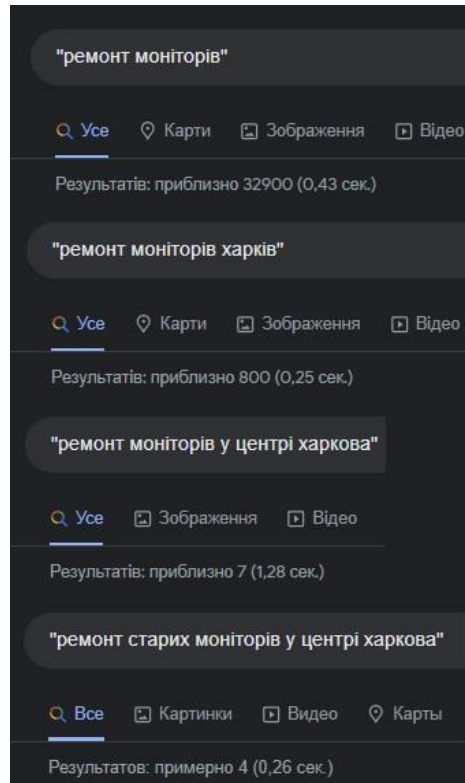


Рисунок 5.3 – Дослідження кількості проіндексованих сторінок за досліджуваними ключовими словами

Для подальшого дослідження необхідно обчислити складність обраних ключових слів, яка дорівнює добутку частотності слова та кількості проіндексованих сторінок, що була встановлена раніше. За допомогою додатку Google Ads можливо отримати дані про частотність ключових слів будь-якою мовою. Відповідно до назви, цей додаток призначений для просування ресурсів шляхом реклами, тож визначення частотності є єдиною його функцією, що була використана під час дослідження.

Таблиця 5.1 – Складність обраних ключових слів

Ключове слово	Частотність (X)	Кількість проіндексованих сторінок (Y)	Складність (XY)
ремонт старих моніторів у центрі харкова	2	4	8
ремонт моніторів у центрі харкова	5	7	35
ремонт моніторів харків	9	800	7200
ремонт моніторів	102	32900	3355800

У таблиці 5.1 наведені результати обчислення складності обраних ключових слів. Остання колонка містить дані, готові для використання у обчисленні коефіцієнту кореляції Пірсона за віссю X. Тепер розглянемо аналіз позицій у видачі зразків, що були успішно проіндексовані роботом Google через певний проміжок часу.

