

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 просування сайтів в пошуковій системі з  
використанням графу знань  
(тема)

Виконав:

студент 2 курсу, групи ІУСТМ-21-1  
Вадим ОРЛОВ  
(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 професор каф. ІУС Оксана ЧАЛА  
(посада, власне ім'я, прізвище)

Допускається до захисту

Зав. кафедри  Костянтин ПЕТРОВ  
(підпис) (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 20 22 р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Орлову Вадиму Юрійовичу  
(прізвище, ім'я, по батькові)

1. Тема роботи: Дослідження методів просування сайтів в пошуковій системі з використанням графу знань  
затверджена наказом університету від 14 листопада 2022 р. № 1490Ст
2. Термін подання студентом роботи до екзаменаційної комісії 26 грудня 2022р.
3. Вихідні дані до роботи: Науково-технічна література, публікації, інформація з інтернет-ресурсів
4. Перелік питань, що потрібно опрацювати в роботі: Дослідження особливостей семантичного пошуку в мережі інтернет, аналіз графу знань, дослідження методів пошукового просування сайтів в мережі інтернет, постановка задач дослідження, аналіз графу знань для пошукового просування, порівняння підходів до аналізу графу знань, обґрунтування вибору інструментальних засобів для забезпечення аналізу просування сайту, експериментальна перевірка аналізу графу знань для пошукового просування.

## КАЛЕНДАРНИЙ ПЛАН


№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз матеріалів з теми роботи	10.10.2022	Виконано
2	Аналіз особливостей графу знань	18.10.2022 – 23.10.2022	Виконано
3	Використання аналізу графу знань при просуванні сайтів в пошуковій системі	24.10.2022 – 26.10.2022	Виконано
4	Порівняння підходів до аналізу графу знань	27.10.2022 – 15.11.2022	Виконано
5	Експериментальна перевірка отриманих результатів	17.11.2022 – 29.11.2022	Виконано
6	Підготовка пояснювальної записки та графічного матеріалу	30.11.2022 – 10.12.2022	Виконано
7	Захист перед екзаменаційною комісією	26.12.2022	Виконано

Дата видачі завдання 21 листопада 2022 р.

Студент \_\_\_\_\_

  
(підпис)

Керівник роботи \_\_\_\_\_

  
(підпис)

професор каф. ІУС Оксана ЧАЛА

(посада, власне ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка містить: 81 сторінок, 23 рисунка, 23 джерела, 1 таблиця, 1 додаток.

АЛГОРИТМ, ГРАФ ЗНАНЬ, МІКРОРОЗМІТКА, ПОШУКОВА ОПТИМІЗАЦІЯ, ПОШУКОВА СИСТЕМА, ПРОСУВАННЯ САЙТІВ.

Предметом дослідження є методи просування сайтів в пошуковій системі з використанням графу знань

Метою даної роботи – дослідження методів просування сайтів в мережі інтернет та варіанти використання графу знань для більшого розуміння між пошуковою системою та програмістом

Об'єктом дослідження є процес просування сайтів в пошуковій системі з використанням графу знань

Використано підхід до пошукового просування сайту на основі аналізу графу знань

## **ABSTRACT**

Explanatory note to revenge: 81 pages, 23 drawings, 23 links, 1 table, 1 addendum.

**ALGORITHM, WEBSITE PROMOTION, SEARCH SYSTEM, KNOWLEDGE GRAPH, SEARCH OPTIMIZATION, MICROMARKING.**

The subject of the study is the methods of promoting sites in the search engine using the knowledge graph

The purpose of this work is to study the methods of promoting sites on the Internet and options for using the knowledge graph for greater understanding between the search engine and the programmer

The object of the study is the process of promoting sites in the search engine using the knowledge graph

The approach to the search promotion of the site based on the analysis of the knowledge graph was used.

## ЗМІСТ

Реферат .....	4
Abstract .....	5
Зміст .....	6
Скорочення та умовні позначки .....	7
Вступ.....	8
1 Аналіз особливостей графу знань .....	10
1.1 Дослідження особливостей семантичного пошуку в мережі інтернет... 10	
1.2 Аналіз графу знань.....	20
1.3 Дослідження методів пошукового просування сайтів в мережі інтернет .....	27
1.4 Мета дослідження .....	40
2 Використання аналізу графу знань при просуванні сайтів в пошуковій системі .....	44
2.1 Побудова каркасу графів знань .....	44
2.2 Аналіз графу знань для пошукового просування .....	47
2.3 Аналіз графу на основі запитів .....	49
3 Порівняння підходів до аналізу графу знань .....	50
4 Експериментальна перевірка отриманих результатів .....	53
4.1 Обґрунтування вибору інструментальних засобів для забезпечення аналізу просування сайта .....	53
4.2 Експериментальна перевірка аналізу графу знань для пошукового просування.....	53
Висновки .....	61
Перелік джерел посилання .....	62
Додаток А Графічний матеріал.....	65

## **СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ**

URL – Universal Resource Identifier

RDF – Resource Description Framework

GMB – Google My Business

SEO – Search Engine Optimization

KRL – Knowledge Representation Learning

KGE – Knowledge Graph Education

## ВСТУП

З моменту свого впровадження Інтернет став важливою платформою для взаємодії бізнесу зі своїми клієнтами. Оскільки кількість користувачів Інтернету постійно зростає щодня і перевищує половину земної популяції в 2017 році, необхідність для компаній забезпечити присутність в Інтернеті стала незаперечною.

В результаті притаманного зростання Інтернет був захаращений інформацією, що перешкоджає ефективній навігації та розташуванню ключових знань. Тому пошукові системи стають все більш важливими для доступу до відповідної інформації, і багато людей використовують їх як відправну точку для своєї веб-навігації. Що свідчить про їхній успіх, пошукова система Google, яка здійснила подвиг одного мільярда унікальних щомісячних відвідувачів як перший веб-сайт у світу в 2011 році. Хоча пошуки охоплюють широкий спектр, найчастіше теми, що повторюються, стосуються продуктів і послуг, що підкреслюють важливість для.

Завдяки впровадженню графу знань у 2012 році Google встановив подальшу оптимізацію для своєї пошукової системи та забезпечив миттєвий технологічний прогрес для майбутнього пошук інформації. Ця технологія машинного навчання, визначена як база даних, сформована за допомогою пошукових запитів користувачів і додатків ключових слів, надає інформацію та рекомендації за межами звичайних пошукових термінів для надання користувачів всеохоплюючий огляд і спонукають до подальшого навчання. Хоча в основному розроблялися для вдосконалення пошуку, підприємства прийняли цей інструмент, і графи знань перетворилися на невідому поширену практику, що використовується у багатьох областях такі як соціальні медіа, надання бізнес-аналітики та голосові помічники.

У світлі постійних змін алгоритмів пошуку пошуковими системами та сучасних методів досягнення успіху, підприємствам та маркетологам

необхідно залишатися в курсі подій, щоб виділитися з мас. Таким чином, пошукова оптимізація (SEO) стала часто використовуваною маркетинговою стратегією та цінним активом, щоб відрізнити продукти від конкурентів, за рахунок підвищення рейтингу в пошуковій системі шляхом використання визначених ключових слів і інтеграції їх у веб-сайт.

Це дослідження зосереджено на пошуковій оптимізації та розширеному підключенні до графів знань. Мета дослідження полягає в тому, щоб створити глибокий огляд часто використовуваних методів у поєднанні з переважно недооціненою темою в бізнес-літературі.

# 1 АНАЛІЗ ОСОБЛИВОСТЕЙ ГРАФУ ЗНАНЬ

## 1.1 Дослідження особливостей семантичного пошуку в мережі інтернет

Останнім часом термін "Граф знань" активно використовується як бізнесом, так і дослідниками для опису комплексної технології, що базується на зберіганні даних та покращенні пошуку. Винайдений і запроваджений компанією Google Knowledge Graph у 2012 році, ця тема набула популярності, незважаючи на те, що компанія не дала чіткого визначення або пояснення її точних процедур.

На початку створення графі знань описуються як системи, що охоплюють методи отримання нових знань з Інтернету. Через поєднання фактів створюється взаємозв'язок для покращення генерації знань. Іншими словами це об'єкти та зв'язки реального світу, об'єднані у формі графа, щоб сформувані ймовірні зв'язки спеціально для того, щоб дозволити зв'язки один з одним в діапазоні різних обмежених тем. На противагу цьому, деякі вчені згадують більш технічне визначення, описуючи граф знань як "Resource Description Framework" або RDF-граф, визначений суб'єктом, предикатом та об'єктом, які виступають в якості трансляторів видобутих знань. Після ретельного аналізу подібностей та проблем у кожному визначенні, Ерлінгер та Весс об'єднали існуючі знання для побудови чіткого опису графів знань як методу генерування та включення інформації в комплексну базу знань, також названу онтологією, де програмне забезпечення робить висновки, щоб припустити нові знання з розглянутої теми.

Зберігаючи великі набори даних, отримані з існуючих даних в Інтернеті, графі знань можуть знаходити зв'язки і пропонувати подальші пошуки або результати і, таким чином, сприяти подальшому навчанню або, в кінцевому рахунку, підвищенню ділової цінності для компаній. Простіше кажучи, граф

знань пов'язує інформацію разом, щоб відобразити найбільш ймовірний результат пошуку, незважаючи на неоднозначність користувача.

Для того, щоб пояснити основу графів знань, потрібно спочатку подивитися на еволюцію пошуку в останні роки. Зі збільшенням обсягів інформації та знань, доступних в Інтернеті, стає все складніше здійснювати пошук і знаходити релевантні результати. Удосконалення практики пошуку є обов'язковим для того, щоб відповідати сучасними стандартами, тому пошук перетворився зі стаціонарного інструменту на інструмент, що включає в себе додаткову інформацію або очевидне інтелектуальне мислення при відображенні його результатів. Це і є ефект семантичного пошуку. Перші згадки про цю концепцію з'явилися в роботі Guha et al. 2003 року, в якій детально пояснюється розвиток семантичної павутини з метою включення подальших зв'язків між термінами, що виходять за рамки чистих гіперпосилань, і, таким чином, систематично закладається основа для притаманних пошуковим системам переваг у вигляді семантичного пошуку. Хоча основа їхніх досліджень була заснована на припущеннях щодо майбутнього розвитку, подальші дослідження з того часу були створені для оптимізації їхніх ідей з метою покращення пошуку інформації. З'являються повідомлення про перші успіхи систем семантичного пошуку з онтологіями, передовими системами управління знаннями, що зберігають численні дані, зокрема, для бізнес-інтрамереж.

Також згадується про відсутність спроб застосування семантичного пошуку в інтернеті та демонструють дослідження, що доводять можливість семантичного пошуку на основі онтологій в необмеженому середовищі, яким є вся всесвітня мережа. Після подальшого розвитку почалось широкомасштабне впровадження семантичного пошуку у всесвітній мережі та описує перехід від результатів окремих веб-сторінок, заснованих на активації ключових слів з ймовірністю отримання релевантної інформації, до пошукового механізму, що надає інтелектуальні результати, які ведуть до прямої відповіді.

Перехід від посилань за ключовими словами до семантичного пошуку виявився корисним для кінцевого користувача, який отримав прямий доступ до веб-сторінок, пов'язаних з пошуковим питанням. Причиною затримки запуску в Інтернеті були, зокрема, труднощі з обробкою великих обсягів даних та масштабуванням.

Для того, щоб семантичний пошук працював над трильйонами існуючих веб-сторінок, він повинен бути, по-перше, автоматичним, а по-друге, забезпечувати необхідну якість. Якщо раніше якість пошуку також залежала від пошукових запитів користувачів, то семантичний пошук зменшує неоднозначність і дозволяє пошуковій системі робити припущення щодо очікувань шукача шляхом розрахунків, заснованих на вподобаннях користувача та попередніх пошуках.

