

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет \_\_\_\_\_ Комп'ютерних наук \_\_\_\_\_  
(повна назва)

Кафедра \_\_\_\_\_ Інформаційних управляючих систем \_\_\_\_\_  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_

Дослідження методів Process Mining в задачах пошукового просування сайтів  
в мережі Інтернет \_\_\_\_\_  
(тема)

Виконав:

студент 2 курсу, групи ІУСТМ-21-1

\_\_\_\_\_ Дмитро ЄРОХІН \_\_\_\_\_  
(власне ім'я, прізвище)

Спеціальність \_\_\_\_\_ 122 Комп'ютерні \_\_\_\_\_  
науки \_\_\_\_\_  
(код і повна назва спеціальності)

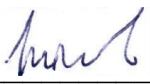
Тип програми \_\_\_\_\_ освітньо-професійна \_\_\_\_\_  
(освітньо-професійна або освітньо-наукова)

Освітня програма \_\_\_\_\_ Інформаційні \_\_\_\_\_  
управляючі системи та технології \_\_\_\_\_  
(повна назва освітньої програми)

Керівник \_\_\_\_\_ проф. кафедри ІУС Оксана ЧАЛА \_\_\_\_\_  
(посада, власне ім'я, прізвище)

Допускається до захисту

Зав. кафедри

  
\_\_\_\_\_ (підпис)

\_\_\_\_\_ Костянтин ПЕТРОВ \_\_\_\_\_  
(власне ім'я, прізвище)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Інформаційних управляючих систем

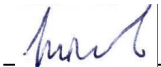
Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва)

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

Освітня програма Інформаційні управляючі системи та технології  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри   
(підпис)

«21» листопада 2022 р.

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Срохіну Дмитру Олексійовичу  
(прізвище, ім'я, по батькові)

1.Тема роботи Дослідження методів Process Mining в задачах пошукового просування сайтів в мережі Інтернет

затверджена наказом університету від 14 листопада 2022 р. № 1490Ст

2.Термін подання студентом роботи до екзаменаційної комісії 14 грудня 2022р.

3.Вихідні дані до роботи задачі пошукової оптимізації сайтів, процес пошукового просування сайтів, аналіз файлів журналу за допомогою Googlebot, процес роботи методу Performance Mining, етап «Моніторинг продуктивності програми», процес аналізу сайту з використанням методу Process Mining, технологія пошукового просування сайтів в мережі Інтернет, фрагмент роботи програми із побудованою моделлю процесів веб-сторінки, приклад log файлу з двома новими булевыми змінними

4.Перелік питань, що потрібно опрацювати в роботі проаналізувати існуючі методи вирішення задачі, проаналізувати способи вдосконалення існуючого методу, удосконалити метод задля отримання результатів, проаналізувати отримані результати


## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз задач пошукового просування сайтів в мережі Інтернет	10.10.22 – 15.10.22	Виконано
2	Аналіз процесу просування сайтів в пошуковій системі в мережі Інтернет	16.10.22 – 22.10.22	Виконано
3	Дослідження методів просування сайтів в пошукових системах в мережі Інтернет	23.10.22 – 29.11.22	Виконано
4	Аналіз методів Process Mining	30.10.22 – 03.11.22	Виконано
5	Опис обраного методу, який буде удосконалюватися	04.11.22 – 08.11.22	Виконано
6	Опис обраного методу, який буде удосконалюватися	09.11.22 – 11.11.22	Виконано
7	Теоретично удосконалений метод	12.11.22 – 15.11.22	Виконано
8	Опис інформаційних технологій, особливості реалізації	16.11.22 – 22.11.22	Виконано
9	Опис програмного забезпечення, яке використовується та реалізація методу	23.11.22 – 25.11.22	Виконано
10	Експериментальна перевірка вдосконаленого методу	26.11.22 – 29.11.22	Виконано
11	Оформлення пояснювальної записки	30.11.22 – 04.12.22	Виконано

Дата видачі завдання 21 листопада 2022 р.

Студент 

(підпис)

Керівник роботи 

(підпис)

проф. кафедри ІУС Оксана Чала

(посада, власне ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка до магістерської кваліфікаційної роботи містить: 69 с., 4 розділи, 12 рис., 7 табл., 33 джерел.

ВЕБ-ПЕРЕГЛЯДАЧ, ВІЗУАЛЬНІ МОДЕЛІ ПРОЦЕСІВ, МЕТОДИ PROCESS MINING, ПОШУКОВА ОПТИМІЗАЦІЯ, ПОШУКОВИЙ РОБОТ, LOG ФАЙЛИ

При виконанні даної магістерської роботи описані методи Process Mining, а саме метод Performance Mining, який буде використовуватися для удосконалення методу пошукового просування сайтів, а саме аналізу сторінки на предмет конкретних атрибутів. Використавши новий метод можна буде зробити ранжування сторінок більш точним, задля відсіювання сторінок, які використовують методи пошукової оптимізації сайтів, які заборонені компанією Google та її пошуковим роботом Googlebot, котрий аналізує сторінки.

Під час дослідження роботи наведені результати: аналіз існуючих методів вирішення задачі, аналіз способів вдосконалення існуючого методу, удосконалення методу задля отримання результатів, аналіз отриманих результатів [1-3].

## **ABSTRACT**

The explanatory note to the master's qualification thesis contains: 69 pages, 4 chapters, 12 figures, 7 tables, 33 sources.

**LOG FILES, METHODS OF PROCESS MINING, SEARCH OPTIMIZATION, SEARCH ROBOT, VISUAL MODELS OF PROCESSES, WEB BROWSER**

During the execution of this master's work, the methods of Process Mining are described, namely the Performance Mining method, which will be used to improve the method of search promotion of sites, namely the analysis of the page for specific attributes. By using the new method, it will be possible to make the ranking of pages more accurate, to filter out pages that use methods of search engine optimization of sites that are prohibited by Google and its search engine Googlebot, which analyzes pages.

During the study of the work, the results are given analysis of existing methods of solving the problem, analysis of ways to improve the existing method, improvement of the method to obtain results, analysis of the obtained results [1-3].

## ЗМІСТ

Скорочення та умовні позначки .....	7
Вступ.....	8
1 Аналіз предметної області .....	9
1.1 Аналіз задач пошукового просування сайтів в мережі Інтернет	9
1.2 Аналіз процесу просування сайтів в пошуковій системі в мережі Інтернет .....	15
1.3 Дослідження методів просування сайтів в пошукових системах в мережі Інтернет .....	20
1.4 Аналіз методів Process Mining.....	27
1.5 Постановка задачі .....	29
2 Гібридний метод пошукового просування сайтів в мережі інтернет ....	30
2.1 Опис методу Performance Mining .....	30
2.2 Удосконалений гібридний метод пошукового просування сайтів на основі логів.....	39
3 Опис інформаційних технологій пошукового просування сайтів в мережі Інтернет .....	46
4 Експериментальна перевірка удосконаленого методу.....	56
4.1 Опис програмного модуля побудови моделі процесів веб-сторінки.....	56
4.2 Експериментальна перевірка вдосконаленого методу на наборі даних	59
Висновки .....	65
Перелік джерел посилання.....	66

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІС – інформаційна система;

КТЗ – комплекс технічних засобів;

ОС – операційна система;

ПЗ – програмне забезпечення;

ПК – персональний комп'ютер

ШІ – штучний інтелект;

HREF – посилання на ресурс;

HTML – мова розмітки сайтів;

HTTPS – захищене з'єднання;

ІТ – інформаційні технології;

LOG – журнал;

PM – процес видобутку;

PR – платна реклама;

SEO – пошукова оптимізація сайту.

## ВСТУП

У сучасному світі дуже гострим питанням постає просування сайтів у мережі Інтернет. Цьому є декілька пояснень, на сам перед більшість контенту, що отримує людина у сучасному світі представлена в електронному форматі та розміщена на веб-сторінках. Чим вище сторінка у пошуковій вибірці, тим вище ймовірність, що потенційний користувач скористається конкретним посиланням. Звідки витікає необхідність просування сайту на фоні інших сторінок. Але у цього процесу є суттєві недоліки, головний з них – сіра та чорна оптимізація. Вони офіційно не підтримуються пошуковими роботами, такими як Googlebot для пошукової системи Google, однак попри ризики, використання таких методів вони надають більший результат та піднімають сайт у видачі пошукової системи більше ефективно.

Актуальність роботи пов'язана з недостатнім використання можливостей ідентифікації спаму та платного піару за допомогою аналізу логів серверу, що погіршує можливості виявлення недозволених методів пошукового просування сайтів.

Об'єктом дослідження виступають процеси пошукового просування сайтів в мережі Інтернет.

Предметом дослідження виступають методи Process Mining.

Метою роботи є удосконалення існуючого методу пошукового просування сайтів за допомогою методів Process Mining. Науковою новизною буде виступати удосконалений метод пошукового просування сайтів за допомогою Googlebot на основі використання атрибутів щодо спаму та платного піару, які виділяються з використанням Process Mining.

До практичних результатів можна віднести підвищення ефективності пошукового просування сайтів з урахуванням можливості спаму та платного піару.



## 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

### 1.1 Аналіз задач пошукового просування сайтів в мережі Інтернет

Пошукове просування сайтів в мережі Інтернет – це сукупність методів, які направлені на підвищення ймовірності, що саме необхідний користувачу сайт буде знаходитись у перших строках при пошуку у браузері. Мережа Інтернет налічує в собі мільярди сторінок у відкритому доступі, серед яких може загубитися потенційно необхідна сторінка. Важливим питанням постає ранжування всієї цієї нескінченної купи веб сторінок для відображення саме потрібних. Основними критеріями такого ранжування можуть стати наступні пункти:

- мова, на якій вводиться запит у браузері;
- доцільність інформації;
- авторитетність видання;
- актуальність сторінки.

Спираючись на ці критерії та використовуючи спеціальні алгоритми пошуку та сортування сайтів отримуємо необхідний, і за частіше дуже влучний, результат із необхідними веб сторінками все серед перших п'яти у пошуковій смузі.

Для досягнення поставленої цілі ранжування сайтів використовують пошукову оптимізацію, або ж SEO. Пошукова оптимізація поділяється на дві невід'ємні частини: SEO безпосередньо на самій сторінці, та SEO за межами сторінки [4-7].

На рисунку 1.1 наведена графічна візуалізація складових кожної веб-сторінки, а також, в свою чергу, SEO.

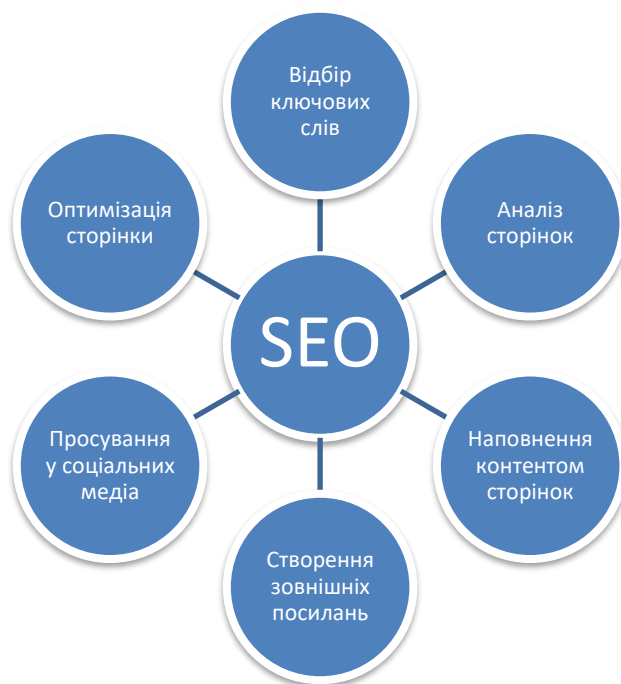


Рисунок 1.1 – Задачі пошукової оптимізації сайтів

Веб-сторінка містить купу елементів, на які спирається SEO, задля подальшої оптимізації та просування сторінки. Сама веб-сторінка, а також і SEO, містить наступні елементи, які зазначені в таблиці 1.1.

Таблиця 1.1 – Опис задач пошукової оптимізації сайтів

Назва складової SEO	Опис
Відбір ключових слів	Один з найважливіших етапів SEO, являє собою набір слів, які загально описують розділи та інформацію в них. Служить для розуміння пошуковою системою, що саме хоче отримати користувач, а з іншого боку для розуміння користувачем, чи знайшов він необхідну інформацію
Аналіз сторінок	Отримання інформації щодо частоти використання певної сторінки користувачами, тобто трафіку; розуміння, чи отримав користувач що хотів саме на обраній веб-сторінці

Кінець таблиці 1.1

1	2
Наповнення контентом сторінок	Що саме знаходиться на сторінці: текст, відео, картинки, аудіо файли тощо
Створення зовнішніх посилань	Посилання на інші авторитетні джерела, а також посилання інших веб-сторінок на задану сторінку
Просування у соціальних медіа	Використання соціальних мереж для реклами веб-сторінки
Оптимізація сторінки	Оптимізація веб-сторінки з точки зору розробки: коректна розмітка, відсутність помилок у консолі браузера, виконання вимог сучасної версії HTML

Пошукова оптимізація являє собою те, як виглядає сторінка безпосередньо для користувача: структура сайту; його дизайн; кількість медіа контенту, таких як фото та відео; мета опис сайту; актуальність наданої інформації та частота її оновлення; та багато інших аспектів[8-9].

SEO за межами сторінки – це авторитетність веб сторінки, кількість посилань на неї з інших джерел, кількість відвідувачів, тощо.

Сучасний інструментарій безкоштовних засобів дозволяє зробити пошукове просування дуже простим процесом, а також не вимагає наймати спеціаліста з просування сайтів. Саме тому SEO являє собою найбільш рентабельний метод для просування сайтів в Інтернеті.

Як зробити сайт більш привабливим для відображення в стрічці пошуку було описано вище, тепер опишемо методи, якими це ранжування здійснюється.

Такі пошукові системи як Google, Bing та інші використовують спеціальних так званих ботів (наприклад, Googlebot). Вони сканують веб сторінки в Інтернеті, перевіряють їх на вимоги коректності оформлення сторінок, доцільність інформації, надійність домену та самого сайту, та ще сотні різних критеріїв, а потім індексують сторінку, що й дозволяє їй отримати

своє місце при відображенні результатів використовуючи конкретний пошуковий сервіс. Після аналізу сайтів ботами створюються log файли, з якими власник сайту може ознайомитись, запросивши їх у свого хоста, визначити проблемні моменти та виправити їх для підвищення рейтингу сайту.