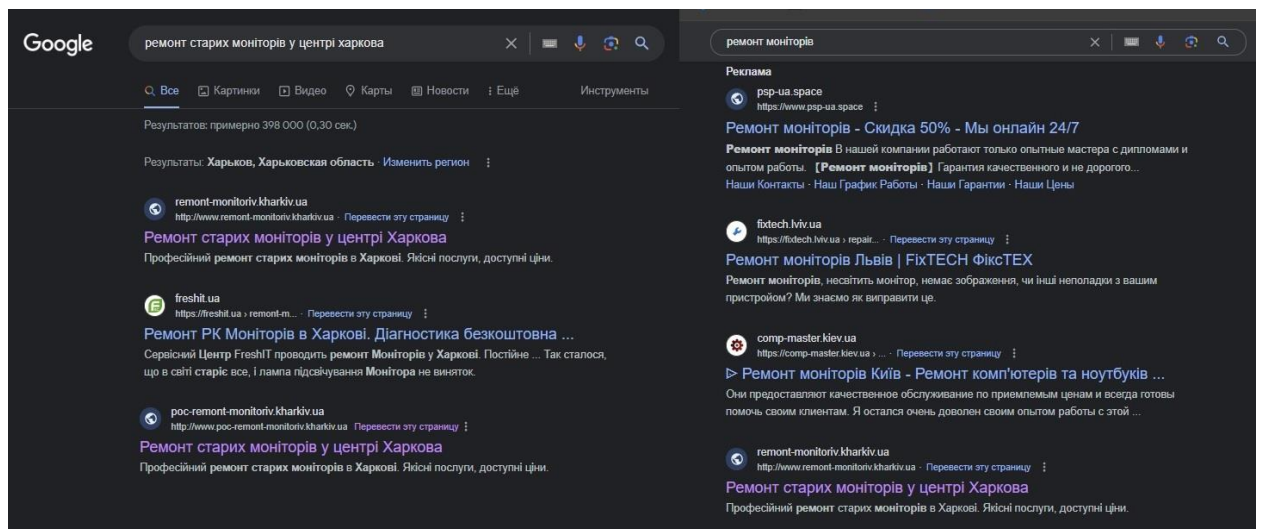


Рисунок 5.4 – Візуальний аналіз позицій зразків у видачі

В першу чергу було проведено ручний аналіз позицій обох зразків у видачі за обраними ключовими словами. Як результат маємо вже з перших днів достатньо високі позиції як для досить молодих сайтів: перший зразок займає першу позицію для ключового слова з найнижчою складністю та трохи більшу за сімдесяту для найвищою складності. Другий зразок має також непогану позицію за ключовим словом з найменшою конкуренцією, втім стрімко втрачає позиції при збільшенні складності запиту.

Враховуючи описані вище візуально підтвержені залежності висуваємо гіпотезу, що складність ключового слова та приріст позиції за рахунок оптимізації мають позитивну залежність. Після отримання точних даних про позиції зразків та проведення обчислень буде можливим оцінити вірність такого припущення та, в разі наявності залежності, її силу.

За кілька днів стає можливим визначити конкретні позиції за допомогою статистики додатку Google Search Console. В першу чергу необхідно визначити період, дані з якого будуть використані для обчислень. Для цього проаналізуємо зміни позицій першого зразка за перші дні появи у видачі (рисунок 5.5).

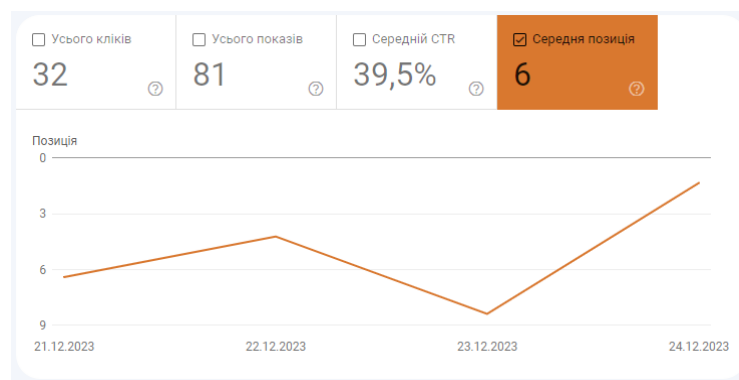


Рисунок 5.5 – Графік зміни середньої позиції у видачі для першого зразка

В результаті дослідження середнього значення позиції першого зразка було вирішення взяти для розгляду дані за 23 грудня. Низьке значення у даній точці обумовлене наявністю достатньою кількістю потраплянь у видачу

за більш конкурентними запитами. Висока середня позиція із показником 1,3 за 24 грудня, у свою чергу підсвічує високу вірогідність того, що зразок потрапляв до видачі в основному за ключовими словами із малою складністю, де він займає лідируючі позиції. Очевидно, така ситуація негативно впливатиме на точність результатів обчислень.

Отже, після аналізу позицій стає можливим знаходження значень другої змінної для обчислення коефіцієнту кореляції Пірсона – різниці позицій другого та першого зразків. В першу чергу формуємо запит до Google Search Console що видасть статистику щодо потрапляння обох зразків до видачі за обраними ключовими словами. Також фільтруємо дані для показу лише позицій зразків, так як інші аспекти (наприклад, відсоток показів, що призвели до переході на сайт) не мають особливого сенсу для дослідження, що проводиться. Отримані дані представлено на рисунку 5.6.

Популярні запити	↓ Позиція	Популярні запити	↓ Позиція
ремонт старих моніторів у центрі харкова	1	ремонт старих моніторів у центрі харкова	3
ремонт моніторів у центрі харкова	6	ремонт моніторів у центрі харкова	14
ремонт моніторів харків	23	ремонт моніторів харків	40
ремонт моніторів	83	ремонт моніторів	152

Рисунок 5.6 – Позиції першого (ліворуч) та другого (праворуч) зразків у видачі пошуковика до вибраних ключових слів за даними GSC

Загалом, отримані дані в деякій мірі збігаються з результатами ручного дослідження проведеного раніше, що підвищує вірогідність правдивості висунутої гіпотези. Тепер заповнимо таблиці 5.2 та 5.3 необхідними для подільших обчислень даними.