Для того, щоб продемонструвати вплив та важливість на побудову Графів знань, необхідно розглянути 3 складові семантичного пошуку, які проклали шлях до його розвитку.

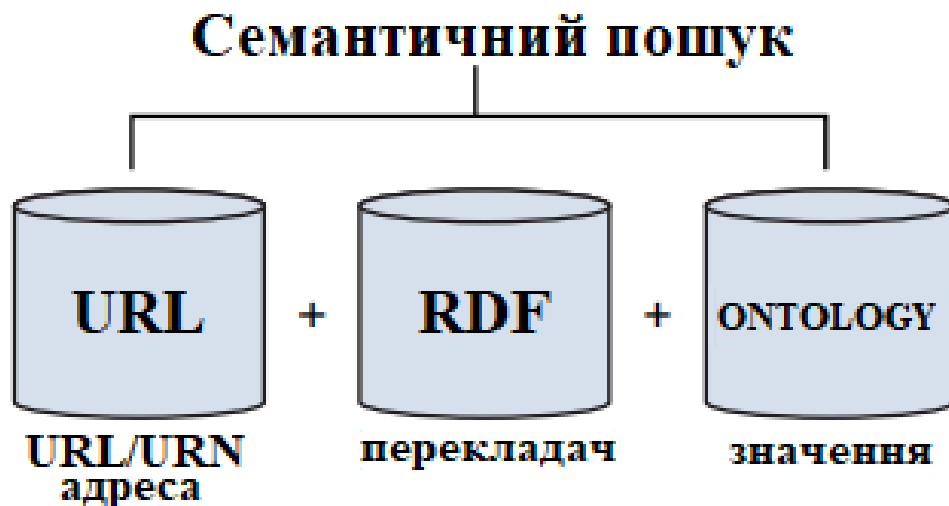


Рисунок 1.1 Семантичний пошук

URI (Universal Resource Identifier) визначає веб-посилання або ім'я суб'єкта, яке може бути встановлене як початковий набір даних для початку пошуку. Він сприяє накопиченню великої кількості первинних даних про

пошуковий запит без подальшого визначення або класифікації його достовірності. Тому його використання обмежене і потребує подальшої класифікації. Це досягається за допомогою RDF або Resource Description Framework, що переводить дані за допомогою встановленого переліку правил з однієї бази даних в іншу. Цей процес описується як надання сенсу необробленим даним, проіндексованим для машинного використання, і переведення їх у зрозумілі для людини терміни. Під час процедури транскрипції технологія може не розрізняти окремі слова, оскільки використання та значення можуть бути різними залежно від відповідного контексту, тому додатковий крок є обов'язковим для здійснення адекватного пошуку.

Для надання обґрунтованої відповіді на неоднозначні запити та забезпечення контексту, що виходить за межі пошукового терміну, використовується онтологія. Це колекція значень у базі даних, заснована на класах з певними правилами для об'єднання даних. Завдяки використанню онтологій дані можуть бути пов'язані логічними зв'язками, які подібні до людського розуміння та інтелекту. Зрештою, поєднання всіх факторів семантичного пошуку призвело до побудови графів знань як розширення існуючої технології.

Для простих запитів користувачів пошукова система може надійно знайти правильний вміст, використовуючи лише відповідність ключових слів. Запит «червоний тостер» відкриває всі продукти з «тостером» у назві чи описі та червоним в атрибуті кольору. Додайте такі синоніми, як бордовий для червоного, і ви зможете підібрати ще більше тостерів. Але швидко все стає складніше: вам доведеться самостійно додати ці синоніми, і ваш пошук також відкриє тостери. Тут на допомогу приходить семантичний пошук.

Семантичний пошук намагається застосувати наміри користувача та значення (або семантику) слів і фраз, щоб знайти потрібний вміст. Він виходить за рамки відповідності ключових слів, використовуючи інформацію, яка може бути присутньою не відразу в тексті (самі ключові слова), але тісно

пов'язана з тим, що хоче шукач. Наприклад, знайти светр за запитом «светр» або навіть «солодший» не становить труднощів при пошуку за ключовими словами, а за запитом «теплий одяг» або «як зігріти тіло взимку?» краще обслуговуються семантичним пошуком.

Як ви можете собі уявити, спроба вийти за межі інформації поверхневого рівня, вбудованої в текст, є складною спробою. Його пробували багато, і він включає в себе багато різних компонентів. Крім того, як і будь-що багатообіцяюче, семантичний пошук – це термін, який іноді використовується для пошуку, який насправді не відповідає назві. Щоб зрозуміти, чи придатний семантичний пошук для вашого бізнесу та як найкраще скористатися перевагами, це допоможе зрозуміти, як він працює та компоненти, які входять до складу семантичного пошуку. Які елементи семантичного пошуку? Семантичний пошук застосовує наміри користувача, контекст і концептуальні значення, щоб зіставити запит користувача з відповідним вмістом. Він використовує векторний пошук і машинне навчання для повернення результатів, які відповідають запиту користувача, навіть якщо немає відповідних слів.

Ці компоненти працюють разом, щоб отримати та ранжувати результати на основі значення. Однією з найбільш фундаментальних елементів, важливих для просування, є контекст.

Контекст, у якому відбувається пошук, важливий для розуміння того, що шукач намагається знайти.

Контекст може бути таким же простим, як мова (американець, який шукає «футбол», хоче щось інше, ніж британець, який шукає те саме), або набагато складнішим.

Інтелектуальна пошукова система використовуватиме контекст як на особистому рівні, так і на рівні групи. Особистісний рівень впливу на результати прийнято називати персоналізацією. Персоналізація використовуватиме інтереси цього окремого шукача, попередні пошуки та попередні взаємодії, щоб повернути вміст, який найкраще відповідає

поточному запиту. Він застосовний до всіх видів пошуку, але семантичний пошук може піти ще далі. На рівні групи пошукова система може змінити рейтинг результатів, використовуючи інформацію про те, як усі шукачі взаємодіють із результатами пошуку, наприклад, які результати натискають найчастіше, або навіть сезонність, коли певні результати популярніші за інші. Знову ж таки, це показує, як семантичний пошук може залучити інтелект до пошуку, у цьому випадку інтелект через поведінку користувача. Семантичний пошук також може використовувати контекст у тексті. Ми вже обговорювали, що синоніми корисні в усіх видах пошуку та можуть покращити пошук за ключовими словами, розширивши збіги для запитів до пов'язаного вмісту. Але ми також знаємо, що синоніми не є універсальними – іноді два слова еквівалентні в одному контексті, а не в іншому. Коли хтось шукає "футболісти", які результати є правильними? У різних державах відповідь буде іншою. Однак для такого запиту, як «футболісти тампа-бей», ймовірно, не потрібно знати, де знаходиться шукач. Додавання загального синоніма, який робив би еквівалентом футбол і футбол, призвело б до поганого досвіду, коли цей пошукач побачив футбольний клуб Tampa Bay Rowdies поруч із Роном Гронковскі.

Це приклад розуміння запиту за допомогою семантичного пошуку. Кінцева мета будь-якої пошукової системи – допомогти користувачеві успішно виконати завдання. Це завдання може полягати в тому, щоб прочитати новинні статті, купити одяг або знайти документ. Пошуковій системі потрібно з'ясувати, що користувач хоче зробити або які наміри користувача. Це можна побачити це під час пошуку на веб-сайті електронної комерції. Коли користувач вводить запит «джорданці», пошук автоматично фільтрується за категорією «Взуття». Це передбачає, що намір користувача полягає в тому, щоб знайти взуття, а не мигдаль (який буде в категорії «Їжа та закуски»).

Випереджаючи наміри користувача, пошукова система може повертати найбільш релевантні результати, а не відволікати користувача елементами, які

відповідають тексту, але не релевантно. Це може бути ще більш актуальним, якщо застосувати сортування на початку пошуку, як-от ціна від найнижчої до найвищої. Категоризація запиту та обмеження набору результатів забезпечать відображення лише відповідних результатів. Різниця між пошуком за ключовими словами та семантичним пошуком.

Хоча пошукові системи за ключовими словами також запроваджують обробку природної мови для покращення відповідності між словами – за допомогою таких методів, як використання синонімів, видалення стоп-слів, ігнорування множини – ця обробка все ще покладається на зіставлення слів зі словами.

Але семантичний пошук може повернути результати, де немає відповідного тексту, але будь-хто, хто знає домен, може побачити, що є просто хороші збіги.

Це пов'язано з великою різницею між пошуком за ключовими словами та семантичним пошуком, тобто тим, як відбувається збіг між запитом і записами. Щоб дещо спростити, пошук за ключовими словами відбувається шляхом зіставлення тексту. «Мильний» завжди відповідатиме «мильному» або «мильному» через збіг тексту.

Існує достатньо відповідних літер (або символів), щоб повідомити системі, що користувач, який шукає один предмет, захоче й інший. Таке ж зіставлення також повідомить механізму, що запит soap більш імовірно відповідає слову «суп», ніж слову «миючий засіб». Якщо власник пошукової системи заздалегідь не повідомив системі, що мило та миючий засіб є еквівалентами, у цьому випадку пошукова система «вдаватиме», що миючий засіб насправді є милом, коли вона визначає подібність. Пошукові системи, засновані на ключових словах, також можуть використовувати такі інструменти, як синоніми, альтернативи або видалення слів запиту – усі типи розширення та послаблення запиту – щоб допомогти з цим завданням пошуку інформації.

Інструменти NLP і NLU, такі як захист від друкарських помилок, токенизація та нормалізація, також покращують пошук. Незважаючи на те, що всі вони допомагають забезпечити кращі результати, вони можуть виявитися недостатніми за допомогою більш розумного зіставлення та зіставлення концепцій.

Семантичний пошук відповідає поняттям. Оскільки семантичний пошук зіставляє поняття, пошукова система більше не може визначати, чи релевантні записи, на основі того, скільки символів поділяють два слова. Або складніші запити, як-от «засіб для прання», «вивести плями з одягу» або «як видалити плями від трави з джинсової тканини?».

Аналогією з цього в реальному світі може бути клієнт, який запитує співробітника, де знаходиться «унітаз незабитий». Співробітник із суто ключовим розумінням запиту відхилив би його, якщо магазин прямо не називає свої плунжери, каналізаційні очищувачі та туалетні шнеки «прочисувачами туалетів».

Але працівник може встановити зв'язок між різними умовами та направити клієнта до потрібного проходу. Наприклад, працівник знає різні терміни або синоніми, які клієнт може використовувати для будь-якого продукту). Стислий спосіб підсумувати те, що робить семантичний пошук, це сказати, що семантичний пошук приносить підвищений інтелект, щоб зіставити поняття більше, ніж слова, завдяки використанню векторного пошуку. Завдяки цьому інтелекту семантичний пошук може виконуватися більш схожим на людину, як шукач, який шукає сукні та костюми, коли шукає модні сукні, не маючи в полі зору джинсів. \

На даний момент семантичний пошук має бути зрозумілим як потужний метод покращення якості пошуку. Таким чином, ви не повинні бути здивовані, дізнавшись, що значення семантичного пошуку застосовується все ширше. Часто такий досвід пошуку не завжди виправдовує назву. І хоча офіційного визначення семантичного пошуку немає, ми можемо сказати, що це пошук, який виходить за рамки традиційного пошуку за ключовими словами. Це

робиться шляхом включення реальних знань для визначення намірів користувача на основі значення запитів і вмісту. Це призводить до висновку, що семантичний пошук — це не просто застосування NLP і додавання синонімів до індексу. Це правда, токенізація вимагає деяких реальних знань про побудову мови, а синоніми застосовують розуміння концептуальних збігів. Однак у більшості випадків їм не вистачає штучного інтелекту, необхідного для того, щоб пошук піднявся до семантичного рівня.