Основними проблемами частіше за все бувають:

- сторінки внутрішні не пов'язані одна з одною, та не мають, наприклад, кнопок “назад”;
- важливі сторінки знаходяться далеко від до кореня веб сторінки, що створює проблеми під час пошуку необхідної інформації;
- сторінки повертають помилки різного характеру (критичні, або ж попередження).

Важливим етапом для просування сайтів є дослідження ключових слів, яке допомагає визначити, на які ключові слова чи теми слід орієнтуватися боту, під час пошуку необхідних сторінок. Веб сторінка ніколи не оцінюватиметься за одним ключовим словом, тому потрібно використовувати теми. Більшість сторінок оцінюються за сотнями або тисячами варіантів ключових слів. Це важливо мати на увазі під час дослідження ключових слів та під час наповнення сайту інформацією[10].

Щоб визначити обсяг пошуку та конкурентоспроможність потенційних ключових слів, потрібно буде скористатися інструментом добору ключових слів.

Виходячи з описаного вище з'являється нескінченний цикл оптимізації процесу просування сайтів та аналізу цих самих веб сторінок ботами пошукових систем.

SEO є фундаментальною частиною цифрового маркетингу, оскільки щороку люди здійснюють трильйони пошукових запитів, часто з комерційною метою, щоб знайти інформацію про продукти та послуги. Пошук часто є основним джерелом цифрового трафіку для брендів і доповнює інші маркетингові канали. Краща видимість і вищий рейтинг у результатах пошуку,

ніж у конкурентів, можуть істотно вплинути на прибуток розробника веб сторінки.

Однак результати пошуку розвивалися протягом останніх кількох років, щоб надати користувачам більш прямі відповіді та інформацію, яка, швидше за все, утримає користувачів на сторінці результатів, а не спрямовує їх на інші веб-сайти.

Також можна зауважити, що такі функції, як розширені результати та панелі знань у результатах пошуку, можуть підвищити видимість і надати користувачам більше інформації про компанію розробника веб сторінки безпосередньо в результатах.

Підсумовуючи, SEO є основою цілісної маркетингової екосистеми. Коли розробник сайтів зрозуміє, чого хочуть користувачі його веб-сайту, він зможе застосувати ці знання у своїх кампаніях (платних і звичайних), на своєму веб-сайті, у своїх ресурсах у соціальних мережах тощо.

Googlebot сканує веб-сторінки за допомогою посилань. Він знаходить і читає новий і оновлений вміст і пропонує, що слід індексувати. Каталог можна розглядати як місце, де Google зберігає інформацію. Google використовує велику кількість комп'ютерів, щоб надсилати свої веб-переглядачі в усі куточки Інтернету, щоб знайти ці сторінки та побачити, що на них міститься. Googlebot – це веб-робот Google, і інші пошукові системи також мають власних пошукових роботів. Розуміння того, як працює Googlebot, має важливе значення для успішної пошукової оптимізації. Робот Googlebot використовує карти сайту та бази даних посилань, виявлених під час попередніх сканувань, щоб визначити, куди перейти під час наступного сканування. Коли роботи знаходять нові посилання на сайті, вони додають їх до списку сторінок для наступного відвідування. Якщо Googlebot знаходить зміни в посиланнях або пошкоджені посилання, він бере це до відома, щоб можна було оновити індекс. Програма визначає частоту сканування сторінок. Щоб переконатися, що Googlebot може правильно індексувати сайт, слід

перевірити сканування сайту. Якщо сайт відкритий для роботів пошукової системи, сайт періодично скануватиметься.

Крім того, якщо Googlebot виявляє зміни непрацюючих чи інших посилань, він робить примітку про оновлення в індексі Google. Тому завжди потрібно переконатися, що веб-сторінки можна сканувати, щоб Googlebot міг їх належним чином індексувати.

Google має багато різних типів сканерів Google, і кожен розроблений для багатьох способів сканування та відтворення веб-сайтів. Певні директиви або мета-команди для певних ботів повинні бути згенеровані веб-сайтом.

Є багато різних роботів. Наприклад, AdSense і Ads Bot контролюють якість реклами, а Mobile Apps Android – програми Android. Для нас це найважливіше:

- Googlebot (комп'ютерна версія);
- Googlebot (мобільна версія);
- Googlebot Video;
- Зображення Googlebot;
- Googlebot News.

Можна переглянути у своїх файлах журналу або відкрити розділ «Сканування» Google Search Console, щоб дізнатися, як часто Googlebot відвідує сайт і що він там робить. Якщо потрібно виконати розширені дії для оптимізації продуктивності сканування свого сайту, можна скористатися такими інструментами, як SEO Log File Analyzer від Kibana або Screaming Frog.

Можна використовувати файл robots.txt, щоб визначити, як Googlebot відвідує частини вашого сайту. Але будьте обережні, якщо зробите це неправильно, потрібно повністю припинити пошук Googlebot. Це видалить сайт з індексу.

## 1.2 Аналіз процесу просування сайтів в пошуковій системі в мережі Інтернет

Об'єктом дослідження виступає процес пошукового просування сайтів в мережі Інтернет.

Просування сайтів, або ж SEO – це маркетингове поняття, яке включає в себе цілий набір аспектів, рішень і технічних елементів, спрямованих на залучення більшого трафіку на сайт і просування сайту, іншими словами, підвищення його рейтингу в системах (наприклад, у пошукових системах Google, Bing, Yahoo). Передбачається, що SEO-просування виведе веб-сайт на перше місце в результатах пошуку за конкурентними ключовими словами. Раніше спеціалісти з просування сайтів дуже цінувалися, але зараз кожний розробник веб-сторінки може власноруч це зробити. Враховуючи ситуацію на сучасному ринку, створення сайту та його просування не можна розділити на два процеси. По-перше, постійно зростає кількість користувачів Інтернету (потенційна аудиторія) і з'являються нові технології (системи зв'язку, замовлення та онлайн-платежі). По-друге, водночас неймовірна швидкість онлайн-покупок і обмежений час, який клієнти мають вибирати, зробили Інтернет потужним маркетинговим інструментом і місцем для здійснення покупок. Створення веб-сайтів і SEO-просування можуть принести користь будь-якій компанії, яка орієнтується на потенційних клієнтів, які замовляють товари та послуги в Інтернеті. Щодня і щогодини неймовірна кількість людей шукає різні продукти в Інтернеті. Питання в тому, чи знайдуть вони саме ту продукцію, яка їх цікавить, на сайті конкретної компанії, або ж на сайті конкурентів, що погано вплине на прибуток та розвиток. Оптимізація веб-сайту – дуже потужна та ефективна зброя в цій нескінченній боротьбі за місце під сонцем у мережі Інтернет [26-28].

Процес просування сайтів у пошуковій системі в мережі Інтернет являє собою цикловану послідовність дій, які потрібно постійно виконувати, щоб

сторінка ставала більш авторитетною та піднімалася вище у видачі пошуковою системою. Як це працює зображено на рисунку 1.2.



Рисунок 1.2 – Процес пошукового просування сайтів

Порядок дій у процесі пошукового просування сайтів може змінюватися, але обов'язково повинен буде циклічним. Опишемо більш детально по кожному пункту:

- аналіз трафіку – етап спрямований на розуміння кількості відвідувачів веб-сторінки, а також кожної сторінки окремо. На цьому етапі також визначається степінь ранжування сторінки, тобто наскільки вона високо знаходиться у пошуковій видачі веб переглядача, а також наскільки конкуруючі веб-сторінки вище або нижче у списку;

- додавання ключових слів – зараз один з найважливіших етапів просування сайтів, тому що саме по ключовим словам пошуковий робот розуміє, що саме хоче отримати клієнт. Але ключових слів не повинно бути занадто багато, щоб вони не спаплюжили інформацію на сторінці;

- створення зовнішніх посилань – посилення на більш авторитетні сторінки, які можуть містити необхідну користувачу інформацію. В свою



чергу, інші ресурси також зможуть посилатися на обрану веб-сторінку, що призведе до підвищення рейтингу обох сторінок;

- просування у соціальних мережах дуже активно розвивається останні десять років та являє собою створення посилань на конкретні сторінки веб-сайту або ж на нього в цілому. Це призводить до збільшення трафіку, а також пошуковий робот знаходить це посилання у соціальних мережах, аналізує кількість його відвідувань та кількість підписників у акаунта, який підтримує це посилання;

- оптимізація сторінки і контенту – найскладніший етап у процесі просування сайтів, бо в ньому потрібно оптимізувати те, як виглядає сторінка для користувача, та те, як вона працює у back-end. Контент можна оптимізувати додаванням актуальної інформації (в тому числі і ключові слова), медіа-файлів та інтерактивних способів надання інформації. Оптимізація в back-end повинна оптимізувати сторінку з точки зору розмітки, стилів та скриптів.

Опишемо роботу оцінки сайтів компанією Google, а саме їх роботом Googlebot. На даний момент Google є основною пошуковою системою в світі. Хоча можна зробити акцент на інші пошукові системи, але це не дасть такого ефекту. Алгоритм, за допомогою якого Google вимірює авторитет веб-сторінки, називається PageRank. Він застосовується до набору документів із гіперпосиланнями, присвоюючи кожному документу числове значення як ідентифікатор його повноважень. Чим більше посилань на сторінку, тим вона важливіша. Пошукові системи – це повністю автоматизовані механізми, вони детально сканують всі сервери (відкриті для сканування) і збирають індекс, іншими словами, дані про елементи та місце (сторінку) розташування. Ця інформація записується в базу даних пошукової системи. Розглянемо основні особливості пошукового просування, такі як:

- точність визначає, наскільки знайдені сторінки відповідають запиту користувача. Наприклад, якщо шукати інформацію про «просування веб-сайту», отримаємо 100 веб-сторінок, з яких 50 містять точні ключові фрази, а

решта містять однакові слова. Чим більше подібність ключових слів та їх кількість – тим актуальніше буде сторінка для користувача;

- актуальність – ця важлива функція пошуку характеризується часом, який минув від онлайн-публікації даних до моменту, коли ці дані були включені в індекс пошукової системи. Основне просування в пошукових системах має «швидку базу даних», яка оновлюється щодня. Також актуальність стосується саме актуальності та підтримки статті, тобто її оновлення та посилання на актуальні джерела інформації;

- швидкість особливо важлива, якщо навантаження на сайт занадто велике, оскільки багато людей використовують двигун щосекунди. Тому і власник сайту, і відвідувач цієї веб-сторінки зацікавлені в швидкому та якісному обслуговуванні.

Таким чином, основний обов'язок пошукового просування полягає в тому, щоб надати користувачам інформацію, яку вони шукають. Ця потужна, величезна, гігантська система починає працювати, коли хтось вводить кілька слів у пошуковому рядку. Щоб потенційний користувач побачив саме необхідну йому сторінку, потрібно брати до уваги очікування цільових клієнтів і переконатися, що пошукове просування відобразить веб-сайт у списку відповідей на запит, а якщо ні, то проаналізувати проблеми та виправити їх.

Просіювання журналів серверів зробило нескінченно кращим роботу SEO. Далі буде пояснення про:

- типові проблеми зі скануванням Googlebot;
- використання журналів сервера, щоб побачити шлях сканування Googlebot;
- справжні проблеми з роботом Googlebot, який витрачає бюджет на сканування, і виправите її;
- аналізатором даних.

Це критично важливо для пошукових систем, оскільки:

- інструменти для веб-майстрів, сторонні сканери та пошукові оператори не дадуть повної інформації;
- як Googlebot поводить на вашому сайті, і це зробить вас кращим SEO.

Кожен рядок у журналі сервера представляє «попадання» на веб-сервер. Кожен сервер за своєю суттю відрізняється реєстрацією звернень, але зазвичай вони надають подібну інформацію, яка організована в поля.

Зокрема щодо SEO, потрібно переконатися, що Google сканує сторінки сайту, які потрібно сканувати, тому що необхідно, щоб вони мали рейтинг. Вже зрозуміло, що можна зробити внутрішньо, щоб допомогти сторінкам позиціонуватися в результатах пошуку, наприклад:

- переконайтеся, що сторінки внутрішні зв'язані;
- зберігайте важливі сторінки якомога ближче до кореня;
- переконайтеся, що сторінки не повертають помилок.

Виділимо конкретні проблеми Googlebot, які найчастіше зустрічаються при його використанні:

- непотрібні повзаючі бюджетні витрати;
- сторінка, яку він вважає важливою/неважливою;
- якщо є пастки для ботів;
- чи робить Google помилки 404, намагаючись скласти URL-адреси (подумайте про JavaScript).
- чи Google намагається заповнити форми.

Слід завантажити набір журналів сервера від хостингової компанії. Суть полягає в тому, щоб спробувати зафіксувати Googlebot, який відвідує сайт, за винятком того, що не знаємо, коли це станеться, тому може знадобитися кілька днів журналів або лише кілька годин.

Приклад домену має PageRank 6, DA 80 і отримує 200 000 відвідувань на день. Журнали серверів IIS складатимуть 4 ГБ на день, але оскільки сайт дуже популярний, Googlebot відвідує його принаймні раз на день [11-13].

Кожного разу, коли Googlebot заходив на сайт, він витрачав більшу частину свого часу на сканування сторінок PPC і внутрішніх сценаріїв JSON. Просто щоб дати уявлення про те, скільки часу та бюджету сканування було витрачено, перегляньте нижче:

Справжня проблема полягає в тому, що на нашому сайті були сторінки, які не були проіндексовані, і це було причиною.

Можна сканувати або відвідати сайт за допомогою агентів користувача Googlebot, і навіть гірше – можна підробити IP-адресу Googlebot. Завжди потрібно двічі перевіряти список IP-адрес та порівнювати із тим, що знаходиться у звіті журналу сервера, і використовую метод, офіційно описаний Google.

Сканування PPC сторінок: спочатку потрібно перевірити, чи ці сторінки не проіндексовано та не отримують жодного трафіку, а потім використав robots.txt, щоб заблокувати лише робота Googlebot на цих сторінках.

### 1.3 Дослідження методів просування сайтів в пошукових системах в мережі Інтернет

Визначимо методи, які використовуються під час пошукового просування сайтів, вони зазначені у таблиці 1.2.