Таблиця 5.2 – Позиції зразків у видачі за обраними ключовими словами

Ключове слово	Позиція першого зразка (X)	Позиція другого зразка (Y)	Різниця позицій (Y - X)
ремонт старих моніторів у центрі харкова	1	3	2
ремонт моніторів у центрі харкова	6	14	6
ремонт моніторів харків	23	40	17
ремонт моніторів	83	152	69

Таблиця 5.3 – Дані для обчислення коефіцієнту кореляції Пірсона

Ключове слово	Складність ключового слова (X)	Різниця позицій зразків (Y)	XY
ремонт старих моніторів у центрі харкова	8	2	16
ремонт моніторів у центрі харкова	35	6	280
ремонт моніторів харків	7200	17	122400
ремонт моніторів	3355800	69	1021545000

В попередньому підрозділі наведено формулу (5.1) коефіцієнта кореляції Пірсона, що є змістовною та короткою, але незручною для

обчислень. Тож для проведення дослідження будуть використовуватися альтернативні формули.

1. В першу чергу знаходиться сума квадратів відхилень значень першого параметра від середнього арифметичного значення за допомогою рівняння 5.2.

$$SS_x = \sum_{i=1}^N X_i^2 - \frac{\left(\sum_{i=1}^N X_i\right)^2}{N} = 2827514554962,25 \quad (5.2)$$

2. Тепер необхідно виконати аналогічні обчислення для другого параметра

$$SS_y = \sum_{i=1}^N Y_i^2 - \frac{\left(\sum_{i=1}^N Y_i\right)^2}{N} = 2304 \quad (5.3)$$

3. Знаходиться так звана скоригована на середні значення сума попарних добутків X та Y, за формулою:

$$SP_{xy} = \sum_{i=1}^N X_i Y_i - \frac{\sum_{i=1}^N X_i \sum_{i=1}^N Y_i}{N} = 79031510 \quad (5.4)$$

4. За допомогою формули 5.5 знаходиться коефіцієнт кореляції Пірсона:

$$R_{xy} = \frac{SP_{xy}}{\sqrt{SS_x SS_y}} = 0,98 \quad (5.5)$$

За отриманим значенням  $R_{xy}=0,98$  можна назвати кореляцію досить

сильною, адже коефіцієнт має майже максимальне значення. Отже, гіпотезу про те, що із зростанням складності ключового слова зростає кількість позицій, на яку підіймається сайт відносно не оптимізованого зразка, можна вважати підтвердженою. Наступним кроком буде побудова рівняння регресії та його графіку для уточнення конкретної залежності між згаданими вище величинами. Таке рівняння має наступний вигляд:

$$y_i = b x_i + a \quad (5.6)$$

де  $b$  – коефіцієнт регресії,  $a$  – коефіцієнт зміщення.

Для отримання рівняння регресії використаємо метод найменших квадратів. Візьмемо складність ключового слова як аргумент та середню кількість позицій, на яку підніметься сайт як значення при побудові функції. Запишемо вихідні дані до таблиці 5.4 та розглянемо процес побудови рівняння.

Таблиця 5.4 – Вихідні дані для побудови рівняння регресії

Складність ключового слова (X)	Різниця позицій (Y)
8	2
35	6
7200	17
3355800	69

1. Знайдемо середні значення  $X$  та  $Y$  (рівняння 5.7):

$$X = \frac{\sum_{i=1}^N X_i}{N} = 839810,75 \quad Y = \frac{\sum_{i=1}^N Y_i}{N} = 24 \quad (5.7)$$

2. Обчислимо коефіцієнт регресії:

$$b = \frac{\sum_{i=1}^N (X_i - X)(Y_i - Y)}{\sum_{i=1}^N (X_i - X)^2} = 1.72939E - 5 \quad (5.8)$$