Технологія Vector Search. Саме цей останній елемент робить семантичний пошук потужним і складним. Як правило, з терміном семантичний пошук існує неявне розуміння того, що існує певний рівень машинного навчання. Майже так само часто це також включає векторний пошук. Векторний пошук працює шляхом кодування деталей про елемент у вектори, а потім порівняння векторів для визначення найбільш схожих. Знову ж таки, навіть простий приклад може допомогти. Візьмемо дві фрази: «Toyota Prius» і «стейк». А тепер давайте порівняємо їх із «гібридом». Які з перших двох більше схожі? Жоден не збігався б текстово, але ви, напевно, сказали б, що «Toyota Prius» більш схожий з двох. Ви можете сказати це, оскільки знаєте, що «Prius» — це тип гібридного транспортного засобу, тому що ви бачили «Toyota Prius» у схожому контексті зі словом гібрид, наприклад «Toyota Prius — це гібрид, який варто розглянути» або «гібрид». такі транспортні засоби, як Toyota Prius». Однак ви майже впевнені, що ніколи не бачили «стейк» і «гібрид» так близько. Побудова векторів для пошуку подібності

Загалом так працює векторний пошук. Модель машинного навчання бере тисячі або мільйони прикладів з Інтернету, книг або інших джерел і використовує цю інформацію, щоб потім робити прогнози. Звичайно, для моделі неможливо порівнювати одну за одною («Чи часто бачать Toyota Prius і гібрид разом? Як щодо гібрида та біфштексу?»), тому замість цього відбувається те, що моделі кодують шаблони що він помічає про різні фрази. Це схоже на те, як ви можете поглянути на фразу і сказати: «цей позитивний» або «цей містить колір». За винятком машинного навчання, мовна модель не

працює так прозоро (саме тому мовні моделі може бути важко налагодити). Ці кодування зберігаються у векторі або довгому списку числових значень. Потім векторний пошук використовує математику, щоб обчислити, наскільки схожі різні вектори. Ще один спосіб подумати про вимірювання подібності, які виконує векторний пошук, це уявити вектори, накреслені. Це неймовірно складно, якщо ви намагаєтеся уявити вектор, розбитий на сотні вимірів. Якщо натомість уявити вектор, зображений у трьох вимірах, принцип буде той самий. При побудові ці вектори утворюють лінію, і виникає питання: які з цих ліній ближче одна до одної? Рядки для «стейк» і «яловичина» будуть ближчими, ніж рядки для «стейк» і «автомобіль», і тому вони більш схожі.

Цей принцип називається векторною, або косинусною, подібністю. Векторна схожість має багато застосувань. Він може давати рекомендації на основі раніше придбаних продуктів, знаходити найбільш схоже зображення та може визначати, які елементи найкраще відповідають семантично в порівнянні із запитом користувача.

Семантичний пошук — це потужний інструмент для пошукових програм, які вийшли на передній план із появою потужних моделей глибокого навчання та апаратного забезпечення для їх підтримки. Хоча ми торкнулися тут ряду різних поширених програм, є ще більше, які використовують векторний пошук і ШІ.

Навіть пошук зображень або вилучення метаданих із зображень може підпадати під семантичний пошук. І все ж його застосування ще рано, а його відома потужність може призвести до неправильного використання терміну. У конвеєрі семантичного пошуку багато компонентів, і важливо правильно визначити кожен із них.

При правильному виконанні семантичний пошук використовуватиме знання реального світу, особливо через машинне навчання та векторну схожість, щоб зіставити запит користувача з відповідним вмістом.

## 1.2 Аналіз графу знань

Граф знань - це структуроване представлення фактів, що складається з сутностей, зв'язків і семантичних описів. Сутності можуть бути об'єктами реального світу та абстрактними поняттями, зв'язки представляють зв'язок між сутностями та семантичними описами сутностей, а їхні зв'язки містять типи та властивості з чітко визначеним значенням. Широко використовуються графи властивостей або графи з атрибутами, у яких вузли та зв'язки мають властивості або атрибути. Термін «граф знань» є синонімом бази знань з незначною відмінністю.

Граф знань можна розглядати як граф, розглядаючи його структуру графа. Якщо він включає формальну семантику, його можна сприймати як базу знань для інтерпретації та висновку щодо фактів. Приклад графа знань проілюстровано на рисунку 1.2, а бази знань на рисунку 1.3.

Знання можуть бути виражені у фактичній трійці у формі (суб'єкт; предикат; об'єкт) у структурі опису ресурсу (RDF), наприклад, Альберт Ейнштейн, Переможець, Нобелівська премія). Він також може бути представлений як орієнтований граф із вузлами як сутностями та ребрами як відношеннями.

Останні досягнення в дослідженнях, заснованих на графах знань, зосереджені на навчанні подання знань (KRL) або вбудовуванні графів знань (KGE) шляхом відображення сутностей і відношень у низьковимірні вектори з одночасним захопленням їх семантичного значення. Конкретні завдання з отримання знань включають заповнення графа знань (KGC), потрійну класифікацію, розпізнавання сутностей і вилучення зв'язків. Моделі з урахуванням знань виграють від інтеграції різномірної інформації, багатих онтологій і семантики для представлення знань і багатомовних знань. Таким чином, багато реальних додатків, таких як системи рекомендацій і відповідей на запитання, досягли процвітання. Деякі реальні продукти, наприклад Satori

від Microsoft і Knowledge Graph від Google, продемонстрували потужну здатність надавати більш ефективні послуги.

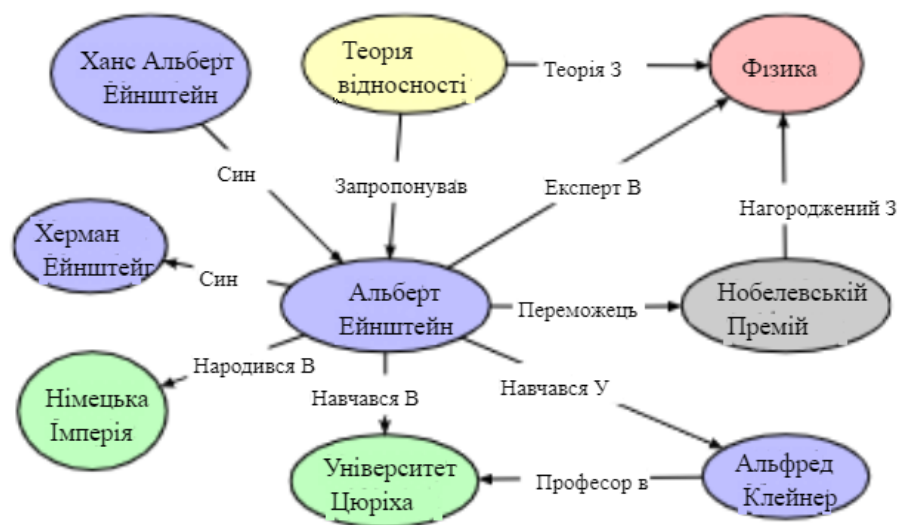


Рис.1.2 Граф Знань

- (Альберт Ейнштейн, Народився В, Німецька Імперія)
- (Альберт Ейнштейн, Син, Херман Ейнштейн)
- (Альберт Ейнштейн, Навчався В, Університет Цюріха)
- (Альберт Ейнштейн, Переможець, Нобелівська премія)
- (Альберт Ейнштейн, Експерт В, Фізика)
- (Нобелівська премія, Нагороджений З, Фізика)
- (Теорія Відносності, Теорія З, Фізика)
- (Альберт Ейнштейн, Навчався У, Альфред Клейнер)
- (Альфред Клейнер, Професор В, Університет Цюріха)
- (Теорія відносності, Запропонована, Альфред Ейнштейн)
- (Ханс Альберт Ейнштейн, Син, Альберт Ейнштейн)

Рис.1.3 База Знань

Навчання репрезентації знань є важливим дослідницьким питанням графа знань, яке відкриває шлях для багатьох завдань на отримання знань та низхідних застосувань.

KRL класифікується на чотири аспекти простору представлення, функцію оцінки, моделі кодування та допоміжну інформацію, забезпечуючи чіткий робочий процес для розробки моделі KRL. До специфічних компонентів відносяться:

- простір представлення, в якому представлені відносини і сутності;
- функція оцінки для вимірювання правдоподібності фактичних трійок;
- моделі кодування для представлення і вивчення реляційних взаємодій;
- допоміжна інформація бути включеним в методи вбудовування.

Навчання репрезентації включає в себе точковий простір, многовид, складний векторний простір, розподіл Гауса і дискретний простір. Показники підрахунку балів, як правило, поділяються на функції оцінки на основі відстані та подібності. Сучасні дослідження зосереджені на моделях кодування, включаючи лінійні білінійні моделі, факторизацію та нейронні мережі. Допоміжна інформація розглядає текстову, візуальну та типову інформацію.

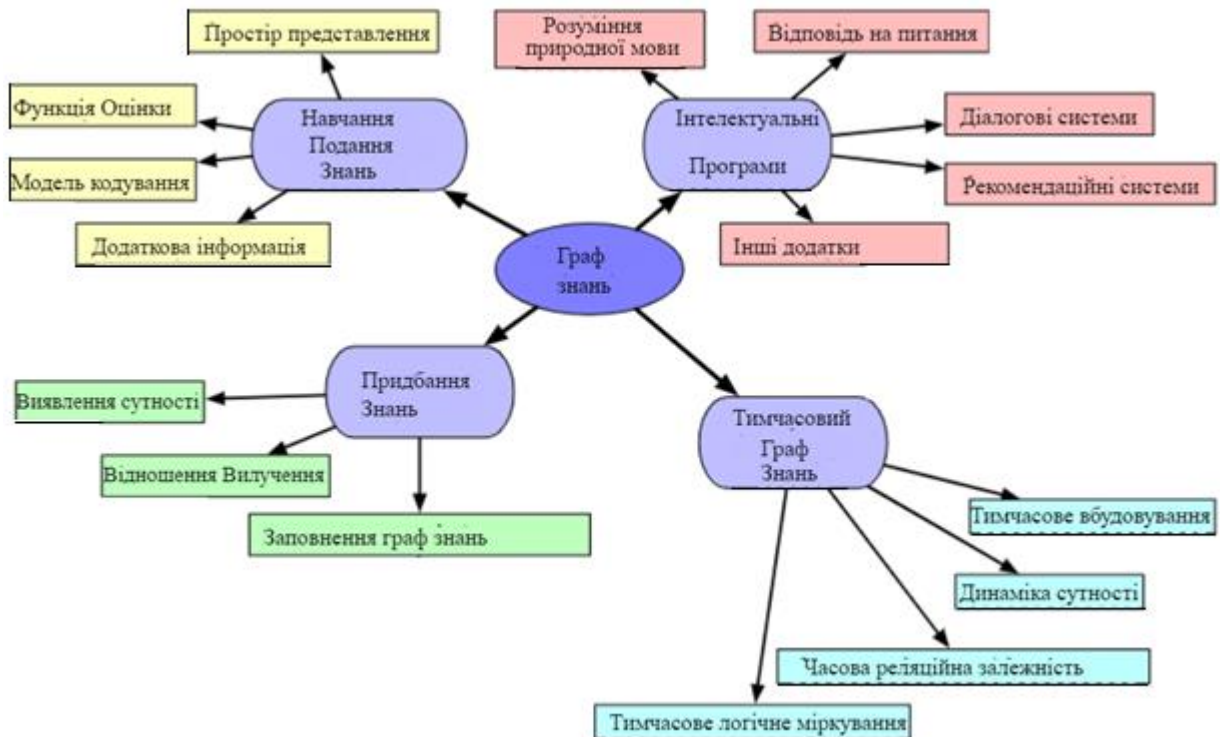


Рисунок 1.4 Категорії досліджень на графах знань

Завдання придбання знань поділяються на три категорії, тобто КГК, вилучення зв'язків та відкриття сутностей. Перший - для розширення існуючих графів знань, а два інших відкривають нові знання (вони ж відносини і сутності) з тексту. КГК поділяється на такі категорії: рейтинг на основі вбудовування, міркування про шлях відношення, міркування на основі правил та метареляційне навчання. Виявлення сутності включає розпізнавання, неоднозначність, введення тексту та вирівнювання. Моделі вилучення зв'язків використовують механізм уваги, графові згорткові мережі (GCN), змагальне навчання, навчання з підкріпленням, глибоке залишкове навчання та трансферне навчання.