Таблиця 1.2 – Методи SEO

Назва методу	Опис методу
White-Hat SEO	Метод, який відповідає умовам основних пошукових систем, включаючи Google. Покращує рейтинг веб-сторінки в результатах пошуку, та дає змогу розробити авторитетний з точки зору пошукового просування сайт

Кінець таблиці 1.2

1	2
Gray-Hat SEO	Більш ризиковий метод, який не призводить до блокування сторінки пошуковим роботом. Досягається за допомогою додавання великої кількості елементів, які мають значення для пошукової системи, але погіршують контент сторінки для кінцевого користувача
Black-Hat SEO	Заборонений у пошукових системах метод, який робить акцент на використанні спаму, платного піару сайту, маскуванню його за інші сторінки тощо

Тепер опишемо недоліки кожного з методів, вони будуть наведені у таблиці 1.3.

Таблиця 1.3 – Переваги та недоліки існуючих методів

Назва методу	Плюси	Мінуси
White-Hat SEO	Більший авторитет; Підтримка пошуковою системою	Мінімальна ефективність; Великі часові затрати
Gray-Hat SEO	Сторінка не може бути заблокована; Ефективність методу значно вища за White-Hat	Контент сторінки менш зрозумілий для користувача; Відсутня підтримка пошуковою системою
Black-Hat SEO	Найбільша ефективність методу; Мінімальні часові затрати	Сторінка з високою ймовірністю буде заблокована; Авторитет такої сторінки викликає сумніви у пошукової системи та користувачів

Опишемо, як компанія Google, а саме Googlebot, індексує сторінки у мережі Інтернет. Він має так звану таблицю факторів SEO, яка складається з наступних пунктів:

- типи факторів успіху пошукової системи – це фактори на сторінці та поза нею, які можуть вплинути на пошуковий рейтинг;

- фактори успіху вмісту та пошукової системи – це елементи веб-сторінки, які містять високоякісний вміст, котрий «сподобається» пошуковим системам і цільовій аудиторії сторінки, що просувається;

- архітектура сайту та фактори успіху пошукової системи – аспекти сайту, які спрощують доступ для пошукових систем і впливають на взаємодію з користувачем;

- HTML-код і фактори успіху пошукової системи – це HTML-теги та структуровані дані, які допомагають упорядкувати інформацію на веб-сторінці та допомогти пошуковим системам зрозуміти її вміст;

- довіра, авторитет, експертиза та рейтинги в пошуку – це те, як користувачі взаємодіють із сайтом, а також його репутація та авторитет допомагають пошуковим системам визначити, чи вартий він показу користувачам;

- побудова посилань і рейтинг у пошукових системах – посилання на інші сторінки, або ж посилання сторінок на сайт, що просувається, важливі та що вони говорять пошуковим системам про вміст;

- персоналізація та рейтинг пошукової системи – це специфічні для користувача елементи, такі як місцезнаходження, використання аудіо та відео, доступ до банківських даних, тобто усі види конфіденційності, які можуть впливати на результати, які бачать користувачі;

- токсини та покарання за спам у пошукових системах – це так звані «ярлики» SEO, які вішаються на сайт, котрий використовує образи, нецензурну лексику, нетолерантність, тощо. Якщо пошукова система спіймає сторінку за такою тактикою, це може призвести до покарання, наприклад

блокуванням деяких частин сайту, або ж зменшенням індексу сайту, або навіть до видалення сторінки з пошукового індексу;

– нові версії пошуку (наприклад, голосовий, розташування, пошук зображень і відео) – це нові способи для користувачів знайти те, що вони шукають. Хоча кожен із них надає нюанси можливостей для брендів, вони все одно базуються на фундаментальних принципах SEO.

Для розвитку існуючого методу просування сайтів у мережі Інтернет можна використати актуальні на цей час моменти, які або зовсім не використовуються при ранжуванні, або дуже мало, це дозволить приділяти більше уваги сторінці та користувацькому досвіду як факторам ранжирування. Можна зробити фокус на такі більш сучасні елементи UX на сторінці, як:

– зручність для мобільних пристроїв – найбільш актуальний на даний момент параметр, який велика кількість розробників веб-сторінок опускають, але у теперішньому світі все більше людей використовують мобільні пристрої, такі як смартфони або планшети, для перегляду контенту у різних кутках світу у різних умовах;

– безпечний перегляд – коли крім стандартних елементів безпеки використовується ще й безпечний перегляд, який дозволяє не повністю вивантажувати сторінку на мобільний девайс, що робить перегляд більш легким та безпечним;

– HTTPS – використання більш сучасних методів безпеки веб-сайтів, а саме створення дійсних сертифікатів, які підходять під всі умови веб-переглядача.

SEO – це широке поле, яке охоплює багато корисних методів. Давайте розглянемо деякі з найважливіших стратегій SEO, щоб полегшити роботу Googlebot. Можна починати, виконавши наведені нижче дії [29-30].

Необхідно зробити сайт видимим для пошукових систем. Можна зробити це за допомогою команди "Allow: /" у файлі robots.txt.

Не можна використовувати тег «nofollow» у внутрішніх посиланнях на вашому сайті або зведіть його до мінімуму. Ці посилання спеціально повідомляють сканерам, таким як Googlebot, не переходити за ними назад до джерела.

Важливий етап – це створення карти сайту. Карта сайту – це список усіх сторінок вашого сайту та важливої інформації про них, упорядкований таким чином, щоб робот Googlebot міг її легко зрозуміти. Якщо у вас є карта сайту, Googlebot звернеться до цього ресурсу, щоб дізнатися про сайт і знайти весь вміст.

Також необхідно використовувати Google Search Console. За допомогою цього набору інструментів можна виконувати багато життєво важливих завдань. Наприклад, можна надіслати свою карту сайту, щоб робот Googlebot міг швидше знаходити та сканувати URL-адреси. Також можна дізнатися, чи є на сторінках помилки, пов'язані зі скануванням, і отримати поради щодо їх усунення.

Чим старанніше працювати над тим, щоб зробити свій сайт зрозумілим для робота Googlebot та інших сканерів, тим більше зростуть трафік, конверсії та продажі вашого веб-сайту.

Googlebot – це пошуковий робот, який відвідує сайт. Якщо зробити технічно правильний вибір для свого сайту, відповідний робот буде приходити на сайт часто. Якщо регулярно додається новий вміст, він з'являтиметься частіше. Іноді, коли вноситься масштабні зміни на свій сайт, може знадобитися викликати цього милого маленького сканера, щоб він прийшов негайно, щоб зміни якомога швидше відобразилися в результатах пошуку.

Тепер опишемо, як використовувати пошукового робота Googlebot для сканування сторінок веб-сайту завдяки широкій індексації веб-сайту Google. Отримання допомоги від експертів SEO Analytica House може бути дуже корисним у цьому складному процесі. Завдяки багаторічному досвіду надання цілісних послуг SEO Analytica House може допомогти вашому веб-сайту досягти максимального потенціалу та орієнтуватися у світі Googlebot.



Робот Googlebot базується на високорозвиненому алгоритмі, який здатний виконувати завдання автономно та побудований на концепції всесвітньої мережі (WWW). Можна уявити всесвітню павутину як велику мережу веб-сторінок (вузлів) і з'єднань (гіперпосилань). Кожен вузол унікально ідентифікується URL-адреса, і до нього можна дістатися через цю веб-адресу. Гіперпосилання на одній сторінці ведуть на інші підсторінки або ресурси в інших доменах. Бот Google може ідентифікувати та аналізувати посилання (посилання HREF) і ресурси (посилання SRC). Алгоритми можуть визначити найефективніший і найшвидший спосіб пошукового робота Googlebot у всій мережі.

Googlebot використовує різні методи сканування. Наприклад, багато потоковий метод використовується для одночасного виконання кількох процесів сканування. Крім того, Google також використовує веб-сканери, які зосереджені на пошуку в певних областях, наприклад скануванні всесвітньої мережі за допомогою певних типів гіперпосилань.

Існують різні способи надання або приховування певної інформації від веб-сканерів. Кожен сканер можна ідентифікувати в полі заголовка HTTP "агент користувача". Для веб-сканера Google специфікацією є "Googlebot", яка походить від адреси хоста googlebot.com. Ці записи агентів користувача зберігаються у файлах журналу відповідного веб-сервера та надають детальну інформацію про те, хто надсилає запити на веб-сервер.

Можна самостійно вирішити, чи хочете заборонити Googlebot сканувати конкретний веб-сайт чи ні. Якщо необхідно виключити Googlebot зі свого веб-сайту, це можна зробити різними способами:

- директива disallow у вашому файлі robots.txt може виключити цілі каталоги вашого веб-сайту від сканування;
- використання nofollow у метатезі robots веб-сторінки повідомляє Googlebot не переходити за посиланнями на цій сторінці;

– також можна використовувати атрибут "nofollow" для окремих посилань, щоб Googlebot не переходив за цими посиланнями (у той час як усі інші посилання на цій сторінці скануються).

Розуміння того, як працює Googlebot і як на нього впливати, особливо важливо для оптимізації пошукових систем. Наприклад, можна використовувати Google Search Console, щоб повідомляти Googlebot про нові сторінки на вашому веб-сайті. Крім того, має сенс створити карти сайту та зробити їх доступними для пошукових систем. Карти сайту надають огляд URL-адрес веб-сайту та можуть прискорити сканування. Однак найголовніше – допомогти Googlebot орієнтуватися на вашому веб-сайті, щоб переконатися, що він знаходить увесь релевантний вміст і не витрачає час на перегляд нерелевантних сторінок. Більше інформації на цю тему можна знайти в нашому блозі про оптимізацію бюджету сканування.

Перехід до сторінок та їхній аналіз на основі різноманітних факторів, зокрема:

- швидкість сторінки;
- візуальна естетика;
- своєчасність і актуальність;
- читабельність і структура;
- кількість і якість цитувань;
- глибина висвітлення спорідненої теми;
- наскільки добре сторінка відповідає меті пошуку.

Якщо сторінці не вистачає будь-якої, багатьох або всіх цих областей, є велика ймовірність, що сторінка може отримати кращий результат і рейтинг.

Необхідно переглянути, скільки зворотних посилань мають сторінки рейтингу, щоб отримати уявлення про діапазон посилань, які потрібно охопити, щоб бути конкурентоспроможними. Крім того, треба переглянути кількість доменів переходу для кожного домену рейтингу – поки йде перевірка

на рівні сторінки до сторінки в SERP, то сторінки в більш авторитетних доменах виграють від цього авторитету.

Однак, якщо знайти сторінку, яка займає рейтинг відносно невідомого або нового сайту, і вона має значну кількість зворотних посилань, ймовірно, це причина її рейтингу, а отримання подібної кількості посилань дасть сторінці хороші шанси отримати рейтинг .

#### 1.4 Аналіз методів Process Mining

У цьому розділі буде описано існуючі методи Process Mining. Існує три ключових підходи до процесу видобутку, вони наведені у таблиці 1.4.

Таблиця 1.4 – Підходи Process Mining

Назва підходу	Опис підходу
Automated Process Discovery	Основна техніка, яка передбачає вилучення та візуалізацію моделей процесів із log файлів
Conformance checking	Порівнює існуючий процес з існуючим шаблоном, задля визначення відповідності
Performance Mining	Зосереджений на розширенні моделі за допомогою додаткової інформації, такої як дані про місцезнаходження, витрати, терміни тощо.

Опишемо більш детально кожний підхід:

– Automated Process Discovery – це підмножина інтелектуального аналізу процесів, яка визначає візуалізацію процесу на основі даних. Він надає інтуїтивно зрозумілий, візуальний та інтерактивний метод дослідження кожного етапу процесу з метою виявлення вузьких місць. Він використовує як машинне навчання, так і штучний інтелект, щоб відстежувати всі можливі

способи, якими може виконуватися процес, і пропонувати способи їх автоматизації. Як результат, виявлення процесів допомагає швидко й ефективно створювати робочі процеси та розгортати автоматизовані процеси;

- Conformance checking – це процес порівняння фактичного процесу або журналу подій з існуючою еталонною моделлю або цільовою моделлю. Журнал подій і модель процесу використовуються як вхідні дані, а на виході – діагностичні дані, що показують відмінності та схожість моделі з журналом;

- Performance Mining – це процес аналізу керованої даними моделі процесу для потенційної оптимізації. Парадигма процесу, керованого даними, використовує журнал прогресу. Залежно від цієї інформації вносяться зміни або вдосконалення. Якщо ці зміни не впроваджувати належним чином, вони можуть мати небажані наслідки для бізнесу. Результати інтелектуального аналізу продуктивності відображають якість аналізу та служать орієнтиром для майбутнього аналізу. Ці результати використовуються в методах виявлення процесу та перевірки відповідності.

Для використання у цій магістерській роботі буде використовуватися підхід Performance Mining, задля розширення інформації у файлах журналів, щоб збільшити ефективність пошукового просування сайтів. Опишемо найпоширеніші методи цього підходу у таблиці 1.5.

Таблиця 1.5 – Методи Performance Mining

Назва методу	Опис методу
Full-load measurement	Вимірювання системи під час повного робочого навантаження, щоб отримати всю інформацію, яку можна виміряти лише під час завантаження системи
Single-transaction measurement	Протягом цього методу система не повинна виконувати жодних інших дій, тобто повинна бути закрита для користувачів

Для використання у роботі можна використовувати обидва методи, бо кожний із них потрібен на різних етапах створення веб-сторінки. Перший метод гарно підходить, коли сторінка вже є продуктивною та використовується користувачами. А другий метод підходить, коли сторінка тільки що створена та потрібно перевірити її перший раз [15-16].

## 1.5 Постановка задачі

Задачею, що вирішується у даній магістерській роботі є удосконалення існуючих методів просування сайтів у мережі Інтернет. У попередньому розділі був описаний метод, який використовує компанія Google, для просування сайтів, а саме Googlebot, а також принципи його роботи. Існуючий метод просування сайтів у мережі Інтернет є дуже досконалим та актуальним, але технології розвиваються кожного року, що потребує модернізацію методів, що використовуються.

Існуючі методи пошукового просування сайтів, а також методи Process Mining, були описані в пунктах 1.3 та 1.4 цієї пояснювальної записка. Спираючись на необхідність удосконалення методу пошукового просування сайтів у мережі Інтернет, необхідно додати у цей процес метод Process Mining під назвою Performance Mining задля розширення можливостей файлів журналу та їхнього аналізу.

Для розв'язання проблеми з використанням log файлів потрібно виконати наступні задачі:

- проаналізувати існуючі методи вирішення задачі;
- проаналізувати способи вдосконалення існуючого методу;
- удосконалити метод задля отримання результатів;
- проаналізувати отримані результати.