3. Обчислимо коефіцієнт зміщення:

$$a = Y - bX = 9.9933157 \quad (5.9)$$

Тоді рівняння регресії матиме вигляд:

$$y = 1.72939E - 5 * x + 9.9933157 \quad (5.10)$$

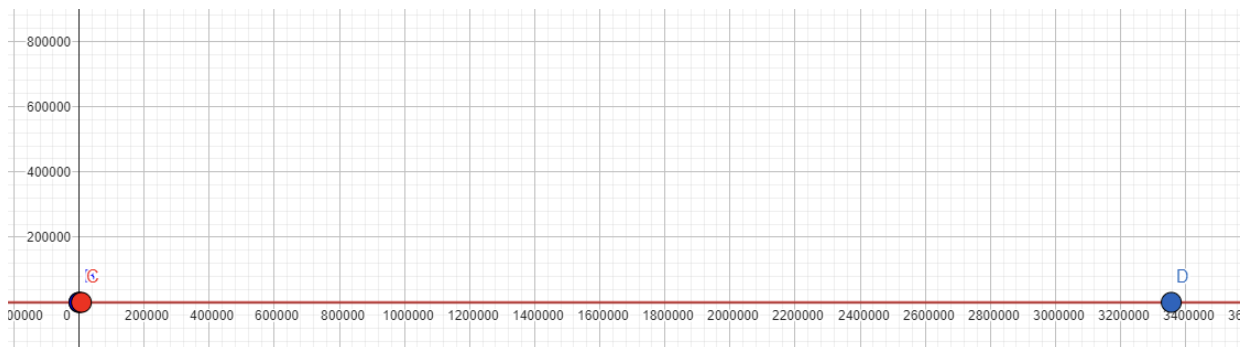


Рисунок 5.7 – Розсіювання параметру та лінія регресії

На рисунку 5.7 наведено графік отриманого рівняння регресії та розсіювання параметрів, що виступали вихідними даними. Графік очікувано лінія регресії помітно полого, близька до прямої, адже коефіцієнт регресії дуже малий. Загалом такий результат цілком очікуваний, адже ще на етапі ручного дослідження позицій зразків у видачі було очевидно, що кількість позицій, на які підніметься сайт або загальна кількість позицій у видачі не

зможе так змінюватись із тою ж швидкістю, що й складність ключових слів.

Загалом, не зважаючи на те, що отриманий графік не надає достатньо інформації своїм зовнішнім виглядом, отримане рівняння регресії може бути використане для обчислення орієнтовного приросту кількості позицій. Однак, можливо покращити візуальне сприйняття графіку змінивши пропорції між одиницями вимірювання на осях, або змінивши одиниці вимірювання осі абсцисс. Задля виключення плутанини з одиницями вимірювання було вирішено змінити пропорції.

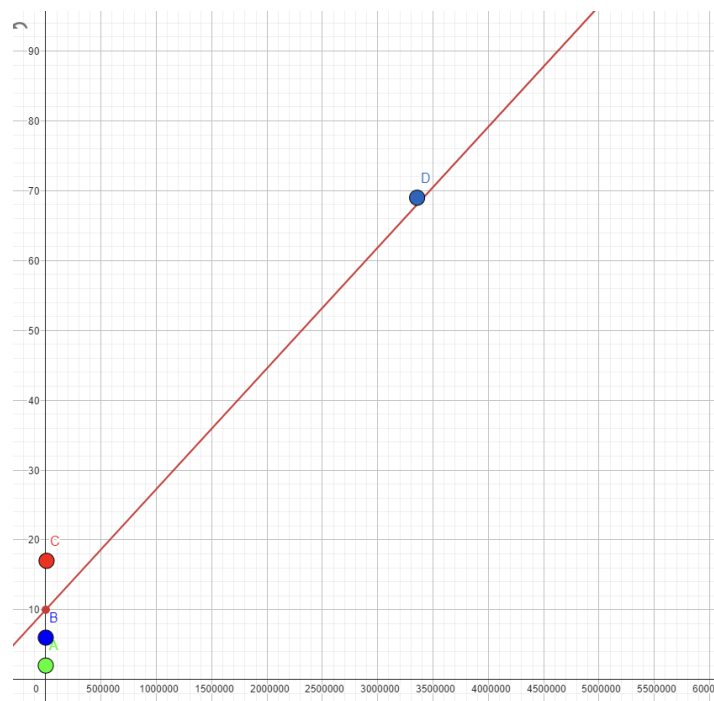


Рисунок 5.8 – Скореговані розсіювання параметру та лінія регресії

Скорегований графік не є вірним математично, адже порушено логіку коефіцієнта нахилу лінійної функції. Проте він надає можливість візуально оцінити розсіювання параметру навколо лінії регресії та можливий приріст позиції ресурсу при просуванні під ключові слова певної складності.