Графи тимчасових знань включають часову інформацію для навчання репрезентації. Це дослідження класифікує чотири дослідницькі галузі, включаючи тимчасове вбудовування, динаміку сутностей, тимчасову реляційну залежність та часові логічні міркування.

Програми, обізнані з знаннями, включають розуміння природної мови (NLU), відповіді на запитання, системи рекомендацій та різні реальні завдання, які вводять знання для покращення навчання репрезентації.

Зі своєю публікацією в блозі в 2012 році Google представив своє найновіше оновлення продукту як граф знань, щоб покращити свій семантичний пошук, щоб включити цифри реального життя, об'єкти, місця або сутності більш захоплююче з підвищеною зручністю для користувачів. Вони розкрили обсяг свого графа знань, який містить понад 500 мільйонів об'єктів і 3,5 мільярда фактів, взаємопов'язаних один з одним при його реалізації. Завдяки оновленню технології на основі пошукових термінів користувачів, частоти та релевантних даних, підкріплених відомими джерелами, Google обіцяє релевантність результатів пошуку. Застосовне як для настільного, так і для мобільного пошуку, це поєднання виявилось новою епохою для пошукових систем. З тих пір публікація в блозі цитувалася понад 600 разів за даними Google Scholar і викликала інтерес до графів знань обох дослідників і підприємств.

The image shows a Google search results page for the query "Шекспір". The search bar at the top contains the text "Шекспір". Below the search bar, there are navigation options like "Все", "Картинки", "Карты", "Видео", "Покупки", and "Ещё". The search results show a prominent card for "Уильям Шекспир" (William Shakespeare) with a portrait and buttons for "Обзор", "Пьесы", and "Видео". Below this card, there are several search results from Wikipedia and Readmodo. On the right side of the page, a "Граф знань" (Knowledge Graph) panel is visible, which is highlighted with a red border. This panel contains a "Сведения" (Information) section with details about Shakespeare's life, including his birth and death dates, place of birth, and family. Below the information section, there are sections for "Пьесы" (Plays) and "Также ищут" (Also search for).

**Граф знань**

**У вигляді панелі знань**

**Сведения**

Уильям Шекспир — английский поэт и драматург, зачастую считается величайшим англоязычным писателем и одним из лучших драматургов мира. Часто именуется национальным поэтом Англии. Дошедшие до нас работы, включая некоторые, написанные совместно с другими авторами, состоят из 38 пьес, 154 сонетов, 4 поэм и 3 эпитафий. [Википедия](#)

**Дата и место рождения:** Стратфорд-на-Эйвоне, Великобритания

**Дата и место смерти:** 23 апреля 1616 г., Стратфорд-на-Эйвоне, Великобритания

**Супруга:** Энн Хатауэй (в браке с 1582 г. до 1616 г.)

**Родители:** Джон Шекспир, Мэри Арден

**Дети:** Хемнет Шекспир, Сьюзен Холл, Друдит Шекспир

**Братья и сестры:** Джоан Шекспир, Зрмунд Шекспир, Маргарет Шекспир, Гилберт Шекспир, Аена Шекспир, Ричард Шекспир

**Пьесы**

Гамлет  
Ромео и Джульетта  
Король Лир  
Сон в летнюю ночь

**Также ищут**

Энн Хатауэй  
Хемнет Шекспир  
Мигель де Серва...  
Сьюзен Холл

Рисунок 1.5 "Шекспір" Сторінка пошуку Google.

Граф знань Google Contains узагальнив інформацію в галузях знань, щоб дати стислий огляд конкретної теми та встановити подальші рекомендації щодо пов'язаних галузей. Нижче пошуковий запит за "Шекспіром" призвів до детальної інформації про життя і творчість Вільяма Шекспіра. Інформація генерується автоматично і відображається безпосередньо без необхідності

подальшого додаткового пошуку. Можна знайти пов'язані пошукові запити, що містять подібних відомих виконавців, які часто використовувалися як продовження початкового пошукового запиту і можуть зацікавити користувача.

За даними Venetis, приблизно 20% усіх пошукових запитів у Інтернеті зосереджені на пошуку місцевої інформації, при цьому це число подвоюється для пошуку з мобільних телефонів. Завдяки передовим технологіям, здатним визначити точне місцезнаходження користувачів, важливість наявності списку компаній, доступних на пошукових платформах, була розширена.. Як правило, локальні пошуки на мобільних телефонах здійснюються людьми, які хочуть швидких рішень та відповідей, крім таких факторів, як репутація, близькість та візуальність Представлення важливі для вибору користувачів. Чіткі списки бізнесу можуть забезпечити основну цінність для компаній і дозволити їм виділитися з широких мас.

Для відображення релевантної інформації про підприємства "Google My Business" використовується для ввімкнення автоматично згенерованих списків бізнесу за допомогою графів знань. Компанії можуть претендувати на ці поля, надаючи дійсне підтвердження права власності для налаштування інформації для своїх клієнтів. Ця платформа включає в себе невідтворення Карт Google, час роботи, розділи запитань і відповідей, огляди та посилання на соціальні мережі. Компанії також мають можливість показувати спеціальні пропозиції, акції та завантажувати відповідні ділові фотографії. Клієнти можуть безпосередньо взаємодіяти з компаніями та задавати питання або подавати свої коментарі у вигляді відгуків з оцінками від 1-5.

На сторінці пошуку "CofeeFactory прага", кав'ярні в Чехії, можна знайти, яка відображає його список Google Мій бізнес, автоматично згенерований Знання Граф. Регулярна взаємодія з питаннями клієнта в розділі запитань і відповідей, а також велика кількість відгуків про Google дають огляд того, чого клієнт може очікувати від відвідування цієї кав'ярні.

Google coffeefactory прага

Уві Карти Зображення Відео Новини Більше Інструменти

Приблизна кількість результатів: 13 100 (0,36 с)

<https://coffeefactory.cz> - Перекласти цю сторінку  
**COFFEE FACTORY - Praha**  
 Coffee Factory is a real island of smiles and magic in the ocean of Pragues' avenues. Here you can find the atmosphere of friendship, energy and love!  
[HOME](#) · [MENU](#) · [DRINKS](#) · [REVIEWS](#)

<https://ru-ru.facebook.com> - Coffee Factory - Facebook  
**Coffee Factory | Prague - Facebook**  
 Coffee Factory, Прага. Отметки "Нравится": 381 · Посетили: 623. Fancy place for delicious breakfast and brunch. Whilst in the evening is bar for...  
 ★★★★★ Оцінка: 4,6 · 49 голосів

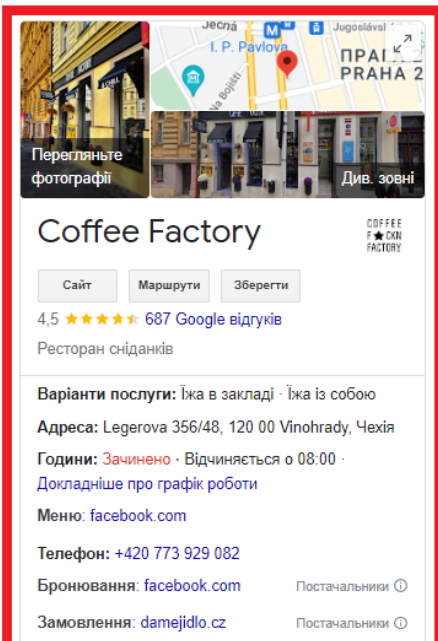
<https://www.tripadvisor.ru> - Прага - Ресторани Праги  
**Coffee Factory, Прага - фото ресторана - Tripadvisor**  
 Coffee Factory, Прага: посмотрите 257 объективных отзывов о Coffee Factory с оценкой 4,5 из 5 на сайте Tripadvisor и рейтингом 94 среди 6 350 ресторанов в ...  
 ★★★★★ Оцінка: 4,5 · 257 відгуків · Діапазон цін: \$

<https://www.instagram.com> - Перекласти цю сторінку  
**coffeefactory.prague - Instagram**  
 Since 2017 • Mon - Sun: 8:00-18:00 • Kitchen: 8:00-16:00 • A perfect place for fancy brunch and delicious breakfast • Catalan wine and cava Portell.

<https://www.agoda.com> - hotel - Перекласти цю сторінку  
**Апартаменты студии 33 кв. м, собственных ванных ... - Agoda**  
 Бронируйте Апартаменты студии 33 кв. м, собственных ванных: 1, в Прага 2 (Coffee Factory House Home) (Чехия) со скидкой. Реальные отзывы гостей и фото.  
 ★★★★★ Рейтинг: 10/10 · 1 відгук · Діапазон цін: room rates from per night (RUB) - We Price Match!

<https://ru.restaurantguru.com> - Перекласти цю сторінку  
**Кафе Coffee Factory, Прага - Меню и отзывы о ресторане**  
 26 лист. 2022 р. — Coffee Factory, №21 среди клубов Прага: 856 отзывов и 21 фотография. Забронировать столик. Это место предлагает вам блюда по цене CZK 94 ...  
 ★★★★★ Оцінка: 4,5 · 856 відгуків · Діапазон цін: \$

**Гаф Знать**  
 у вигляді Google My Business



**Coffee Factory**  
 Сайт Маршрути Зберегти  
 4,5 ★★★★★ 687 Google відгуків  
 Ресторан сніданків

Варіанти послуги: Їжа в закладі · Їжа із собою  
 Адреса: Legerova 356/48, 120 00 Vinohrady, Чехія  
 Години: Зачинено · Відчиняється о 08:00 · Докладніше про графік роботи  
 Меню: facebook.com  
 Телефон: +420 773 929 082  
 Бронювання: facebook.com  
 Замовлення: damejdl.cz

Запитання й відповіді  
 Переглянути всі запитання (7)

Години пік  
 ПН ВТ СР ЧТ ПТ СБ НД  
 11:00: Зазвичай достатньо людей  
 Потрібно чекати до 15 хв

Рисунок 1.6 "CoffeeFactory" сторінка пошуку Google

### 1.3 Дослідження методів пошукового просування сайтів в мережі інтернет

Пошукова оптимізація (SEO) - це процес, який покращує кількість і якість трафіку на веб-сайт через пошукові системи, і його метою є досягнення видимості веб-сайту за допомогою використання органічних результатів

пошукової видачі. Ці результати з'являються в неспонсорованому розділі пошукової видачі, і на відміну від платної реклами, коли користувач натискає на результат і переходить на веб-сайт, маркетингові витрати для конкретної компанії відсутні. Важливість хорошого місця в пошуковій видачі за певними термінами не викликає сумнівів, оскільки користувачі, як правило, переглядають результати тільки в межах перших трьох сторінок або тільки першої сторінки пошуку.