## 2 ГІБРИДНИЙ МЕТОД ПОШУКОВОГО ПРОСУВАННЯ САЙТІВ В МЕРЕЖІ ІНТЕРНЕТ

### 2.1 Опис методу Performance Mining

Для пошукового просування сайтів в мережі Інтернет використовуються так звані log файли, або ж файли журналу. Файли журналу – це записи про те, хто заходив на веб-сайт і до якого вмісту вони отримували доступ. Вони містять інформацію про те, хто зробив запит на доступ до веб-сайту. Це може бути бот пошукової системи, наприклад Googlebot, або особа, яка переглядає сайт. Записи файлу журналу збираються та зберігаються веб-сервером сайту або ж хостом, і вони зазвичай зберігаються протягом певного періоду часу.

У файлах журналу можна знайти наступну інформацію:

- IP клієнта;
- мітка часу з датою та часом запиту;
- метод доступу до сайту (наприклад, GET або POST);
- запитувана URL-адреса, яка містить сторінку, до якої користувач хотів отримати доступ;
- код статусу запитуваної сторінки, який відображає успішне або невдале виконання запиту;
- агент користувача, який містить додаткову інформацію про клієнта, який надсилає запит, включно з браузером і ботом (наприклад, якщо він надходить із мобільного пристрою чи комп'ютера).

З аналізу файлів журналу можна отримати велику кількість інформації, яку в свою чергу використовую Googlebot для ранжування та показу її у пошуковій виборці. Зібравши статистику Googlebot може відобразити інформацію по конкретних критеріях:

- типи файлів;
- відповіді сервера;

- тип пристрою;
- контент, яким наповнена сторінка;
- чи була сторінка проіндексована вперше, або ж переіндексована.

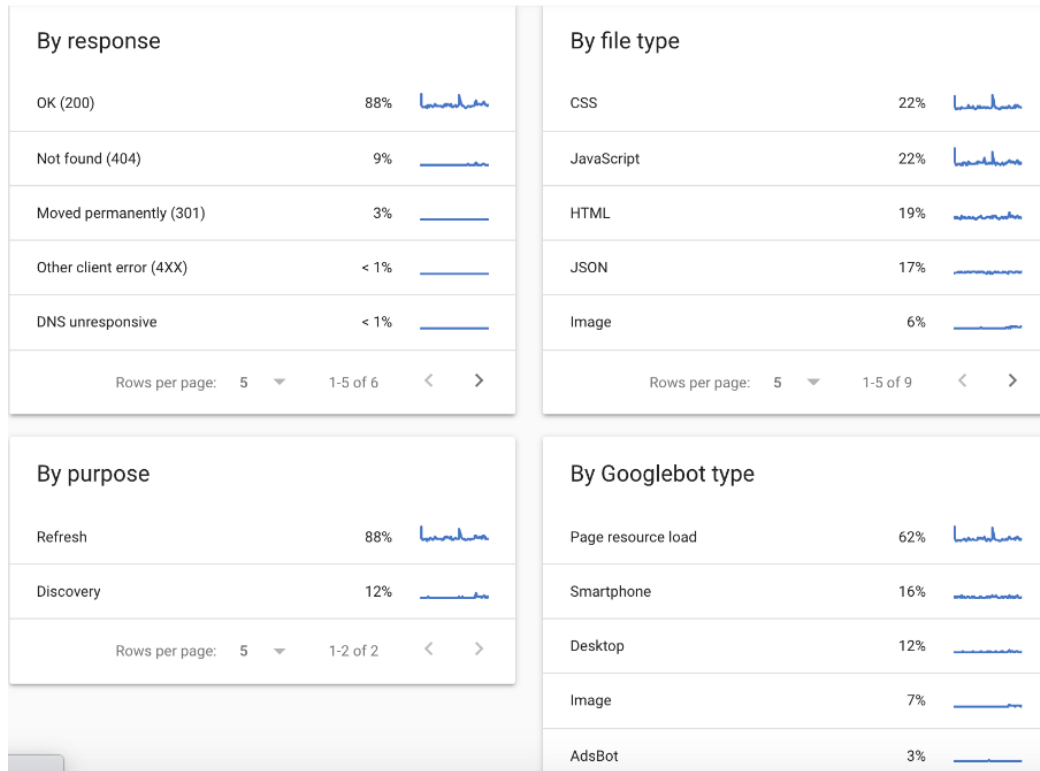


Рисунок 2.1 – Аналіз файлів журналу за допомогою Googlebot

Використовуючи файли журналу можна оптимізувати сторінку та підняти її позицію у вибірці під час пошуку. Але веб-сторінки, які були оптимізовані методами Gray-Hat SEO та Black-Hat SEO будуть мати більшу вагу. З цього постає проблема аналізу log файлів, які не враховують, що сторінка була оптимізована за допомогою чорних методів [21-24].

Для вирішення цієї проблеми гарним рішенням буде використовувати метод Process Mining під назвою Performance Mining. Опишемо принципи його роботи.

Метод за своєю структурою схожий на Automated Process Discovery, однак його суттєвий плюс – додавання необхідних параметрів аналізу. Призначений для аналізу даних про використання та користувачів, які

збираються за допомогою інструментів журналювання та зберігаються в журналах (тобто log файлів). Прототип складається з трьох скриптів, які відповідають трьом основним завданням аналізу: моніторинг продуктивності на серверах, моніторинг продуктивності програм, аналіз продуктивності програм.

На рисунку 2.2 наведено послідовність роботи методу Performance Mining.



Рисунок 2.2 – Процес роботи методу Performance Mining

Загальний опис кожного етапу методу Performance Mining зазначено на списку нижче:

- перевірка продуктивності сервера – перевірка роботи сервера та програми, яка буде використовувати обраний метод Process Mining, відбір даних, які будуть аналізуватися та їх підготовка;

- моніторинг продуктивності сервера – цей етап створює файл журналу подій, ініціює атрибути та створює записи у цей самий журнал. Більше детально цей етап буде описано далі;



- побудова моделі процесів сервера – створення моделі, які відображає роботу процесів на практиці, тобто послідовний список дій, які може виконувати система;
- аналіз продуктивності сервера – аналіз отриманого log файлу та отриманої моделі задля оптимізації та покращення деяких аспектів сервера або його переробка в цілому.

Опишемо покроково другий етап цього методу, а саме моніторинг продуктивності програми. На рисунку 2.3 буде візуально зображений цей процес.



Рисунок 2.3 – Етап «Моніторинг продуктивності програми»

Процес його роботи більш детально:

- створення порожнього журналу – створення порожнього log файлу, в який будуть додаватися записи;
- ініціалізація атрибутів – найголовніший етап методу, який полягає в створенні вибірки необхідних змінних, по яких буде проводитися аналіз;
- створення запитів – виконання конкретної дії (наприклад, запит до конкретної сторінки);

- створення часових міток – запис до файлу журналу часу, коли був створений запит;
- додавання записів до журналу – по строкове створення інформаційних даних, які відображають атрибути та значення для них;
- сформований log файл – заповнений записами журнал, який буде використовуватися для подальшого аналізу.

Що таке SEO було описано у попередніх пунктах, але визначимо це у більш широкому сенсі. За своєю суттю сенс пошукової оптимізації (SEO) полягає в збільшенні видимості вашого веб-сайту в звичайних результатах пошуку основних пошукових систем.

Щоб отримати таку видимість необхідно розуміти три основні компоненти пошукового просування сайтів:

- які типи вмісту хочуть або потребують користувачі пошукової системи та клієнти веб-сторінки;
- як пошукові системи виявляють, індексують і показують вміст на сторінках результатів пошуку;
- як правильно просувати та оптимізувати сайт, щоб розповісти про нього пошуковим системам більше.

Хоча пошукові системи та технології постійно розвиваються, є деякі базові елементи, які залишаються незмінними з перших днів SEO [17-19].

Як же саме працює SEO? SEO – це сфера, яка швидко та динамічно розвивається у сучасному світі. Іноді не розуміння цього принципу може засмучувати, особливо якщо для просування сайту був обраний застарілий метод, яка більше не працює або не дає максимум ефекту для сторінки, що просувається.

Ось чому необхідно бути добре поінформованим і постійно навчатися. Пошукові системи постійно оновлюють свої алгоритми, щоб надавати користувачам якісні результати пошуку.

Штучний інтелект постійно вдосконалює алгоритми, щоб забезпечити кращий досвід користувачів. Це означає, що сьогодні пошукова оптимізація є більш складною, ніж будь-коли. І така тенденція продовжується із року в рік, або навіть частіше.

Найчастіше просування сайтів використовують маркетологи, які повинні продовжувати навчання SEO протягом усієї своєї кар'єри, щоб не відставати від того, яка тактика працює зараз, а яку тактику потрібно виключити з плану пошукового маркетингу.

Потрібно не тільки створення посилань, а й створення будь-якого контенту та додавання кількох ключових слів, щоб покращити рейтинг веб-сторінки у звичайному пошуку та збільшити видимість бізнесу чи бренду, який просувається.

Для просування сайтів у мережі Інтернет потрібно відслідковувати та розуміти наступні принципи:

- нові тенденції розвитку веб-сторінок (наприклад, тотальний перехід на мобільні пристрої, або ж використання голосового пошуку) або ж алгоритмів їх просування;
- нові функції в продуктах і інструментах, пов'язаних із пошуком. Це можуть бути інструменти або ж самого веб-переглядача, або веб-сторінки, яка просувається;
- технологічний прогрес, до якого можна віднести, наприклад, штучне навчання;
- аудиторія веб-сторінки, тобто перелік зацікавлених користувачів, яким потрібна саме сторінка, що розробляється та товари або є послуги, яка вона пропонує.

Отже, звідси виникає питання, як розробити стратегію SEO, щоб домінувати над конкурентами в, наприклад, Google або інших пошукових системах? Зрештою, пошукова оптимізація – це не лише пошук у пошукових системах і залучення трафіку на веб-сайті. Йдеться саме про надання чудового досвіду та залучення потенційних клієнтів і прибутку.

Щоб створити позитивну взаємодію з користувачами та залучити потенційних клієнтів із пошуку, потрібно зробити більше, ніж націлити правильні ключові слова. Необхідно розуміти наміри користувача пошуку та розробляти вміст, який пропонує рішення для їхніх проблем.

Коли зрозумілі наміри користувача, можна створювати вміст, який відповідає потребам користувачів пошуку та оптимізований для того, щоб пошукові системи виявляли та індексували його.

Існує три основних типів в пошукових системах, усі спрямовані на те, щоб допомогти отримати кращу видимість у результатах пошуку для веб-сторінки, що розробляється:

- локальний SEO;
- органічний SEO;
- спонсовані результати пошуку.

Основні відмінності між цими трьома:

- що спонукає вміст відображатися в результатах пошуку;
- як швидко можна почати відображати веб-сайт в результатах пошуку;
- основні фактори, які впливають на здатність користувача досягти першої сторінки.

Для їх подальшого порівняння опишемо що саме представляє собою кожний тип:

– локальна пошукова оптимізація – це тип оптимізації пошукових систем, який спрямований на те, щоб інформація про компанію розробника веб-сторінки відображалася в результатах на карті Google якнайближче до верху пошукової стрічки;

– органічне SEO – це процес оптимізації пошукової системи, завдяки якому вміст веб-сайту стає помітнішим у неоплачуваних нелокальних результатах;

– спонсовані результати пошуку – це форма онлайн-реклами, яка розроблена Google. Рекламодавці роблять ставки на слова та фрази, які

спонукають до відображення їхніх оголошень у спонсорованих результатах пошуку. Рекламодавці будуть сплачувати комісію щоразу, коли користувач веб-сторінки натискає на одне з їхніх оголошень. Такий спосіб є дуже затратним та не дуже популярним, але якщо вгадати на ставках, то принесе куди більший результат, ніж класичні варіанти просування сайтів.

Є чотири способи, якими великі дані змінюють SEO. Це можна зрозуміти через такі пункти:

- вміст перетворюється на дані з експоненціальною швидкістю – це, як правило, контент – це не що інше, як опублікована інформація. Однак у зв'язку з тим, що Google стає одним із основних кураторів великих даних, інструменти аналізу даних Google постійно розглядають вміст як об'єкт, який піддається кількісному вимірюванню. Таким чином, перетворюючи вміст у дані, пошукові системи можуть легко аналізувати дані та надавати користувачам більш відповідні відповіді. Таким чином результати від Google стають більш структурованими. Детальний аналіз семантичної інформації допоміг Google розробити розширені фрагменти, пакети локального пошуку, а також інші унікальні результати, які надають більш цінні результати для шукачів;

- великі дані дають глибше розуміння SEO – Google та інші пошукові системи побоюються перекомпонувати вміст у кількісно визначені дані. Це полегшує маркетологам розуміння типу даних, які шукають користувачі. За допомогою великих даних цифрові маркетологи можуть легко відстежувати, а також аналізувати ключове слово, зворотні посилання, оптимізацію на сторінці, а також інші важливі області пошуку для оптимізації своїх зусиль;

- спеціальна аналітика стає більш точною – кінцевою метою SEO або онлайн-маркетолога є збільшення конверсії. Тому для них важливо знайти кореляцію між елементами та іншими аналітичними даними, такими як конверсії, трафік і перегляди сторінок для бізнесу, а також для конкурентів. За допомогою Bing Webmaster Tools, Google Search Console, а також інших систем аналітики цифрові маркетологи можуть перевірити, як їхній веб-контент розміщено у великій мережі даних контенту. Це допомагає вносити

відповідні коригування вмісту на основі його продуктивності. Дані Google Analytics полегшують цифровим маркетологам відстеження натискань, конверсії, часу перебування на сторінці, зворотних посилань, а також інших важливих факторів пошуку;

– дані з соціальних мереж стають важливим фактором пошуку – соціальні мережі є ще одним важливим джерелом отримання інформації про великі дані. Facebook має близько одного мільярда користувачів, тоді як Twitter має понад п'ятсот мільйонів користувачів. Крім того, у мережі розміщено близько 156 мільйонів публічних блогів, які прагнуть створити мережу на таких платформах. Оскільки Google є великою компанією, що обробляє понад 20 ПБ даних щороку, вона не може дозволити собі ігнорувати таку інформацію. Таким чином, сигнали з соціальних мереж є важливим фактором для пошуку, і пошукові системи не можуть їх ігнорувати.