## ВИСНОВКИ

Розкрито питання про те, як параметри сайту впливають на його позицію у видачі пошукової систем? Чи є залежність між складністю ключового слова та покращенням позиції сайту у видачі? На ці запитання отримані відповіді шляхом ретельного вивчення зв'язку між складністю ключового слова та кількістю позицій, на яку підніметься сайт після застосування принаймні базових методів оптимізації. Результат цього дослідження дуже важливий, оскільки привів нас до більш глибокого розуміння поведінки алгоритму ранжування google.

Висунута та перевірена гіпотеза про позитивну залежність між складністю ключового слова та вагомістю результатів SEO-оптимізації. За результатами обчислення коефіцієнту кореляції гіпотеза підтвердилась, а також сила кореляції виявилася значно більшою за очікувану.

Отримано рівняння регресії, за яким можна прогнозувати потенційну вигоду від інвестиції часу у просування сайту за ключовими словами певної складності. Візуальне визначення орієнтовних значень за отриманим графіком ускладнене малим кутовим коефіцієнтом отриманої лінійної функції. Для можливості без точних обчислень швидко оцінити орієнтовне значення побудовано скоригований графік, що не є математично вірним але краще сприймається неозброєним оком.

Також дане дослідження має перспективу покращення результатів. Є досить висока вірогідність що досліджувана залежність має експоненційний характер та рівняння може бути покращене для більш точного прогнозування. Таке покращення матиме максимальний ефект для надвеликих значень складності ключових слів, тож із часом стане ще актуальнішим, адже наявна тенденція до постійного збільшення цього параметру.

Збільшення частотності ключових слів за рахунок збільшення

користувачів Інтернету та об'єму за рахунок постійного поповнення баз пошукових систем новими web-сторінками призведе до підвищення актуальності даного дослідження та будь-яких його продовжень.

Також на основі вивченої для коректного проведення досліджень інформації сформовано декілька рекомендацій щодо просування ресурсу.

По-перше, якщо сайт маленький, це може означати, що цей сайт ще молодий і недостатньо розвинений. Подібні сайти не слід орієнтувати на ключові слова з високою конкурентоспроможністю, оскільки в цьому випадку Google віддасть перевагу великим сайтам.

По-друге, великі сайти можуть досягти хороших результатів при ранжуванні за конкурентоспроможними ключовими словами, якщо також врахують і низку інших важливих факторів ранжування таких як: кількість посилань, контент та авторитетність;

□ По-третє, єдине, що Ви можете зробити зараз для Вашого сайту, щоб у майбутньому розраховувати на успішне ранжування, це наповнювати та оновлювати контент. Це допоможе значно збільшити розмір Вашого сайту, що є позитивним фактором під час ранжування за ключовими словами з високою конкурентоспроможністю.

По-четверте, великі сайти мають певну перевагу для отримання природних посилань та трафіку. Внутрішні сторінки вашого сайту згодом набудуть важливості, і також залучатимуть трафік за відповідними пошуковими запитам.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Лебеденко М. С. ВебOMETричний ранг як показник ефективності електронного ресурсу підприємства / М. С. Лебеденко // Економічний вісник Національного технічного університету України "Київський політехнічний інститут". - 2014. - № 11.
2. Enge E., Spencer S., Stricchiola J. The Art of SEO: Mastering Search Engine Optimization. 3rd ed. Sebastopol : O'Reilly Media, 2015. 994 p.
3. Google [Електронний ресурс]: Офіційний веб-сайт Google Analytics – Вебаналітика та звітування – Google Analytics – Режим доступу: <https://www.google.com.ua/intl/uk/analytics/index.html> - Дата доступу: 16.05.2023.
4. Google [Електронний ресурс]: Про сегменти - Analytics Довідка – Режим доступу: <https://support.google.com/analytics/answer/3123951?hl=uk> - Дата доступу: 13.05.2023.
5. Google [Електронний ресурс]: Список функцій Google Analytics – Google Analytics – Режим доступу: [https://www.google.com.ua/intl/uk\\_ALL/analytics/features/index.html](https://www.google.com.ua/intl/uk_ALL/analytics/features/index.html) - Дата доступу: 16.05.2023
6. Calicchio S. A simple approach to SEO : EBook. 2022. 163 p.
7. Parkianon M. Basic SEO explained :EBook. 2021. 37 p.
8. Duong V. Seo management. Oxford : Wiley, 2020. 208 p.