Це може означати, що якщо веб-сайт не знаходиться на першій сторінці пошукової видачі або в першій 30-ці, він майже не має шансів на те, що на нього перейде користувач.

SEO можна розглядати як процес створення веб-сайту, який ефективно ранжується за обраними ключовими словами в результатах органічного пошуку і одночасно за допомогою цих ключових слів покращує якість і обсяг трафіку на веб-сайті. Існує кілька способів використання SEO, і основні способи пов'язані з внутрішніми та зовнішніми аспектами. За допомогою внутрішньосторінкового SEO вимірюється все, що відбувається всередині веб-сайту і що контролюється власником веб-сайту, наприклад, мета-опис, заголовки, контент та зображення. Для того, щоб зайняти більш високі позиції та отримати значущий трафік від пошукових систем, веб-сайт повинен бути зручним для користувача, містити унікальний контент та мати добре описаний метаопис. Крім того, чим довший контент на веб-сайті, тим більше часу користувач витрачає на його читання, що, в свою чергу, збільшує час перебування на сайті. Це явище має тенденцію піднімати сайт вище в результатах пошуку. Позасторінкові методи SEO здебільшого пов'язані із зовнішніми аспектами, які мають суттєвий вплив на ранжування в Google. [1]

Існує два типи методів SEO, як можна досягти видимості веб-сайту без спонсорованої пошукової системи, розробляючи сайт, щоб бути більш зручним для споживачів або зосереджуючись на стратегії, які впливають тільки на процес ранжирування якості пошукової системи. Вони називаються White Hat SEO та Black Hat SEO.

У перші роки існування пошукових систем управління алгоритмами було простіше, що часто призводило до зловживань маркетологами пошуковою системою з метою підвищення рейтингу сайту. Black Hat SEO не зосереджується на якості веб-сайту і не має наміру об'єднувати вбудований вміст відповідно до потреб користувачів, а має намір лише «обдурити» пошукові системи, щоб імітувати це. Такі методи, як наповнення ключовими словами, що стосується розміщення кількох нерелевантних ключових слів на сайті лише для підвищення рейтингу за цими ключовими словами, надмірне тегування та використання неприродних посилань, придумані, щоб представити сторінку в хорошому світлі, але суперечать підосві мета існування пошукових систем.

Існують правила, а точніше офіційні вказівки, опубліковані пошуковими системами, метою яких є роз'яснення прийнятних і небажаних методів, і в серйозних випадках пошукові системи можуть видалити веб-сайти, які використовують дії з пошукової системи, із звичайного списку. Пошукові системи виправдовують цю сувору помсту тим, що маніпуляції результатами пошуку знижують рівень задоволеності клієнтів і погіршують результати «чесних» веб-сайтів.

Тактика SEO, яка не відповідає вказівкам і правилам, наданим пошуковими системами, і не шкодить рейтингу веб-сайту, називається White Hat SEO. Ці методи можуть не дати швидких результатів, але, врешті-решт, повільно покращує рейтинг SERP, і менше шансів на відхилення веб-сайту пошуковими системами. Крім того, трафік, що надходить від White Hat SEO на веб-сайт, часто кращий, а оптимізація триває протягом більш тривалого часу.

Основною та найбільш часто використовуваною тактикою SEO White Hat є генерація ключових слів та обмін посиланнями. Ключові слова – це база SEO, яка надає ваги генерації ключових слів. Це процес вибору найбільш точних ключових слів, які будуть виявлені пошуковими системами і які відповідають змісту сайту. Існує кілька онлайн-інструментів, таких як

інструмент добору ключових слів Google Ad і трекери Word, які генерують релевантні ключові слова відповідно до конкретних потреб.

Іншим методом є обмін посиланнями, який розміщує посилання на інших веб-сайтах для збільшення трафіку. Пошукові системи вважають зовнішні посилання надійними та якісними, що автоматично підвищує рейтинг сайту.

Вирівнювання зовнішніх посилань називається лінк білдингом, тобто створенням декількох доріг на сайт. Іншими ефективними позасторінковими факторами є ведення блогів та використання соціальних мереж, функціонал яких також базується на підборі ключових слів та побудові посилань. Цей метод схожий на співпрацю між різними веб-сайтами, оскільки обидві сторони отримують вигоду від кліків посилань, покращуючи рейтинг у пошуковій системі.

Короткий зміст найбільш поширених методів пошукової оптимізації представлено в таблиці 3, де вони діляться в залежності від того, чи належать вони до Білого капелюха або категорія чорного капелюха.

Таблиця 1 - Відмінності між White Hat SEO та Black Hat SEO

<b>White Hat SEO</b>	<b>Black Hat SEO</b>
Методи дотримання вказівок пошукової системи	Методи, що порушують правила пошукової системи
Аналіз і вибір ключових слів	Наповнення невідповідними ключовими словами
Обмін/побудова посилань	Маніпуляції посиланнями
Високоякісний контент для користувачів сайту	Крадений/автоматично сгенерований контент

При порівнянні SEO White Hat і Black hat SEO разом очевидно, що White Hat SEO є більш вигідним, оскільки він зосереджений на задоволенні

споживачів, стимулює на покращення фактичного вмісту на сторінці, менше шансів, що веб-сайт буде покараний або заборонений пошуковими системами.

The image shows a Google search result for "готель holiday inn київ". The search bar at the top contains the query. Below the search bar, there are navigation options like "Все", "Карты", "Картинки", "Новости", "Видео", "Ещё", and "Инструменты". The search results section shows several entries:

- Information about results:** "Результатов: примерно 250 000 (0,71 сек.)" and "Информация про результаты".
- First result:** "Holiday Inn Hotel Kiev - Kyiv Отели - IHG Hotels & Resorts". It includes a description, a 4.7 rating from 822 reviews, and a price range. A "Відповідний результат" label is present.
- Second result:** "Холідей Інн Київ | Цены на выбранные даты" from Google My Business. It features a bar chart showing prices for different dates, with a highlighted date "ср, 28 дек." and a price of "от 2257 грн.". A "ПОКАЗАТЬ ЦЕНЫ" button is visible.
- Third result:** "Holiday Inn - Kyiv, an IHG Hotel, Киев - Booking.com". It includes a description, an 8.9/10 rating from 892 reviews, and a price range.
- Fourth result:** "HOLIDAY INN КИЕВ - Hotel-Ds.Com". It includes a description, an 8.9/10 rating from 892 reviews, and a price range.
- Fifth result:** "HOLIDAY INN КИЕВ - отзывы, фото и сравнение цен" from TripAdvisor. It includes a description, a 4.5 rating from 450 reviews, and a price range.

A detailed card for "Холідей Інн Київ" is highlighted with a red border. It shows:

- Rating:** 4,6 stars (1 600 отзывов на Google), 4 звезды.
- Buttons:** "Сайт", "Маршрут", "Сохранить", "Вызов", and a large blue "ЗАБРОНИРОВАТЬ НОМЕР" button.
- Address:** Адрес: Велика Васильківська, 100, Київ, 03150. Телефон: 044 363 3000.
- Price Comparison:** "Сравнение цен" with buttons for "вт, 20 дек.", "ср, 21 дек.", and "2".
- Options:** "Все варианты" with "Destinia" (2537 грн.) and "BookingBargains.com" (2527 грн.). A "Показать ещё 4" button and "От 2633 грн." are also present.
- Description:** "Этот элегантный отель расположен в финансовом районе Киева, в 8 минутах ходьбы от ближайшей станции метро, в 3 км от построенного в XI веке Софийского собора и в 6 км от Национального ...".
- Amenities:** "Бесплатный Wi-Fi", "Платный завтрак", "Платная парковка", "Подходит для инвалидов", "Кондиционер", "Прачечная".

Рисунок 1.7 Сторінка результатів пошуку Google.

Для повного розуміння що таке SEO і для чого воно потрібно розуміти що таке пошукова ситстема і з яких компонентів вона складається. Веб-пошукова система - це програмне забезпечення, таке як Google і Bing, яке

збирає дані з веб-сайтів, які зберігаються в базах даних, і звідти забезпечує найбільш релевантний результат для шуканої теми. Пошукова система показує сторінки результатів пошукової системи (SERP), що складаються зі списку проіндексованих веб-сайтів за певними ключовими словами. На рисунку 4. зображено як відображається та оцінюється пошуковою системою конкретний веб-сайт.

Виділені розділи на сторінці результатів пошуку Google:

- поле пошуку - це розділ введення пошукової системи, де користувачі можуть розмістити своє слово, фразу або речення. Пошук також можна здійснити, використовуючи промову, записавши та натиснувши зображення мікрофона в правому куті вікна пошуку;

- інформація про результати показує інформацію про загальну кількість результатів, що мають відношення до конкретного пошукового слова або фрази;

- платна реклама також відома як Ціна за клік (PPC) показує результати, які купуються компаніями в пошуковій системі. Це розділ реклами, де результати показані на основі таких змінних, як релевантність, сума ставки та рейтинг кліків;

- органічні результати з'являються в цьому розділі виключно на основі їх якості та релевантності;

- google Мій бізнес – це інструмент, який пропонує Google, який допомагає компаніям, надаючи йому публічну ідентичність і присутність із списком. Надана інформація з'являється в Пошуку Google і на Картах Google;

- граф знань, як уже обговорювалося, являє собою автоматизоване зведення відомостей про суб'єкта господарювання.

Компоненти пошукової системи в основному пропонують дві функції: процес індексації та процес запитів. Процес індексування створює дані з можливістю пошуку, а процес запиту використовує ці дані для складання списку результатів і оскільки користувач розглядає лише кілька обраних, для компаній дуже важливо бути на вершині цих списків.

Пошукові системи можна класифікувати на три різні категорії: пошукові системи на основі обхідників, каталоги на основі людини та гібридні інструменти пошуку. Сучасні пошукові системи працюють з програмами-сканерами, які автоматично створюють бази даних, відновлюючи дані з місць призначення, більш старомодний каталог, що працює на людях, покладається на ручну роботу, і не пропонує ефективні результати, коли пошук дуже конкретний.

Третій, гібридні інструменти пошукових систем, використовує як обхідну, так і ручну роботу для створення списків результатів. Деякі веб-індекси на основі сканера, як-от Google, використовують сканери як основний компонент і ручний скринінг як допоміжний інструмент.

Пошукові системи важливі, оскільки вони пропонують ефективний доступ до широкого кола тем та інформації та надають найбільш релевантні результати пошуку для користувача. Для компаній правильне використання пошукових систем є життєво важливим, оскільки нові клієнти часто заходять на веб-сайт компанії через пошукові системи під час пошуку інформація про продукт або послугу.

Одним із найважливіших факторів, які впливають на SEO є вибір ключових слів. Розуміння моделей і тенденцій пошуку користувачів має вирішальне значення при ідентифікації попиту на конкретний зміст і сприяння його поліпшенню. Ключові слова - це слова, які користувачі вводять у вікно пошуку в пошукових системах, і ці слова класифікуються Google під час сканування та індексації веб-сторінок. Ключові слова є основним фактором, який відповідає пошуковій системі та веб-сайту разом, а це означає, що алгоритми пошукової системи піднімають сторінки з правильними ключовими словами на список результатів пошуку.

Найкращі ключові слова для компанії залежать від самої галузі та яких пошукових термінів клієнти використовують при використанні пошукових систем. Наприклад, компанія, що займається спортивним обладнанням, повинна вибрати ключові слова, щоб охопити всі різні спортивні жанри, і

прагнути розмістити якомога вище місце в пошуковій видачі. Ключові слова, такі як " спортивне обладнання в Інтернеті" та "магазин спортивного обладнання", занадто загальні, тому хорошим способом покращити список було б використання конкретних ключові слова, які відповідають потребам клієнта. На закінчення, ключові слова - це слова, які споживачі використовують при пошуку конкретних речей в інтернеті. Під час вибору ключових слів компаніям потрібно визначити ці слова та додати їх до свого вмісту, щоб пошукові системи могли класифікувати їх на основі покриття і придатність.