Важливість аналізу даних у SEO – це інтелектуальний аналіз даних, який має першочергове значення для оптимізації пошукових систем, оскільки за допомогою інтелектуального аналізу даних бізнес може легко зрозуміти модель поведінки своїх клієнтів і націлити їх відповідно. Нижче наведено деякі моменти, які показують рівень важливості аналізу даних для бізнесу:

– допомагає краще зрозуміти клієнта – використовуючи переваги інтелектуального аналізу даних, бізнес може легко зрозуміти своїх клієнтів. Це можна зробити шляхом збору корисної інформації про клієнтів на основі їх попередньої історії пошуку та вподобань щодо покупок. Це дає гарне уявлення про потреби та бажання клієнтів. Краще зрозумівши клієнта, можна докласти правильних зусиль, щоб спонукати клієнтів купувати продукти чи послуги, що пропонує поточний сайт;

– дає більше інформації про цільову аудиторію – бізнес може легко націлитися на цільову аудиторію за допомогою аналізу даних. Інтелектуальний аналіз даних дає чітке уявлення про релевантних і цільових клієнтів. Знаючи та розуміючи симпатії та антипатії клієнтів, компанія може

краще націлюватися на них за допомогою індивідуальних методів. Це також покращує досвід користувача та рівень задоволеності клієнтів;

– допомагає збільшити продажі – процес інтелектуального аналізу даних допомагає бізнесу ідентифікувати клієнтів, яких приваблює певний продукт, і чи можуть вони придбати цей продукт чи ні. Отримавши знання про те, хто і що таке лояльність клієнтів, можна використовувати інтелектуальний аналіз даних для вилучення даних і підвищення рівня залучення клієнтів. Це також покращує маркетингові дії та допомагає залучати нових клієнтів.

Такі пошукові системи, як Google, добре приховують більшість своїх аналізів від широкого загалу, що є головною причиною того, чому багато маркетологів не усвідомлюють справжньої важливості великих даних для SEO. Використання інтелектуального аналізу даних у веб-пошуковій системі допомагає аналізувати вміст і водночас надавати результати, релевантні для користувачів. У результаті цифрові маркетологи, які зосереджені на створенні цінного контенту для користувачів, обов'язково отримають вигоду від впливу аналізу даних на SEO.

## 2.2 Удосконалений гібридний метод пошукового просування сайтів на основі логів

У даній магістерській роботі буде вдосконалюватися крок аналізу сайту за допомогою пошукової роботи з першого етапу, який був описаний у попередньому підрозділі даного розділу. Причиною вдосконалення саме цього кроку можна назвати недостатню інформацію, що міститься у log файлах. При розширенні інформації в файлах журналу можна буде збільшити точність аналізу сторінок.

Дуже важливо додати перевірку спаму та платного піару на сайті, зробити це можна за допомогою удосконаленого методу Process Mining.

Спамом можна вважати будь який контент на сторінці, який призводить до нескінченного дратування користувача задля отримання від нього необхідного результату. Наприклад, використання непотрібних сповіщень; забагато реклами, яка закриваю собою увесь контент; забагато раптових переходів до інших сторінок. Платний піар витікає зі спау – це створення посилань або ж реклами іншої веб сторінки (найчастіше не пов’язаної з поточною) задля підвищення її відвідувань та створення зовнішніх посилань, які підвищують авторитетність та ранг сторінки у пошуковій системі при ранжуванні сторінок у мережі Інтернет.

Описавши метод Process Mining у попередньому етапі можна зробити висновок, що його буде добре пов’язати з пошуковим просуванням файлів, яке теж використовує log файли. Вдосконаливши метод пошукового просування сайтів методом Performance Mining можна досягти результату при перевірці спау та платного піару сторінки.

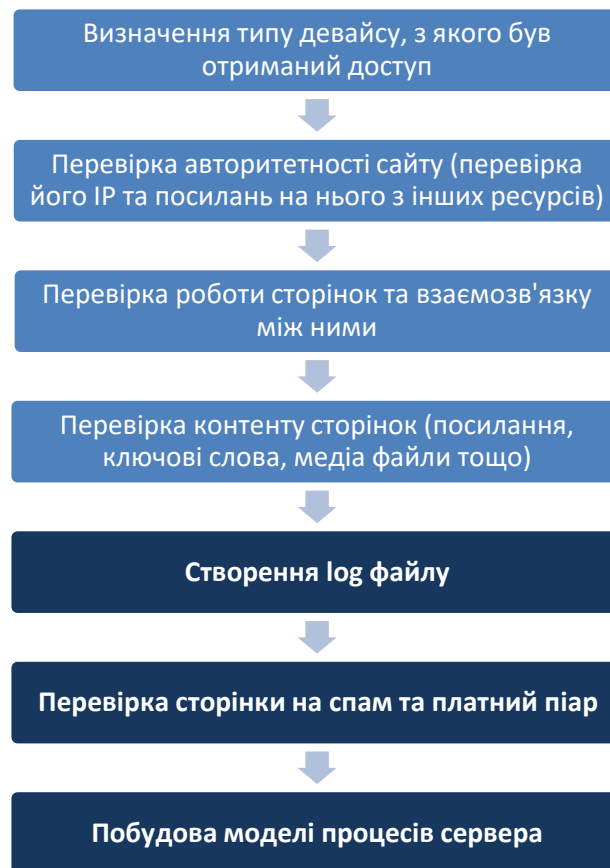


Рисунок 2.4 – Процес аналізу сайту з використанням методу Process Mining



Як видно на картинці вище, метод Performance Mining добре «накладається» на процес аналізу сайту. Цей метод дозволяє вибирати атрибути, які потрібно використовувати при аналізі, тому використовуючи стандартний набір, описаний у пункті 2.1, та додавши атрибути, які позначають спам та платний піар – можна розширити можливості просування сайтів.

Перевіряти сайт на спам можна за допомогою додавання в log файли інформації, що була зроблена переадресація на інший ресурс, який ніяк не пов'язаний з поточним. В кінець кожного запису можна додати булеву змінну, яка буде визначати, чи був перехід на інший ресурс з цієї сторінки або ж не було. Також, задля перевірки платного піару можна використовувати вже існуючий аналіз ключових слів ресурсу, на який була зроблена переадресація, з ціллю порівняння контенту поточної сторінки та нової. Це також можна позначати як булеву зміну у записах log файлів. Якщо сайт хоче отримати доступ до місцезнаходження, або надсилати сповіщення, але його ключові слова не містять ніякого зв'язку із чим – це буде означати спам, який можна також додавати у файли журналів, задля розуміння пошуковою системою потенційно небезпечного сайту [20-21].

Для формування записів у log файлі можна використати математичну формулу (2.1):

$$\sigma_i = \langle e_1, \dots, e_m \rangle, \quad (2.1)$$

де  $\sigma_i$  – це запис у файлі журналу;

$e_m$  – набір атрибутів, за якими виконується перевірка.

На формулі 2.2 зображений математичний процес, за яким в кожний запас файлу журналу додаються активні пов'язані з ними.

$$\{\sigma_i | 1 \leq i \leq \Pi_{act}(\sigma_i)\}, \quad (2.2)$$

де  $\sigma_i$  – це запис у файлі журналу;

$i$  – номер запису;

$\Pi_{act}$  – кількість активностей, за якими перевіряється сторінка.

На формулі 2.3 наведений приклад отримання часової мітки для запису у файлі журналу.

$$ct(\gamma_k) = \Pi_{ts}(\gamma_k^{end}) - \Pi_{ts}(\gamma_k^{start}), \quad (2.3)$$

де  $ct(\gamma_k)$  – це часова мітка за конкретний цикл часу  $\gamma_k$ ;

$\Pi_{ts}(\gamma_k^{start})$  – це час початку активності;

$\Pi_{ts}(\gamma_k^{end})$  – кінцевий час активності.

Заробити першу сторінку в рейтингу складно – щоб створити оптимізовані сторінки, які можуть створювати рейтинг, потрібна комплексна стратегія SEO, яка включає:

- контент-аудити;
- дослідження ключових слів;
- конкурентний аналіз;
- технічні SEO аудити;
- прогнозування та прогнозування;
- дослідження ніші та аудиторії;
- ідея та створення контенту;
- знання та розуміння історії вашого веб-сайту.

Контент, сканований Googlebot-Mobile, використовуватися в основному для покращення взаємодії з користувачем під час мобільного пошуку. Наприклад, веб-сканер може виявити вміст, спеціально оптимізований для перегляду на смартфонах, а також перенаправлення для смартфонів при відкритті цієї сторінки.

Однією з основних функцій Googlebot-Mobile, є виявлення сигналів перенаправлення для сторінок, оптимізованих для смартфонів. Коли

виявляється URL-адреса в результатах пошуку, яка спрямовує користувачів смартфонів на інший URL з оптимізованим вмістом, це сильно підіймає індекс сайту набагато у гору. Також ця функція підміняє URL у результатах пошуку саме на той, який відкриється у користувача. Це усуває додаткову затримку, яку вводить переспрямування, що призводить до економії в середньому до однієї секунди під час відвідування цільової сторінки для таких результатів пошуку.

Зараз кількість користувачів смартфонів стрімко зростає, все більше і більше веб-сайтів оптимізує свої сторінки для перегляду на смартфонах. Компанія Google використовує для індексації та перевірки сайтів Googlebot, а для мобільних сайтів спеціальне відгалуження, яке називається Googlebot-Mobile. Він сканує за допомогою агента користувача смартфона веб-сторінку на предмет функцій, які зав'язані виключно на використання мобільними пристроями. Це зроблено для того, щоб збільшити охоплення вмісту смартфонів і забезпечити кращий пошук для користувачів смартфонів при перегляді веб-сторінок. Сторінки результатів пошукової системи (SERP) – це гори, на які намагаються піднятися як пошуковці, щоб досягти вершини (позиція номер один). Але ці гори призначені не лише для сходжень – із пошукової видачі можна отримати багато «самородків» інформації, які можуть допомогти нам під час подорожі до вершини гори.

Зараз все більше розробників веб-сторінок використовують новітню індексацію сайтів, яка спрямована виключно на функції мобільних телефонів та їх підтримку. Саме із-за актуальності та великої цікавості є сенс вдосконалити метод ранжування сайтів, які максимально роблять оптимізацію під мобільні девайси.

Індекс сайту може бути в діапазоні  $[0, 100]$ , на практиці 0 та 100 – це крайні значення, які неможливо отримати. Тобто якщо сайт дуже погано оптимізований, має погану репутацію та зовсім не використовується у лінках з інших сайтів – його значення не буде нуль, але буде максимально наближене до нього, наприклад 0,005. З розумних причин такий сайт буде знаходитись у

самому кінці списку, тобто вірогідність того, що його побачить користувач пошукової системи дорівнює майже нулю. Аналогічна ситуація із індексом 100, наскільки б гарно та ідеально сайт не був оптимізований, він ніколи не отримує максимальну оцінку, але цього буде достатньо, щоб обігнати інші сторінки у пошуковій стрічці.

Для успішного SEO потрібно багато роботи та досліджень. На щастя, більшу частину цієї інформації можна отримати з результатів пошукової видачі, на які орієнтуються, що, у свою чергу, інформуватиме вашу стратегію та допоможе приймати кращі рішення.

Три основні сфери досліджень, які можуть бути корисними для аналізу SERP:

- дослідження ключових слів;
- створення контенту;
- конкурентний аналіз.

Будь-яка надійна стратегія SEO побудована на ретельному дослідженні ключових слів. Без дослідження ключових слів просто сліпо створюються сторінки та сподіваєтесь, що Google їх ранжує.

По-перше, аналіз SERP допоможе зрозуміти намір (або, принаймні, намір, який Google сприймає) за цільовими ключовими словами чи фразами. Побачивши сторінки продуктів або інформаційний вміст, чи є сторінки типу порівняння чи списку та чи існує різноманітність сторінок, які обслуговують різні потенційні наміри?

Вивчення цих сторінок покаже, яка сторінка на сайті або ще не створена сторінка буде влучна. Наприклад, якщо результати є розширеними посібниками, не можна зробити сторінку свого продукту там рейтингом (якщо, звичайно, SERP не обслуговує кілька цілей, включаючи транзакції). Необхідно проаналізувати намір пошуку, перш ніж почати оптимізацію за ключовими словами, і немає кращого ресурсу для оцінки намірів шукача, ніж самі результати пошуку.

Також можна дізнатися багато про потенційний трафік, який можна б отримати від рейтингу в певній пошуковій видачі, переглянувши її структуру та потенціал натискань.

Звичайно, веб-сторінці необхідно займати позицію номер один (а іноді й нульову позицію), оскільки загальноприйнята думка вказує на те, що це наш найкращий шанс заробити цінні кліки. І нещодавнє дослідження SISTRIX підтвердило це, повідомивши, що позиція перша має середній рейтинг натискань (CTR) 28,5% – це значно більше, ніж позиції друга (15,7%) і третя (11%).

Але найцікавішою статистикою в рамках дослідження було те, як макет пошукової видачі може вплинути на CTR.

Деякі основні моменти дослідження включають:

- результати пошукової видачі, які містять посилання на сайти, мають підвищення CTR на 12,7%, що вище середнього;
- перша позиція в пошуковій видачі з пропонуваним фрагментом має на 5,2% нижчий CTR, ніж середній;
- позиція 1 у результатах пошуку, яка містить панель знань, спостерігає падіння CTR на 11,8%, нижче середнього;
- результати пошукової видачі з оголошеннями Google Shopping мають найгірший CTR: на 14,8% нижче середнього.

### 3 ОПИС ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПОШУКОВОГО ПРОСУВАННЯ САЙТІВ В МЕРЕЖІ ІНТЕРНЕТ

У цьому розділі буде наведений процес, як відбувається процес аналізу веб-сторінок та отримання після нього файлів журналу. Загальна схема наведена на рисунку 3.1.