Чим більш релевантні слова використовує компанія, тим краще у неї є те, що вона піднімається на вершину сторінки результатів і заманює клієнти повинні натиснути на посилання на потрібний веб-сайт. [2]

Двома найпоширенішими типами ключових слів є короткі та довгі ключові слова, які мають певну функцію та призначення.

Ключові слова з коротким хвостом, як впливає з назви, є пошуковими термінами, що складаються з одного слова, і дуже популярні у використанні через численні щоденні пошуки. Через високий рівень пошукових запитів конкуренція з короткими ключовими словами є жорсткою між гравцями, які хочуть отримати якомога вищий рейтинг за допомогою ключового слова в SERP.

Важливість високого розміщення в результатах пошуку очевидна, оскільки користувачі частіше натискають результати на першій сторінці. Крім того, з популярними ключовими словами, швидше за все, не конвертується в продажі, оскільки користувачі, які шукають за короткими ключовими словами, зазвичай в основному шукають інформацію про тема, продукт чи послуга, а не прийняття ефективних рішень.

Довгі ключові слова складаються з 2 або більше описових слів і є більш орієнтованими на клієнта, з нижчою конкуренцією, ніж слова з коротким хвостом. Оскільки вони більш конкретно шукаються, вони, як правило, залучають більше якісних клієнтів і конвертують більше продажів.

Ключові слова з довгим хвостом використовуються рідше, ніж ключові слова з коротким хвостом, і є менше пошукових запитів щомісяця, що полегшує високі позиції в результатах пошуку, оскільки конкуренція менша, тому користувач, який прагне придбати певний продукт або послугу, має більшу ймовірність ввести довге ключове слово, і це може бути сигналом до рішення про покупку.

Постійний розвиток розробниками, такими як Google, для посилення алгоритмів пошукових систем для боротьби з наслідками SEO і можливість обману системи шляхом генерації покращив результати пошуку для бізнесу, переріс у “гонку озброєнь” між обома залученими сторонами. Щоб залишатися надійним і високим стандартом якості, пошукова система повинна розробляти все більш складні алгоритми для встановлення конкурентоспроможності за допомогою експлуаторських дій. Навпаки, SEO-фахівці зобов'язані стати винахідливими зі своїми стратегіями і покладатися на експерименти та спостереження, як і до сих пір, жодній компанії не вдалося повністю виявити деталі рейтингу пошукових систем. [3]

З графом знань Google представив новий інструмент для збору інформації, від якого SEO може отримати користь. Хоча до цих пір дуже мало академічних досліджень було спрямовано на з'єднання графів знань та SEO. Ця нова технологія дозволить збільшити охоплення, маркетингове зростання, авторитет, довіру та репутацію.

Крім того, цінність взаємозв'язку бренду та надання підтримки вище звичайних засобів є благодійниками гарної присутності в Інтернеті, розробленої графами знань.

Для цього дослідження було вирішено зосередитися виключно на стратегіях White Hat SEO, зокрема «Вибір ключових слів», «Генерація ключових слів» та «Посилання» Обмін".

Ці методи є актуальними та юридичними стратегіями щодо SEO для бізнесу в наш час і підпадають під затвердження всіх рекомендацій пошукових систем, що суперечать Техніки чорного капелюха.

В кінцевому підсумку вигодою є сприяння задоволенню клієнтів і поліпшення використання. Використовуючи ці стратегії та впроваджуючи їх безпосередньо в граф знань, наприклад, через “Google My Business” (GMB), компанії можуть покращити свої слід в Інтернеті та отримати ділову цінність.

Відомо, що вплив GMB на місцевий бізнес збільшується до 25% при правильному використанні SEO.

Теоретична база дослідження побудована для аналізу зв'язку між графами SEO та графами знань. З аналізом поширених методів SEO, а саме «Вибір ключових слів», «Генерація ключових слів» і «Обмін посиланнями», відомими компаніями створюється уявлення про поточну бізнес-практику щодо SEO та її переваги.

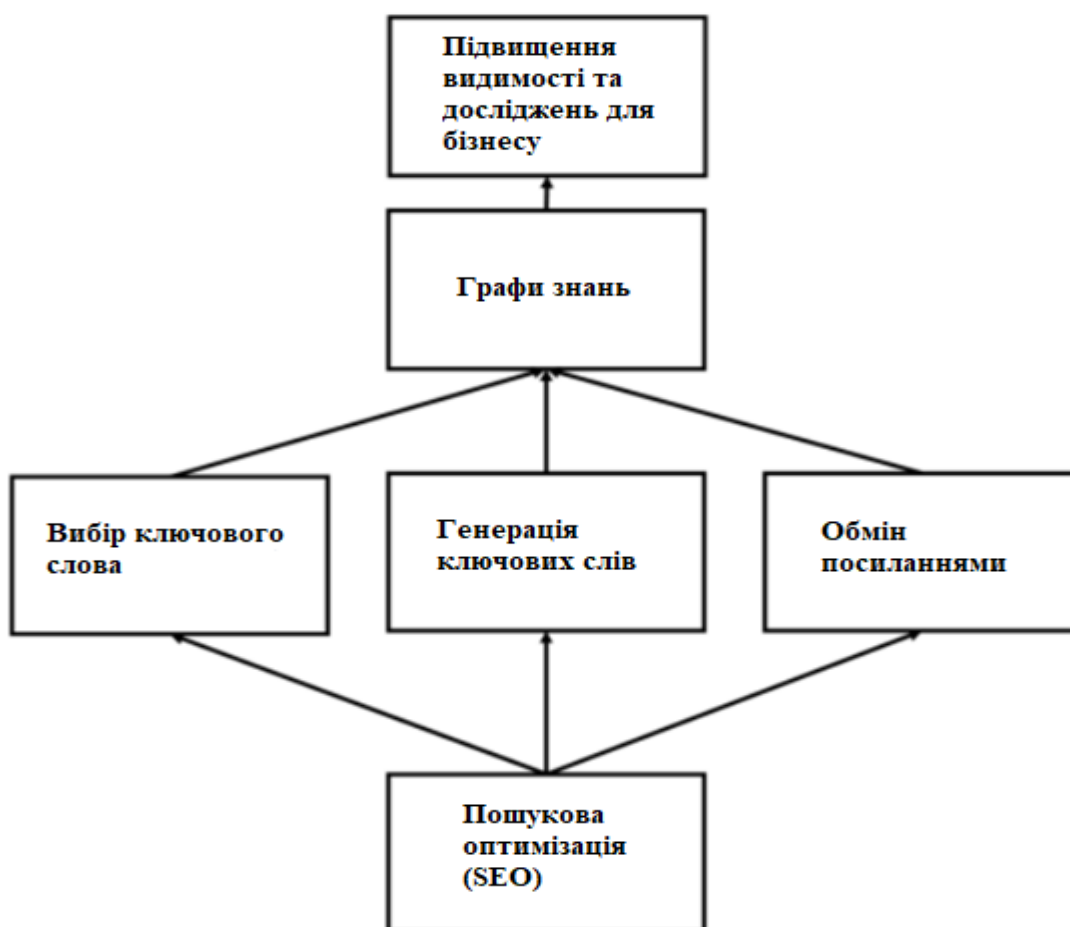


Рисунок 1.8 Пошукове просування сайту з використанням графу знань.

Ще одним важливим інструментом для оптимізації сайту, який побічно стосується і теми цієї роботи, є схема розмітки сайту. Вона існує для того щоб під час написання сайту використовуючи елементи верстки указати пошуковій системі які компоненти є найбільш вагомими і які потрібно індексувати в першу чергу.

Найпопулярнішим сервісом який допомагає допомогти знайти спільну мову між програмістом та пошуковій системою є Schema.org. Це спільні зусилля багатьох програмістів та ентузіастів спрямовані на покращення інтернету шляхом створення схеми розмітки структурованих даних, яка підтримується основними пошуковими системами.

Більшість веб-майстрів знайомі з тегами HTML на своїх сторінках. Зазвичай теги HTML повідомляють браузеру, як відображати інформацію, що міститься в тегу.

Наприклад, `<h1>Аватар</h1>` повідомляє браузеру відобразити текстовий рядок «Аватар» у форматі заголовка 1.

Однак тег HTML не надає жодної інформації про те, що означає цей текстовий рядок — «Аватар» може стосуватися надзвичайно успішного 3D-фільму або може посилатися на тип зображення профілю – і це може ускладнити пошук механізми для інтелектуального відображення релевантного вмісту для користувача.

Наприклад, на рисунку 1.7 один із семантично вірних варіантів розмітки стандарту HTML5, в якому присутні відповідними тегами розділу та статті, які допомагають визначити початок і кінець самого тексту, так і його розділів.

Для більшого розуміння самої структури розмітки вона зображена на рис 1.8 у вигляді схеми.

Розмітка на сторінці допомагає пошуковим системам розуміти інформацію на веб-сторінках і надавати повніші результати пошуку. Спільний словник розмітки полегшує веб-майстрам вибір схеми розмітки та отримує максимальну вигоду від своїх зусиль.

Пошукові системи хочуть полегшити людям пошук потрібної інформації в інтернеті.

```
1
2 <!-- СЕМАНТИЧЕСКАЯ РАЗМЕТКА ТЕКСТА HTML5: -->
3
4 <!-- Начало текста -->
5 <article>
6 <h1>Заголовок</h1>
7 <p>Абзац...</p>
8 <!-- Начало раздела 1 -->
9 <section>
10 <h2>Заголовок 2 уровня</h2>
11 <p>Абзац...</p>
12 <ul>
13 <li>Элемент 1</li>
14 <li>Элемент 2</li>
15 <li>Элемент 3</li>
16 </ul>
17 <!-- Начало раздела 1.1 -->
18 <section>
19 <h3>Заголовок 3 уровня</h3>
20 <p>Абзац...</p>
21 </section>
22 <!-- Конец раздела 1.1 -->
23 </section>
24 <!-- Конец раздела 1 -->
25 <!-- Начало раздела 2 -->
26 <section>
27 <h2>Заголовок 2 уровня</h2>
28 <p>Абзац...</p>
29 </section>
30 <!-- Конец раздела 2 -->
31 </article>
32 <!-- Конец текста -->
33
```

Рисунок 1.9 Розмітка тексту

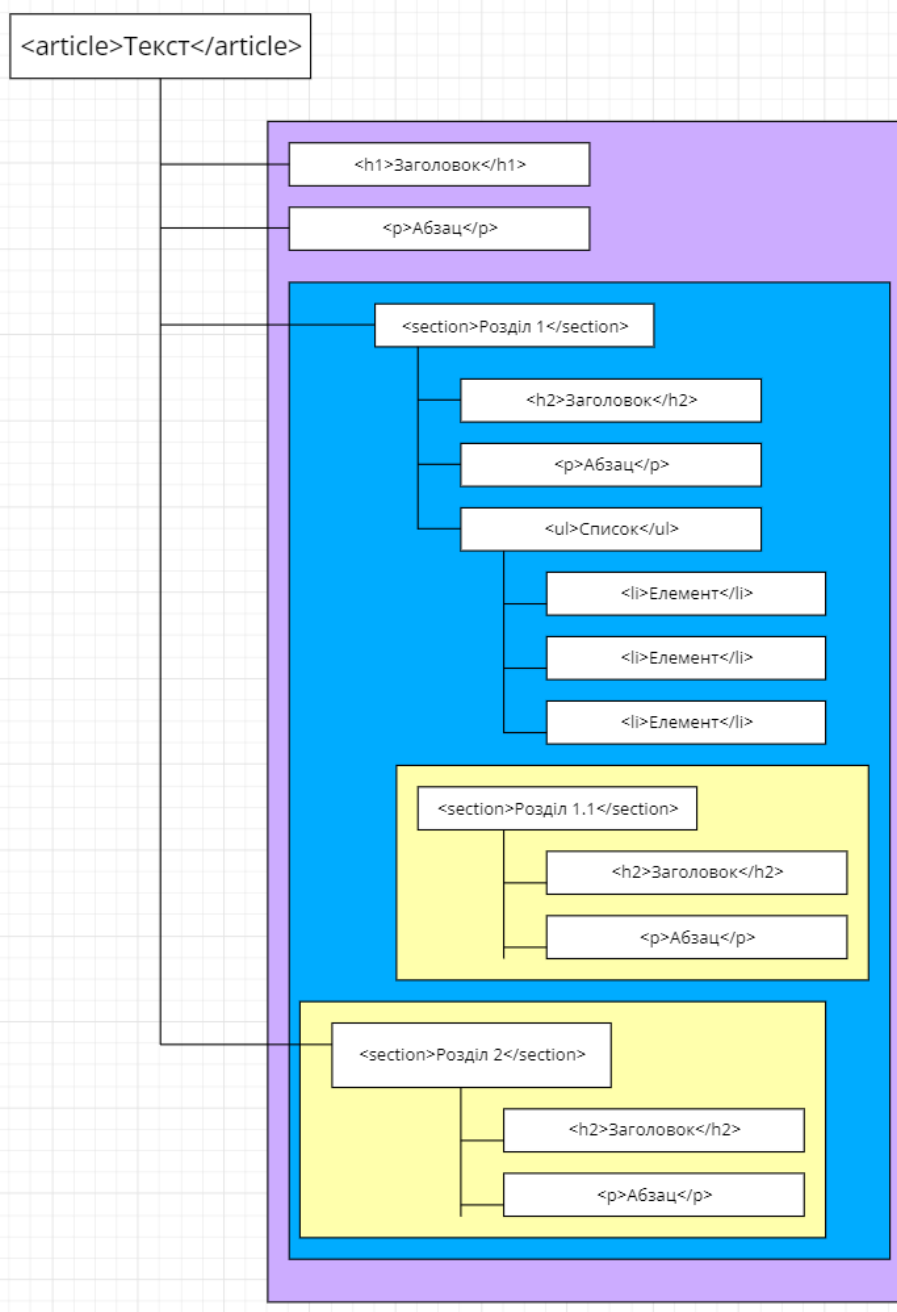


Рисунок 1.10 структура розмітки в HTML5

Розмітка також може використовувати нові інструменти та програми, які використовують цю структуру.

В даний час існує багато стандартів і схем для розмітки різних типів інформації на веб-сторінках. Як наслідок, веб-майстрам важко визначитися з найбільш актуальними та підтримуваними стандартами розмітки для використання.

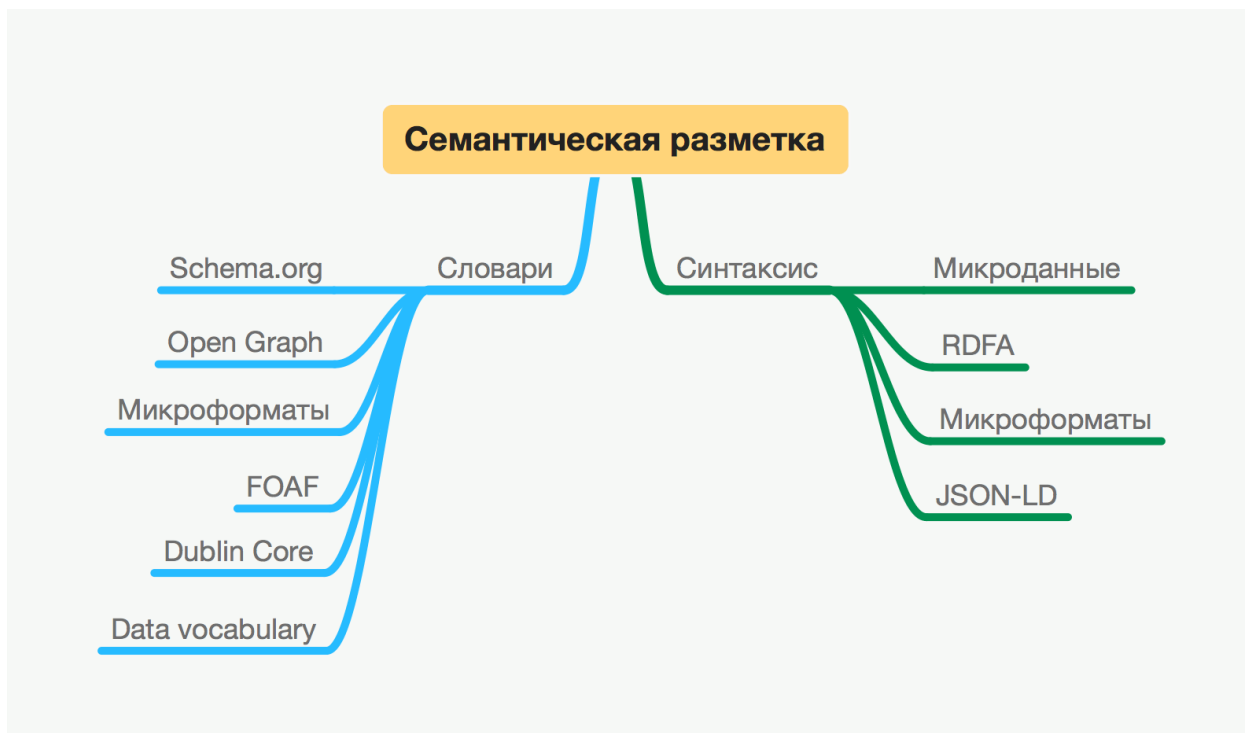


Рисунок 1.11 Компоненти семантичного пошуку

Створення схеми, яка підтримується всіма основними пошуковими системами, полегшує веб-майстрам додавання розмітки, що полегшує пошуковим системам створення розширених функцій пошуку для користувачів. Саме тому конкуруючі між собою пошукові системи об'єднались у цьому питанні і обрали єдиний стандарт.

#### 1.4 Мета дослідження

Протягом останнього десятиліття різні інструменти цифрового маркетингу модифікували сферу маркетингу. Завдяки вдосконаленню технологій стає легше орієнтувати споживачів на більш конкретний та персональний контент, що робить маркетинг більш ефективним та економічно вигідним. Оскільки ідея графа знань полягає в тому, щоб збирати, зберігати та використовувати дані про ключові слова для надання точних та релевантних

результатів, це полегшує застосування цільового маркетингу на практиці. За допомогою графа знань існує величезна кількість зібраних даних, які можна використовувати і повторно використовувати, і проблема полягає в тому, щоб вибрати, як використовувати їх найбільш ефективним чином.

Для того, щоб компанії з'явилися на графах знань, вони повинні мати трафік на своєму веб-сайті та більшість кліків на результатах пошуку. Це означає, що для компаній вкрай важливо, щоб їхня пошукова оптимізація працювала для того, щоб пошукові системи знаходили створений контент і робили його доступним на сторінках результатів пошуку. Компанії можуть покращити свою видимість та рейтинг у результатах пошуку за допомогою SEO, і оскільки це пов'язано з графами знань, сучасні маркетологи повинні знати, як використовувати цей зв'язок. Графи знань не розглядалися з точки зору бізнесу та маркетингу в багатьох академічних дослідницьких журналах, а в основному в блогах галузевих експертів. Не існує досліджень щодо того, наскільки широко цей інструмент відомий серед сучасних маркетологів. Досі графи знань досліджувалися переважно з точки зору ІТ, аналізу інтеграції та розвитку технологій. Ми виявили прогалину в дослідженнях через відсутність академічних досліджень щодо використання графів знань з точки зору бізнесу. Крім того, ми вважаємо, що SEO має позитивний вплив на бізнес на графах знань, на якому буде зосереджено увагу в цьому дослідженні.[4]

Метою цього дослідження є аналіз використання SEO та графів знань.. Також це дослідження охоплює загальні практики SEO та встановлює поточне та потенційне використання графів знань для підвищення видимості та охоплення бізнесу. Це дослідження демонструє корисний зв'язок між цими двома темами.

Провівши великий огляд літератури в галузі пошукової оптимізації, з'являється чітке усвідомлення існування графів знань та їх зростаючий вплив на сучасне життя з точки зору споживача. Мотивація цього полягає в тому, щоб, по-перше, продемонструвати маркетингову стратегію пошукової оптимізації, вбудовування ключових слів, посилань, зображень у веб-сайти

для генерування трафіку веб-сайту через підвищення рейтингу в пошукових системах. По-друге, зв'язати тему з графами знань - новим інструментом, який автоматизує збір даних з веб-сайтів за допомогою алгоритмів, щоб надати користувачам пошукових систем швидкий і більш широкий огляд їх пошукового запиту.

На сьогоднішній день графи знань є повсюдними і забезпечують основу для пошуку в Google та багатьох поточних рекомендацій, наприклад, на YouTube, Netflix, Facebook, LinkedIn та багатьох інших. Попередні дослідження в цій галузі були проведені на графах знань в основному з точки зору IT, аналізуючи інтеграцію та розвиток технологій, благодійників застосування штучного інтелекту, стислий обмін інформацією через графи знань та соціальні медіа. Незважаючи на те, що було проведено багато досліджень з точки зору бізнесу та маркетингу у вигляді публікацій у блогах, практично немає рецензованих академічних наукових журналів, які б зосередилися на цій темі, що дає унікальний кут зору та можливість долучитися до неї.

Об'єктом дослідження є процес просування сайтів в пошуковій системі з використанням графу знань

Предметом дослідження є методи просування сайтів в пошуковій системі з використанням графу знань

Сучасні пошукові системи використовують графи знань для визначення результатів пошуку, на відміну від традиційних підходів, граф знань додає можливість отримати результат на основі семантичного пошуку. Цей семантичний пошук базується на аналізі тексту на сторінках відповідного веб-сайту. Тому побудова і аналіз графу знань для веб-сайту дає можливість забезпечити його пошуковим просуванням.

Використано підхід до аналізу графу знань для пошукового просування сайтів з використанням графу знань Підхід до пошукового просування сайту на основі аналізу графу знань

Для досягнення цієї мети необхідно:

- провести аналіз особливостей семантичного пошуку;
- дослідити методи пошукового просування;
- проаналізувати граф знань;
- експериментально перевірити аналіз графу знань для пошукового просування.

## 2 Використання аналізу графу знань при просуванні сайтів в пошуковій системі

### 2.1 Побудова каркасу графів знань

Для широкомасштабної графової аналітики були запропоновані різні структури, часто в розподілених (кластерних) налаштуваннях. Серед них: Apache Spark (GraphX), GraphLab, Pregel, Signal–Collect, Shark тощо. Ці графові паралельні структури застосовують систолічну абстракцію на основі орієнтованого графа, де вузли є процесорами, які можуть надсилати повідомлення іншим вузлам уздовж країв.

Тоді обчислення є ітераційним, де в кожній ітерації кожен вузол читає повідомлення, отримані через внутрішні ребра (і, можливо, свій власний попередній стан), виконує обчислення, а потім надсилає повідомлення через зовнішні ребра на основі результату. Потім ці структури визначають систолічну обчислювальну абстракцію поверх графа даних, що обробляється: вузли та ребра на графу даних стають вузлами та ребрами на систолічному графу.