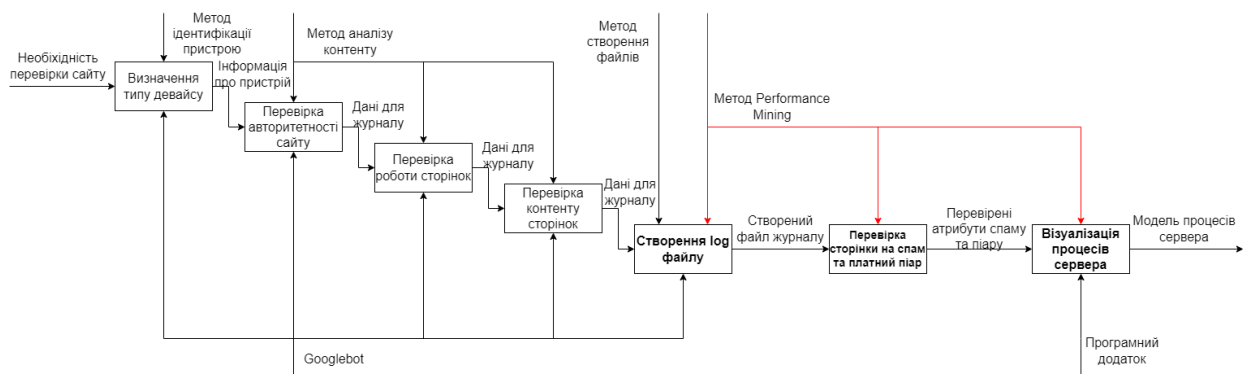


Рисунок 3.1 – Технологія пошукового просування сайтів в мережі Інтернет

Далі буде наведений список етапів з картинки вище та короткий опис стосовно кожного:

- визначення типу девайсу – на вхід подається необхідність перевірки сайту пошуковим роботом. Пошуковий робот, а саме Googlebot, виступає механізмом. Для цього використовується метод ідентифікації пристрою, який перевіряє UDID пристрою. На виході отримуємо інформацію про пристрій, з якого був отриманий доступ;

- перевірка авторитетності сайту – на вхід поступає інформація про пристрій для з'ясування того, який буде тип контенту на сторінці. Механізм – Googlebot, а методом використовується аналіз контенту на сторінці, який в цілому перевіряє увесь зміст;

- перевірка роботи сторінок – на вхід отримуємо дані з попереднього кроку, механізм та метод аналогічні. У даному випадку метод перевіряє взаємозв'язок між сторінками та їхню HTTP відповідь;
- перевірка контенту сторінок – аналогічні дані на вхід, що і у попередньому етапі, однак тут перевіряється сам контент сторінки на наявність ключових слів, помилок та тавтології;
- перевірка сторінки на спам та платний піар – на вхід та механізм ідентичний попередньому, однак методом виступає використання розширеної моделі Process Mining, яка дозволяє додати нові атрибути для перевірки;
- візуалізація процесів сервера – один із результатів роботи методу пошукової перевірки сайту, створює діаграму у вигляді ієрархічної структури усіх можливих дій на сторінці задля подальшого використання у аналізі, наприклад бізнес аналітиками;
- створення log файлу – другий результат роботи методу, який формує log файл з інформацією перевірки сторінки, а також з додатковими атрибутами, які були додані за допомогою методу Performance Mining.

Інтелектуальний аналіз даних застосовує науку про дані для виявлення, перевірки та вдосконалення робочих процесів. Поєднуючи інтелектуальний аналіз даних і аналітику процесів, організації можуть отримувати журнальні дані зі своїх інформаційних систем, щоб зрозуміти ефективність своїх процесів, виявити вузькі місця та інші сфери вдосконалення. Інтелектуальний аналіз процесів використовує керований даними підхід до оптимізації процесів, що дозволяє менеджерам залишатися об'єктивними у прийнятті рішень щодо розподілу ресурсів для існуючих процесів.

Інформаційні системи, такі як інструменти Enterprise Resource Planning (ERP) або Customer Relationship Management (CRM), забезпечують контрольний слід процесів із відповідними даними журналу. Інтелектуальний аналіз процесів використовує ці дані з ІТ-систем для створення моделі процесу або графіка процесу. Звідси розглядається наскрізний процес, а також

окреслюються його деталі та будь-які варіації. Спеціалізовані алгоритми також можуть дати розуміння основних причин відхилень від норми. Ці алгоритми та візуалізації дозволяють керівництву побачити, чи їхні процеси функціонують належним чином, а якщо ні, вони озброюють їх інформацією для обґрунтування та розподілу необхідних ресурсів для їх оптимізації. Вони також можуть відкрити можливості для включення роботизованої автоматизації процесів у процеси, прискорюючи будь-які ініціативи з автоматизації для компанії.

Інтелектуальний аналіз процесів зосереджується на різних перспективах, таких як контрольний потік, організаційний, кейс і час. Хоча більша частина роботи навколо процесного аналізу зосереджена на послідовності дій, тобто контрольний потік – інші перспективи також надають цінну інформацію для управлінських команд. Організаційні перспективи можуть виявити різні ресурси в рамках процесу, такі як окремі робочі ролі або відділи, а часова перспектива може продемонструвати вузькі місця, вимірюючи час обробки різних подій у процесі.

Певною мірою «великі дані» та інтелектуальний аналіз даних стали загальними термінами, які підсумовують нововідкриту реальність: усі цифрові види поведінки тепер є як керованими даними, так і діями, керованими даними. Практика інтелектуального аналізу даних базується на аналізі великих наборів інформації для виявлення закономірностей і якостей, які потім можна використовувати для оптимізації ефективності або нових можливостей в організації [24-27].

До типів Process Mining можна віднести:

– виявлення – виявлення процесу використовує дані журналу подій для створення моделі процесу без зовнішнього впливу. За цією класифікацією не існувало б попередніх моделей процесів, які б інформували про розробку нової моделі процесу. Цей тип процесного майнінгу є найбільш поширеним;

– відповідність – перевірка відповідності підтверджує, чи задумана модель процесу відображається на практиці. Цей тип інтелектуального аналізу



процесу порівнює опис процесу з існуючою моделлю процесу на основі даних журналу подій, виявляючи будь-які відхилення від запланованої моделі;

– покращення – цей тип аналізу процесів також називають розширенням, організаційним аналізом або аналізом продуктивності. У цьому класі інтелектуального аналізу процесу додаткова інформація використовується для вдосконалення існуючої моделі процесу. Наприклад, результати перевірки відповідності можуть допомогти виявити вузькі місця в моделі процесу, дозволяючи менеджерам оптимізувати існуючий процес.

Інтелектуальний аналіз процесів знаходиться на перетині управління бізнес-процесами (BPM) і аналізу даних. Хоча інтелектуальний аналіз процесів і аналіз даних працюють з даними, обсяг кожного набору даних відрізняється. Інтелектуальний аналіз процесів спеціально використовує дані журналу подій для створення моделей процесів, які можна використовувати для виявлення, порівняння або вдосконалення даного процесу. Сфера інтелектуального аналізу даних набагато ширша, і вона поширюється на різноманітні набори даних. Він використовується для спостереження та прогнозування поведінки, використовуючи програми для відтоку клієнтів, виявлення шахрайства та аналізу ринкового кошика, щоб назвати декілька.

Інтелектуальний аналіз процесів використовує більш керований даними підхід до BPM, яким історично керували більше вручну. Зазвичай BPM збирає дані більш неформально за допомогою семінарів та інтерв'ю, а потім використовує програмне забезпечення для документування робочого процесу як карти процесу. Оскільки дані, які інформують ці карти процесів, є більш якісними, аналіз процесів забезпечує більш кількісний підхід до проблеми процесу, детально описуючи фактичний процес через дані подій.

Збільшення продажів – не єдиний спосіб отримати дохід. Методології Six Sigma та Lean також демонструють, як скорочення операційних витрат може збільшити рентабельність інвестицій (ROI). Інтелектуальний аналіз процесів допомагає підприємствам зменшити ці витрати шляхом кількісного визначення неефективності їхніх операційних моделей, що дозволяє

керівникам приймати об'єктивні рішення щодо розподілу ресурсів. Виявлення цих вузьких місць може не тільки зменшити витрати та прискорити вдосконалення процесів, але також може сприяти розвитку інновацій, якості та кращому утриманню клієнтів. Однак, оскільки процес видобутку все ще є відносно новою дисципліною, їй все ще потрібно подолати деякі перешкоди.

Картографування вашого маршруту в Google, розміщення публікацій у Facebook, замовлення в Seamless, трансляція Netflix – усі ці дії запускають нові потоки даних, які їхні платформи збирають, аналізують і використовують для прогнозування вашого наступного кроку, тобто Інтернет, ймовірно, може передбачити наступні кроки.

Програми аналізу даних можуть включати:

- виявлення моделей марнотратства, які призводять до заходів скорочення витрат;
- прогнозування купівельної поведінки та стимулювання продажів за допомогою відповідних пропозицій щодо продуктів;
- визначення більш повільних періодів для правильного планування часу простою;
- розкриття нових можливостей у сегменті клієнтів.

Методи інтелектуального аналізу даних, які колись були прерогативою IT-спеціалістів, кількісних спеціалістів або аналітиків моделювання ризиків, тепер використовуються майже в усіх галузях чи професіях, які мають доступ до великих наборів даних. Подібно до золотошукача під час золотої лихоманки, ваше завдання полягає в тому, щоб пробиратися крізь потоки інформації в пошуках невеликого шматка даних, які справді можуть допомогти.

Будучи новатором у створенні великих баз даних, Amazon започаткував спосіб інтеграції збору великих даних і аналітики у свої операції та ДНК, пропонуючи масштабовані рішення для сховищ даних, аналізу натискань, виявлення шахрайства, механізмів рекомендацій, аналізу на основі подій тощо. Їхня інновація проклала бізнесам шлях до перевірки власних рівнів

доступу до даних і маркетингових можливостей, які надає доступ. Сучасним споживачам не тільки подобається, коли їхня поведінка в Інтернеті документується, але й прямо очікують, що компанії, з якими вони взаємодіють, спрощують цю взаємодію за допомогою аналізу даних. Ця здатність передбачити тепер є дуже важливою для багатьох організацій, які стикаються зі споживачами, і якість аналізу даних є ключовим компонентом для підтримки конкурентної переваги в галузі.

З такою великою кількістю інформації в розпорядженні компаній немає виправдання для прийняття рішень на основі інтуїції чи звички. Внутрішні зацікавлені сторони тепер повинні об'єднатися, щоб не лише виявити шаблони даних, але й відстояти свій шлях через бюрократичні перешкоди до статусу дієвого. Організації повинні використовувати своє краще розуміння споживачів, щоб стимулювати обслуговування клієнтів, задоволеність продуктом, успішний маркетинг і, зрештою, зростання.

Пошукові системи є найкращою організацією, орієнтованою на споживачів, оскільки користувачі диктували еволюцію бізнес-моделі з моменту появи онлайн-пошуку. Місія Google, Bing та їхніх партнерів полягає в тому, щоб надавати релевантні відповіді своїм користувачам, але, як і будь-який інший бізнес, вони повинні підтримувати прибуткову модель, щоб продовжувати роботу, яка в їхньому випадку залежить від залучення трафіку до найрелевантнішої інформації можливий у той самий момент, коли це потрібно користувачам для прийняття рішення. Google називає цю точку нульовим моментом істини (ZMOT). Не дивно, що ця бізнес-модель приховано впливає на те, як цифрові маркетологи підходять до оптимізації пошукових систем, а також на те, як аналізуємо дані, отримані з аналітичних платформ.

SEO-діяльність інтелектуального аналізу даних можна підсумувати як аналіз великих наборів даних з метою виявлення нових моделей трафіку та розкриття нішових можливостей. Ці нішові тенденції потім використовуються для кращого просування послуги чи продукту сегменту користувачів.

Розумна та дієва тактика SEO значною мірою покладається на інтелектуальний аналіз даних, який передбачає:

- отримання даних із Google Analytics, Omniture, Web trends та інших інструментів;
- виявлення відхилень у трафіку, поведінці чи моделях переходів;
- розуміння того, що ці відхилення означають для клієнтів і їхніх бізнес-цілей.

Аномалії, які необхідно шукати, включають джерела трафіку, прості та довгі ключові слова, які спрямовують людей на сайт, і тенденції трафіку з часом. Наприклад, річне зростання, сезонність і те, як усі ці фактори пов'язані з джерелами трафіку.

Після того як були оприлюднені основні тенденції на основі великих наборів даних, потрібно змінити свою стратегію SEO, щоб розповісти правильну історію на основі отриманих даних. Якісний аналіз даних може відкрити безліч можливостей для оповідання, але мати всі ці можливості не завжди добре.

Щоб переконатися, що процес зосереджений на правильній історії потрібно:

- бути достатньо конкретними у націлюванні пошукових термінів;
- бути чесним та не розповідати неправдиві або оманливі історії;
- мати на увазі кінцеву ціль або тип метрики для аналізу даних.

Візьміть під контроль процес і націлюйтеся лише на шаблони, коли вони справді існують, замість того, щоб шукати сумнівні зв'язки просто заради цього.

Щоб налаштувати сторінку на успіх потрібно встановити ключові показники ефективності, щоб порівняти свою ефективність із цілями, які важливі для клієнтів і залишаються актуальними для сфери органічних придбань. Потім потрібно переконатися, що постійно відстежуються прогресивна стратегія, коли здається, що вона не підходить.

Уникайте аналізу за двотижневий період або навіть аналізу за місяцями під час вивчення та звітності в Google Analytics. Якщо не потрібно виміряти короткостроковий вплив зміни на сторінці або оцінити, чи впливає сезонність, завжди потрібно дивитися на ширшу картину, а отже, на ширший часовий проміжок. Саме тоді дані стають достатньо великими, щоб бути значущими та корисними.

Далі буде описано етапи розробки методу, а також відповідальних осіб та засоби, які будуть використовуватися під час виконання кожного з них:

- аналіз існуючих методів – на вхід отримуємо необхідність вдосконалення, аналізом буде займатися аналітик, пункт являє собою поставлену задачу для вдосконалення методу просування сайтів, детальніше ця інформація була надана у розділах 1 та 2 цієї роботи. Основною необхідністю вдосконалення методу пошукового просування сайтів є більший акцент на оптимізацію сайтів для мобільних телефонів. На виході буде поставлена задача на вдосконалення існуючого методу, яка буде передана розробникам;

- вдосконалення існуючого методу – основний робочий процес, містить в себе розробку елементів для вдосконалення існуючого методу. На виході отримаємо вдосконалений метод;

- тестування методу – процес включає в себе тестування вдосконаленого методу відділом тестування на наявність багів. Процес на виході має два варіанти: виправлення помилок та імплементація методу;

- виправлення помилок – якщо були знайдені проблеми під час тестування продукту, цей етап виконує розробник, якщо проблеми таки були знайдені. Після виконання цього етапу метод повертається на повторне тестування;

- імплементація методу – фінальний етап після етапу розробки та тестування вдосконаленого методу, являє собою готовий до використання метод, який можна вбудовувати у вже існуючий Googlebot Mobile задля

використання звичайними користувачами мобільних веб-переглядачів на усіх актуальних платформах.

Стислий перелік інструментальних засобів за допомогою яких можна програмно реалізувати удосконалення методу (більше детально цей список буде описаний у розділі 4.1):

- середа розробки – IntelliJ IDEA Community version та Python IDE версії не нижче 3.11;
- веб-переглядач – Google Chrome актуальної версії;
- програма перегляду log та txt файлів – Notepad;
- мобільна операційна система – Android та iOS;
- мобільний веб переглядач – Google Chrome Mobile.

На основі вимог реалізувати технологічний процес планування IT проекту як складання опису змісту системи за кожною конкретною вимогою:

- визначати проектні цілі та обмеження – ціллю проекту є удосконалення існуючих методів пошукових систем;
- встановлювати декомпозицію робіт, засновану на системній архітектурі, що розвивається – необхідно розбити проект на етапи, а саме етап планування, розробки та тестування;
- визначати та підтримувати графіки робіт у рамках проекту, ґрунтуючись на цілях проекту та оцінках здійсненності робіт – проект повинен бути реалізований приблизно за два робочих тижні;
- визначати витрати на проект та планувати бюджет – бюджет складатиме близько 50-60 тисяч гривень;
- складати плани щодо забезпечення якості проекту – тестуванням проекту буде займатися відділ тестування, який отримає фінанси та час на перевірку вдосконаленого методу.

Для підтримки проекту на етапі його реалізації та ведення звітності функціоналу, що розробляється та знайдених багів на етапі тестування буде використовуватися Gemini – програма, написана на .NET для управління

проектами, що включає систему відстеження помилок. Gemini може бути використана як для класичного управління проектами, так і в рамках методології Agile або Scrum. Причиною цього вибору стала його відносно невисока вартість, а також достатній для цього проекту набір функцій.

Звичайно, дослідження ключових слів корисне, лише якщо використовувати його для створення правильного вмісту.

Однією з основних частин інформації, отриманої при дослідженні SERP, є те, які типи вмісту ранжуються – і оскільки сторінка також хоче ранжуватися там, ця інформація буде корисною для створення власної сторінки.

Наприклад, якщо в SERP є виділений фрагмент, то Google хоче відповісти на запит швидко й лаконічно для шукачів – зробити це на своїй сторінці. Результати відео відображаються в пошуковій системі? Ймовірно, слід включити відео на свою сторінку, якщо потрібно отримати там рейтинг. Треба визначитись, які зображення можуть бути пов'язані з сторінкою та як вони відобразатимуться.

Також можна переглянути сторінки рейтингу, щоб отримати уявлення про те, які формати добре працюють у цьому SERP. Якщо можна визначити тенденцію у форматі, то потрібно мати уявлення про те, як слід структурувати (або реструктурувати) веб-сторінку.

Деякі пошукові видачі можуть обслуговувати кілька намірів і відображати суміш із вищевказаних типів сторінок. У цих випадках подумайте, яку мету потрібно використовувати на сторінці, і зосередитися на сторінці рейтингу, яка відповідає цій меті, щоб отримати ідеї щодо створення вмісту.

## 4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА УДОСКОНАЛЕНОГО МЕТОДУ

### 4.1 Опис програмного модуля побудови моделі процесів веб-сторінки

У цьому розділі будуть наведені приклади роботи та сформований log файл. Метод Process Mining, який був доданий у метод пошукового просування сайтів описаний в розділах 2.1 та 2.2 цієї роботи, однак опишемо більш детально, як він впливає на роботу аналізу веб-сторінок пошуковим роботом.

Даний метод дозволив додати нові атрибути, які використовуються при перевірці сторінки Googlebot. Ці параметри, а також інші стандартні, наведені на рисунку 4.1 із зазначенням частоти їхнього використання та графіків. Також побудована ієрархія дій, а саме сторінок, між якими пересувався пошуковий робот збираючи інформацію про заданий веб-сайт.

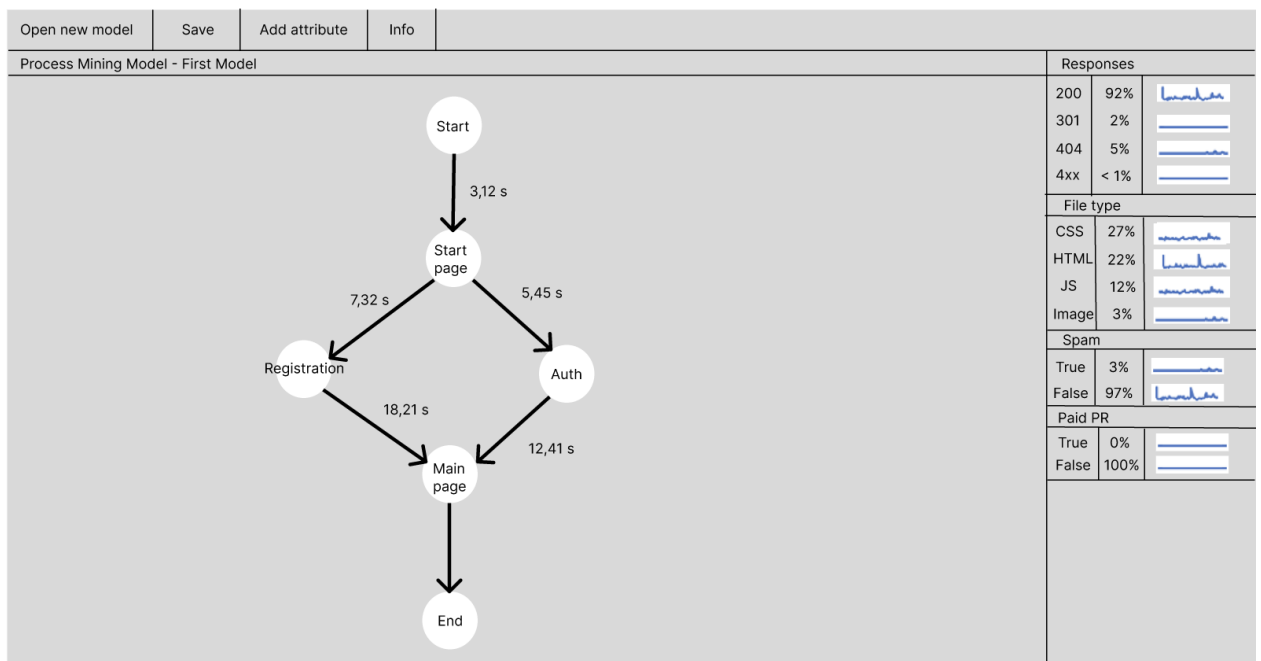
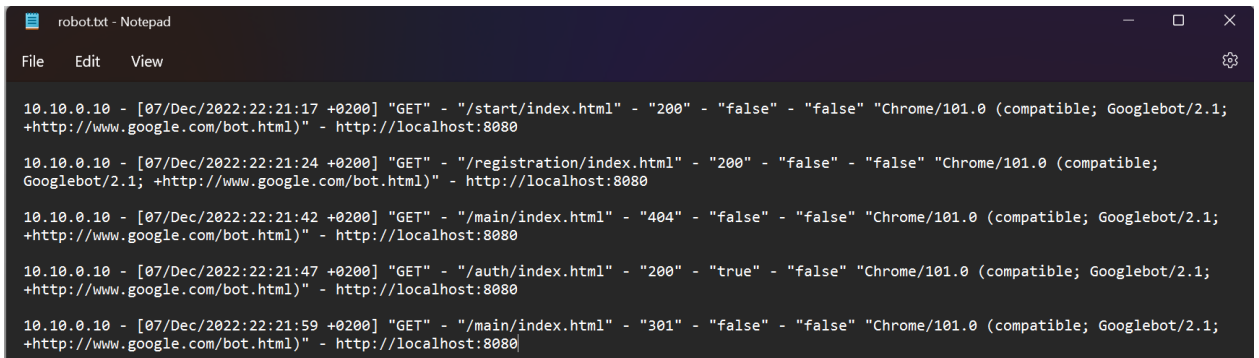


Рисунок 4.1 – Фрагмент роботи програми із побудованою моделлю процесів веб-сторінки



Далі буде наведено фрагмент log файлу, на якому зазначені стандартні атрибути, а також дів нові булеві змінні, які повертають false, якщо на сторінці не було знайдено спаму чи платного піару, або true, якщо ж таки був використаний заборонений метод просування. Приклад файлу наведено на рисунку 4.2.



```

10.10.0.10 - [07/Dec/2022:22:21:17 +0200] "GET" - "/start/index.html" - "200" - "false" - "false" "Chrome/101.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" - http://localhost:8080
10.10.0.10 - [07/Dec/2022:22:21:24 +0200] "GET" - "/registration/index.html" - "200" - "false" - "false" "Chrome/101.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" - http://localhost:8080
10.10.0.10 - [07/Dec/2022:22:21:42 +0200] "GET" - "/main/index.html" - "404" - "false" - "false" "Chrome/101.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" - http://localhost:8080
10.10.0.10 - [07/Dec/2022:22:21:47 +0200] "GET" - "/auth/index.html" - "200" - "true" - "false" "Chrome/101.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" - http://localhost:8080
10.10.0.10 - [07/Dec/2022:22:21:59 +0200] "GET" - "/main/index.html" - "301" - "false" - "false" "Chrome/101.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" - http://localhost:8080

```

Рисунок 4.2 – Приклад log файлу з двома новими булевими змінними

Які дані містить log файл було описано у розділі 2.1, в даному випадку важливі лише дві змінні, які зазначені після HTTP коду відповіді сторінки. Це є ті самі атрибути, які додані за допомогою розширеного методу Performance Mining, а також використання цього методу для побудови нового файлу журналу.

```

3 public class DiagramCreation {
4
5     public void main() throws IOException {
6
7         String[][] vals = {
8
9             ["200", "css", "10.10.0.10", "3.12", "false", "false"],
10            ["200", "java", "10.10.0.10", "7.32", "false", "false"],
11            ["404", "html", "10.10.0.10", "18.21", "false", "false"],
12            ["200", "css", "10.10.0.10", "5.45", "true", "false"],
13            ["301", "png", "10.10.0.10", "12.41", "false", "false"]];
14
15
16         var dataset = new HistogramDataset();
17         dataset.addSeries("key", vals, 50);
18
19         JFreeChart diagram = ChartFactory.createHistogram("First Model",
20             values, dataset);
21
22         ChartUtils.saveChartAsPNG(new File("diagram.png"), diagram, 450, 400);
23     }
24 }

```

Рисунок 4.3 – Фрагмент коду програми

На рисунку 4.3 був зображений фрагмент коду програми, який виконує побудову моделі з використанням вхідних даних.

Задля розв'язання поставленої задачі вдосконалення існуючого методу пошукового просування сайтів в мережі Інтернет необхідно використовувати ряд програмних засобів. Програми були обрані зважаючи на їх безкоштовність та доцільність використання, а також необхідний функціонал.

Спочатку опишемо сам Googlebot, який буде вдосконалюватися по ходу виконання даної магістерської роботи. Він написаний на мові програмування Java та Python, та використовує найактуальніші версії цих мов програмування задля досягнення максимального рівня безпеки та використання найсучасніших технологій. Googlebot – це загальна назва веб-сканера Google та загальна назва для двох різних типів сканерів: настільний сканер (просто Googlebot), який імітує користувача на комп'ютері (Googlebot Mobile, який удосконалюється у даній роботі), і мобільний сканер, який імітує користувача на мобільному пристрої.

Для удосконалення існуючого методу буде використовуватися наступний набір програмного забезпечення:

- середою розробки буде виступати IntelliJ IDEA в версії Community edition, тому що це безкоштовний варіант, функціонала якого вистачить на усі потреби. Допоміжною середою розробки буде Python IDE актуальної версії 3.11;

- для перегляду log файлів та їхнього аналізу буде використовуватися безкоштовна програма Notepad, вона більш зручна та функціональна, ніж звичайний блокнот від компанії Microsoft. Аналіз файлу журналу може допомогти точно визначити ті області, які потребують уваги та вдосконалення, і допомогти у досягненні цілі відображення у більш високій позиції під час пошуку у мережі Інтернет. Кожен файл журналу містить інформацію, яку більше ніде неможливо знайти. Це цінні записи, які показують, хто має доступ до сайту, що розробляється, і автоматично зберігаються на веб-сервері, який використовується. Із файлів можна вичерпати таку інформацію, як частота

веб-сканування сайту або ж індексація за настільним комп'ютером або мобільним методом. Другий пункт саме і є важливим у даній магістерській роботі;

- для відображення результатів пошуку на комп'ютерній версії буде використовуватися веб-переглядач Google Chrome, тому що акцент при розробці робиться саме на Googlebot;

- для перегляду результатів на мобільному пристрої потрібен девайс с актуальними на даний момент операційними системами – це Android та iOS, бажано декількох версій: актуальної, передостанньої, три версії назад та версію, яка є мінімальною з точки зору підтримки;

- також для перегляду результатів на мобільному пристрої потрібні веб-переглядачі Google Chrome Mobile для кожної з операційних систем, задля перевірки відмінностей між комп'ютерною версією та мобільними пристроями.

#### 4.2 Експериментальна перевірка вдосконаленого методу на наборі даних

Для отримання практичних результатів потрібен інстальований в систему під керуванням операційної системи Windows (8.1 або вище) веб-переглядач Google Chrome, а також аналогічний додаток на мобільних пристроях під керуванням операційних системам Android (версії 7 або вище) та iOS (версії 12 або вище). Сам веб-переглядач повинен бути версії не нижче 77.

Наведемо приклад, як виглядають індекси сторінок до імплементації методу Process Mining у пошукову оптимізацію сайтів, вони зазначені на таблиці 4.1.

Таблиця 4.1 – Результати індексації сайтів до використання удосконаленого методу

URL-адреса	Індекс сторінки
http://localhost:8080	87
http://localhost:8030	45
http://localhost:8034	92
http://localhost:8038	74
http://localhost:8052	35

Після індексації сторінок список при видачі буде виглядати наступним чином:

- http://localhost:8034;
- http://localhost:8080;
- http://localhost:8038;
- http://localhost:8030;
- http://localhost:8052.

Однак після перевірки сторінок за допомогою удосконаленого методу пошукового просування сайтів за допомогою Process Mining таблиця буде мати вигляд так, як зазначено на таблиці 4.2.

Таблиця 4.2 – Результати індексації сайтів після використання удосконаленого методу

URL-адреса	Індекс сторінки
http://localhost:8080	20
http://localhost:8030	45
http://localhost:8034	92
http://localhost:8038	74
http://localhost:8052	35

Після індексації сторінок повторно з використанням удосконаленого методу змінився індекс для сторінки <http://localhost:8080>, на ньому були знайдені ознаки спаму, тому його індекс значною мірою впав та відображається останнім. Сторінку не було зовсім видалено із пошукової видачі, бо на ній міститься корисна для користувачів інформація, але вірогідність того, що на цю сторінку зайдуть буде менша. Після таких перевірок власникам сторінок, які використовують заборонені методи просування сайтів, доведеться або відмовлятися від таких принципів, або зовсім закривати сторінку, якщо вона менш релевантна за більш інших ресурсів.

Графік змін індексації наведено на рисунку 4.4.

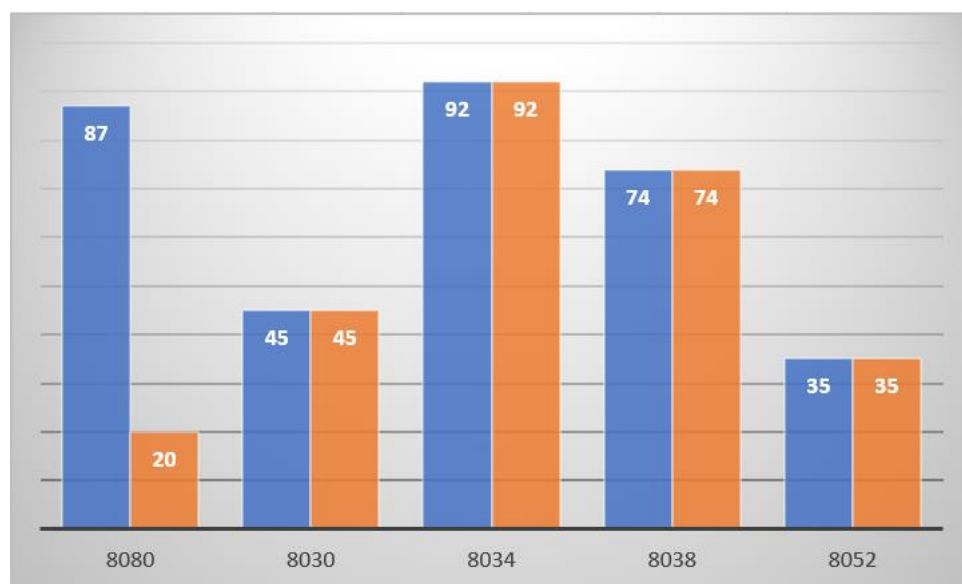


Рисунок 4.4 – Графік змін індексації

Почнімо з перегляду даних на комп'ютерній версії Google Chrome. Для цього потрібна панель «Мережа» використовується, щоб переконатися, що ресурси витягуються та завантажуються. Саме на цій панелі можна перевірити метадані, заголовки http, вміст та інше для кожної окремої URL-адреси, завантаженої під час запиту сторінки.

Отримавши з цих результатів файл robots.txt, який і є аналізом Googlebot потрібної веб-сторінки можна витягнути необхідну інформацію щодо оптимізації під мобільні пристрої.

Цей результат виник у результаті вдосконалення існуючого методу індексації сайтів, а саме додавання нових змінних у вигляді атрибутів спау та платного піару.

Перший крок – визначити релевантні пошукові запити: яку аудиторію необхідно залучити, що вони будуть шукати, їхні інтереси та інше.

У цьому тематичному дослідженні зосередимося на понятті «видобуток тексту», оскільки люди, які шукають ці терміни, швидше за все, зацікавляться продуктами та послугами Nodus Labs (особливо інструментами для обробки та кращого розуміння текстових даних).

Якщо необхідно, щоб люди, які шукають щось, пов'язане з «видобутком тексту», відвідали NodusLabs.Com і дізналися про нові інструменти візуалізації текстової мережі, які вони можуть використовувати у своїй роботі. Щоб зробити це, нам потрібно створити контент для аналізу тексту та мережевого аналізу, однак спочатку нам потрібно визначити семантичне поле, пов'язане з цим запитом, щоб краще знати, що потрібно аудиторії, що вона вже отримує під час пошуку результатів і чого не вистачає.

Наступний крок – краще зрозуміти тип вмісту, який люди отримують, коли шукають «видобуток тексту».

Для цього потрібно використовувати Infra Nodus SEO Keyword Research Tool, який імпортує перші 40 результатів пошуку за певним запитом у Google і візуалізує найважливіші теми та зв'язки між ними.

Слова представлені у вигляді вузлів, а їх співпадіння є зв'язками між ними. Після представлення тексту у вигляді графіка таким чином можна ідентифікувати спільноти (контексти) вузлів (слів), які частіше використовуються разом (теми), і найбільш впливові слова в тексті для циркуляції значень (використовувати міру центральності проміжності, яка

показує, як часто слово з'являється між різними контекстами, присутніми в текстових даних.

Як бачимо, коли користувачі шукають у Google «текстовий аналіз», вони отримують наступні тематичні кластери:

- текст – видобуток – аналітика;
- обробляти – аналізувати – великі;
- дані – неструктуровані – аналіз;
- шаблон – черга – цікавість.

Таким чином, більшість вмісту, який пропонується для цього пошукового запиту, пов'язана з обробкою великого корпусу неструктурованих текстових даних і виявленням шаблонів у цих даних.

Тепер потрібне уявлення про те, який контент отримують люди, коли шукають «видобуток тексту», тож тепер давайте подивимося, що вони насправді шукають.

Наступний крок – краще зрозуміти, яка інформація потрібна людям, коли вони шукають «видобуток тексту». У Google AdWords є інструмент планування ключових слів, який показує саме це: пов'язані пошукові запити, щоб можна було краще визначити, що дійсно потрібно нашій цільовій аудиторії. Просто введіть пошуковий запит у Планувальник ключових слів (виберіть «Ідеї ключових слів»), і Google покаже усі запити, пов'язані з оригінальним.

Потім можна натиснути посилання «Завантажити ідеї ключових слів» в інструменті «Планувальник», який створить файл CSV. Відкрийте цей файл CSV у таблицях Google, Open Office або Excel і скопіюйте дані в перший стовпець, у якому перераховано всі різні пошукові запити та словосполучення.

Потім його потрібно вставити в новий граф контексту в Infra Nodus і візуалізуємо його як мережу, щоб можна було побачити, які пошукові запити, пов'язані з «видобутком тексту», є найпомітнішими, як вони всі пов'язані між собою та теми вони утворюють.

Також можна використовувати інструмент дослідження ключових слів Infra Nodus, щоб порівняти результати пошуку (інформаційну пропозицію) з пошуковими запитами (інформаційний попит) і знайти розрив між ними.

Цей аналіз вказує на те, що люди шукають, але чого не вистачає в результатах пошуку щодо «видобутку тексту», це деякі веб-пошукові програми та реалізації алгоритмів, які можна використовувати для аналізу текстових даних.



## ВИСНОВКИ

Під час виконання магістерської роботи були розглянуті наступні питання: проаналізувати існуючі методи вирішення задачі, проаналізувати способи вдосконалення існуючого методу, удосконалити метод задля отримання результатів, проаналізувати отримані результати.

Під час аналізу існуючих методів вирішення поставленої задачі був наведений загальний опис пошукової оптимізації сайтів у мережі Інтернет та методів Process Mining, які будуть використовуватися для удосконалення існуючого методу.

Аналіз способів вдосконалення існуючого методу наведений у розділі два та описує загальну роботу методу Performance Mining, принципи його удосконалення та як це вплине на вирішення поставленої задачі пошукового просування сайтів.

Удосконалений метод наведено у пункті 4.1, а саме як він працює на реальних даних та які дозволяє отримати результати для подальшого аналізу нових атрибутів.

Проаналізовані результати наведені у пункті 4.2 та являють собою перелік експериментально доведених результатів, які були досягнені додаванням у пошукову оптимізацію методів Performance Mining із розширенням моделі, а саме атрибутів для спаму для платного піару на веб-сторінках.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи (для студентів усіх форм навчання другого (магістерського) рівня програми "Інформаційні управляючі системи та технології) / Упоряд.:Петров К.Е., Левикін В.М., Чалий С.Ф., Євланов М.В., Саєнко В.І., Міхнов Д.К., Міхнова А.В., Чала О.В. - Харків: ХНУРЕ,2021.- 30с
2. Чалий С.Ф. Метод побудови інтервальної моделі процесу вирішення завдання у складі прецеденту на основі аналізу журналу подій / С. Ф. Чалий, І. В. Левикін // Наукові праці ВНТУ, 2016, № 4. – С. 1-8;
3. Левикін В. М., Чала О.В. Виділення елементів контексту знаних бізнес-процесів на основі аналізу логів. Технологічний аудит та резерви виробництва. 2016. № 5/2 (31). С. 65-71. DOI: 10.15587/2312-8372.2016.80989;
4. Левикін В. М., Чала О.В. Підхід до виявлення аномальної поведінки процесів в системах процесного управління на основі аналізу логів. *Вісник Національного технічного університету «ХПІ». Системний аналіз, управління та інформаційні технології.* 2017. № 55(1276). С. 77-81;
5. Chalyi S., Levykin I., Biziuk A., Vovk A., Bogatov I. Development of the technology for changing the sequence of access to shared resources of business processes for process management support. *Eastern-European Journal of Eenterprise Technologies*, 2020. Vol 2, NO 3 (104). С. 22-29;
6. R. M. Haralick, «Decision Making in Context», *IEEE Transactions on pattern analysis and machine intelligence PAMI*, vol. 5, no. 4, pp. 417–428, 1983, doi: 10.1109/TPAMI.1983.4767411;
7. В. М. Левикін та О. В. Чала, «Оцінювання характеристики інформаційних об'єктів на прикладі елементів знання-ємних бізнес процесів», *Науково-виробничий журнал «Метрологія та прилади*», т. 6, м №. 62, с. 48–53, 2016;

8. E. M. Clarke, O. Grumberg, and D. A. Peled, *Model checking*. Cambridge, Mass: MIT Press, 1999;
9. R. B. Doorenbos, *Production matching for large learning systems*. Dissert. Eng, Carnegie Mellon University, 1995;
10. K. Stenzel, H. Grandy, and W. Reif, «Verification of Java Programs with Generics», in *Algebraic methodology and software technology*, vol. 5140, J. Meseguer and G. Roşu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 315–329. doi: 10.1007/978-3-540-79980-1\_24;
11. M. Bidoit and R. Hennicker, «An algebraic semantics for contract-based software components», in *Algebraic methodology and software technology*, vol. 5140, J. Meseguer and G. Roşu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 216–231. doi: 10.1007/978-3-540-79980-1\_17;
12. P. Menzies and H. Price, «Causation as a Secondary Quality», *The British journal for the philosophy of science*, vol. 44, no. 2, pp. 187–203, 1993, doi: 10.1093/bjps/44.2.187;
13. F. Eberhardt, «Introduction to the foundations of causal discovery», *International journal of data science and analytics*, vol. 3, no. 2, pp. 81–91, 2017, doi: 10.1007/s41060-016-0038-6;
14. Chalyi S., Leshchynskyi V. Method of constructing explanations for recommender systems based on the temporal dynamics of user preferences. *EUREKA: Physics and Engineering*. 2020. Vol. 3. P. 43-50;
15. E. Platanios, H. Poon, T. M. Mitchell, and E. Horvitz, «Estimating accuracy from unlabeled data: a probabilistic logic approach», in *31st conference on neural information processing systems (NIPS 2017)*, Long Beach, CA, USA. 2017, pp. 4364–4373;
16. Wade и Hulland, «Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research», *MIS Quarterly*, vol. 28, no. 1, pp. 107-142, 2004, doi: 10.2307/25148626;
17. G. Zholtkevych, H. Khalil, E. Zein, and L. Polyakova, «Category Methods for Analysis of Two Approaches to Modelling Logical Time Based on Concept of

Clocks», in Proceedings of the 14th International Conference, ICTERI, 2018, doi: 10.1007/978-3-030-13929-2\_5;

18. Н. Ф. Хайрова, и Н. В. Шаронова, *Автоматизированные информационные системы: задачи обработки информации*. Харьков: ХГУ «НУА», 2002;

19. R. Eshuis and P. Van Gorp, «Synthesizing object life cycles from business process models», Conceptual modeling. Springer, pp. 307–320, 2012;

20. A. Nigam and N. S. Caswell, «Business artifacts: An approach to operational specification», IBM Systems Journal, vol. 42, no. 3, pp. 428–445, 2003;

21. Process Mining [Электронный ресурс] – Режим доступа до ресурсу: <https://www.ibm.com/cloud/learn/process-mining>;

22. What is Process Mining [Электронный ресурс] – Режим доступа до ресурсу: <https://www.celonis.com/process-mining/what-is-process-mining/>;

23. What Is SEO / Search Engine Optimization [Электронный ресурс] – Режим доступа до ресурсу: <https://searchengineland.com/guide/what-is-seo>;

24. Search Engine Optimization (SEO) Starter Guide [Электронный ресурс] – Режим доступа до ресурсу: <https://developers.google.com/search/docs/fundamentals/seo-starter-guide>.

25. The Beginner's Guide to SEO [Электронный ресурс] – Режим доступа до ресурсу: <https://moz.com/beginners-guide-to-seo>;

26. Search Engine Optimization – Learn to Optimize for SEO [Электронный ресурс] – Режим доступа до ресурсу: <https://www.wordstream.com/seo>.

27. Performance Mining with BIC Process Mining [Электронный ресурс] – Режим доступа до ресурсу: <https://www.gbtec.com/software/bic-process-mining/performance-mining/>.

28. Introduction to process mining [Электронный ресурс] – Режим доступа до ресурсу: <https://www.qpr.com/process-mining>.

29. B.F. van Dongen. Process Mining: Fuzzy Clustering and Performance Visualization / B.F. van Dongen, A. Adriansyah;

30. Methods of performance analysis [Електронний ресурс] – Режим доступу до ресурсу: <https://www.ibm.com/docs/en/cics-ts/5.1?topic=techniques-methods-performance-analysis>.

31. How A Data Mining Approach for Search Engine Optimization Works [Електронний ресурс] – Режим доступу до ресурсу: <https://www.smartdatacollective.com/how-data-mining-approach-for-search-engine-optimization-works/>.

32. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура і правила оформлювання. . – Чинний від 22.06.2015. – Київ: ДП «УкрНДНЦ», 2016. – 31 с.

33. ДСТУ 8302:2015. Інформація та документація. Бібліографічні посилання. Загальні положення та правила складання. – Чинний від 04.03.2016. – Київ: ДП «УкрНДНЦ», 2016. – 20 с.