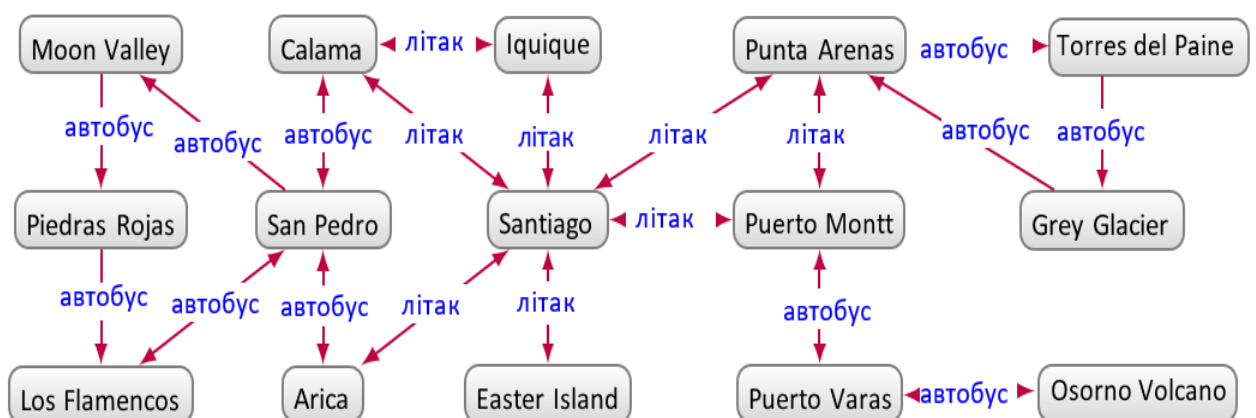


Рисунок 2.1 Граф даних сайту із транспортною інформацією в Чілі

Наприклад, якщо потрібно обчислити місця, до яких найбільше (або найменш) легко дістатися за маршрутами, показаними на графу на рисунку 2.1 гарним способом вимірювання є використання центральності, де ми вибираємо PageRank, який обчислює ймовірність того, що турист, випадково слідує маршрутам, показаним на графу, опиниться в певному місці після заданої кількості переходів.



Рисунок 2.2 Приклад систолічної ітерації PageRank у фазі повідомлення (Msg)

Вузли ініціалізуються оцінкою  $1/|V|=1/6$ , де припущено, що турист має рівні шанси почати в будь-який момент. У фазі повідомлення (Msg), рисунок 2.2, кожен вузол  $v$  проходить оцінку  $dR_i(v)/|E(v)|$  на кожному з його вихідних країв, де позначено  $d$  як постійний коефіцієнт демпфування, який використовується для забезпечення конвергенції (зазвичай  $d = 0,85$ , що вказує на ймовірність того, що турист випадковим чином перейде в будь-яке

місце), через  $R_i(v)$  оцінку вузла  $v$  на ітерації  $i$  (ймовірність перебування туриста у вузлі  $v$  після  $i$  стрибків), і за  $|E(v)|$  кількість вихідних ребер  $v$ .

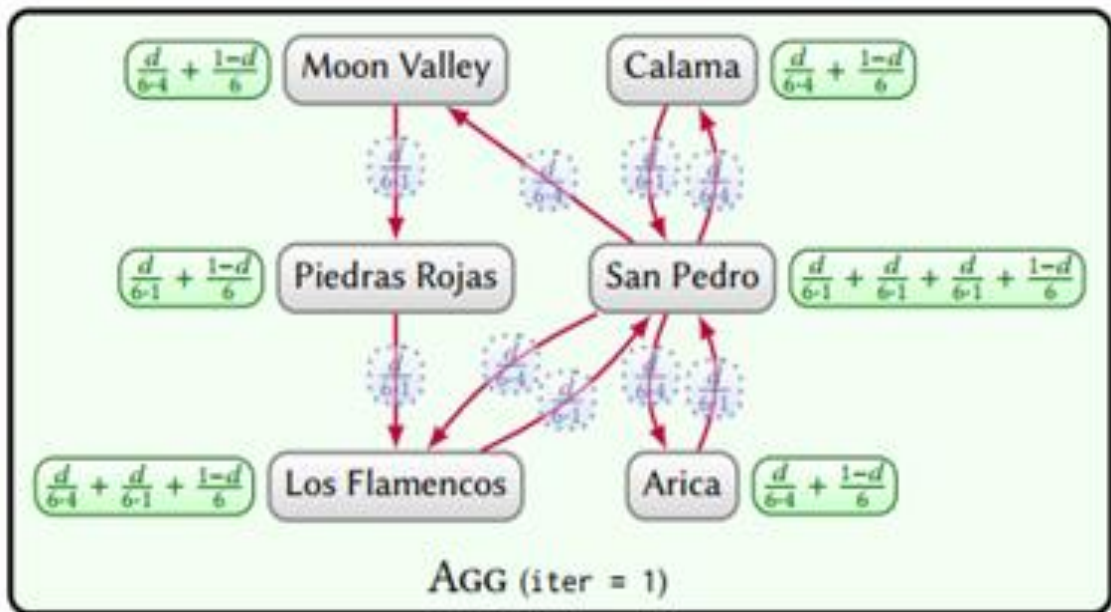


Рисунок 2.3 Приклад систолічної ітерації PageRank у фазі агрегації (Agg)

Потім фаза агрегації (Agg) для  $v$  підсумовує всі отримані вхідні повідомлення разом із постійною часткою коефіцієнта демпфування  $(1-d)/|V|$ , щоб обчислити  $R_{i+1}(v)$ .

Потім слідує фаза повідомлення наступної ітерації, продовжуючи до тих пір, поки не буде досягнуто певного критерію завершення (наприклад, кількість ітерацій або залишкове порогове значення тощо) і не буде виведено остаточні оцінки.

Хоча наведений приклад стосується PageRank, систолічна абстракція є достатньо загальною, щоб підтримувати широкий спектр аналізу графів, у тому числі згаданих раніше. Алгоритм у цій структурі складається з функцій для обчислення значень повідомлень у фазі повідомлення (Msg) і для накопичення повідомлень у фазі агрегації (Agg). Фреймворк займається розповсюдженням, передачею повідомлень, відмовостійкістю тощо.

Однак такі фреймворки – засновані на передачі повідомлень між сусідами – мають обмеження: не всі типи аналітики можна виразити в таких фреймворках. [5]

## 2.2 Аналіз графу знань для пошукового просування

Як згадувалося вище, більшість аналітичних засобів, представлених до цього моменту застосовні для неорієнтованих або орієнтованих графів без крайових метаданих – тобто, крайових міток або пар властивість–значення – типових для графових моделей даних. Можна застосувати кілька стратегій, щоб зробити графи даних підданими аналітиці цієї форми.

**Проекція.** Передбачає просте проєкціювання неорієнтованого або орієнтованого графа шляхом вибору суб-графа з графа даних, з якого видаляються всі краєві метадані; наприклад, малюнок 4 може бути результатом виділення підграфа, викликаного шиною міток краю та польотом, із більшого графа даних, де мітки потім відкидаються, щоб створити орієнтований граф.

**Зважування.** Передбачає перетворення крайових метаданих у числові значення відповідно до певної функції. Багато з вищезгаданих методів легко адаптуються до випадку зважених (орієнтованих) графів; наприклад, ми могли б розглянути вагові коефіцієнти на графу на малюнку 2.1, що позначають тривалість поїздки (або ціну, трафік тощо), а потім обчислити найкоротші шляхи, додавши тривалість кожного відрізка.<sup>21</sup> За відсутності зовнішніх вагових коефіцієнтів ми можемо швидше зіставляти мітки ребер із вагами, призначаючи однакову вагу всім краям польоту, усім краям шини тощо на основі певних критеріїв.

**Трансформація.** Передбачає перетворення графа в модель нижчої арності. Перетворення може бути втраченим, тобто вихідний граф неможливо відновити; або без втрат, що означає, що вихідний граф можна відновити. На малюнку 2.4 наведено приклад перетворення з втратами та без втрат від

орієнтованого графа з мітками ребер до орієнтованих графів. У перетворенні з втратами ми не можемо сказати, наприклад, чи вихідний граф містив ребро Iquique рейс Сантьяго або Iquique рейс Arica тощо. Перетворення без втрат має вводити нові вузли (подібно до реифікації) для збереження інформації про спрямовані помічені ребра. Обидва перетворені графи намагаються зберегти спрямованість вихідного графа.

Налаштування. Передбачає зміну аналітичної процедури для включення крайових метаданих, як це було у випадку пошуку шляху на основі виразів шляху. Інші приклади можуть включати структурні показники подібності вузлів, які враховують не лише спільних сусідів, але й спільних сусідів, з'єднаних ребрами з однаковою міткою, або сукупні показники центральності, які враховують важливість ребер, згрупованих за мітками, тощо.

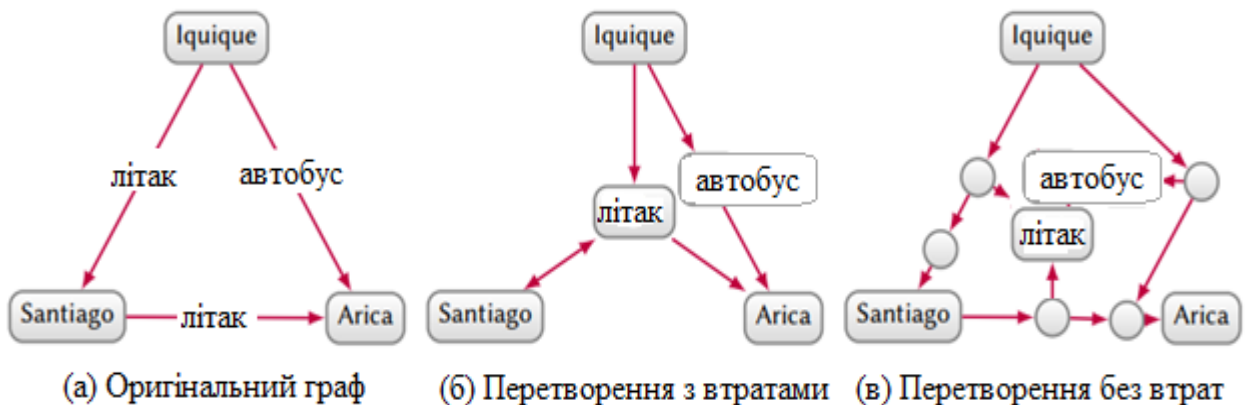


Рисунок 2.4 Перетворення орієнтованого реброво-міченого графа в орієнтований граф

Результати аналітичного процесу можуть різко змінитися залежно від того, яка з попередніх стратегій була обрана для підготовки даних для аналізу. Цей вибір може бути нетривіальним, щоб зробити його апріорі, і може потребувати емпіричного підтвердження. Потрібні додаткові дослідження, щоб більш загально зрозуміти вплив таких стратегій на результати різних аналітичних методів.

### 2.3 Аналіз графу на основі запитів

Розглянемо різні способи, якими мови запитів і аналітика можуть доповнювати один одного. По-перше, це потрібно розглянути можливість використання мов запитів для проектування або перетворення графа, придатного для конкретного аналітичного завдання, наприклад для вилучення графа на малюнку 2.1 із більшого графа даних.

Такі мови запитів, як SPARQL, Cypher і G-CORE, дозволяють виводити графи, де такі запити можна використовувати для вибору підграфів для аналізу. Ці мови також можуть виражати деяку обмежену (нерекурсивну) аналітику, де агрегації можна використовувати для обчислення ступеня центральності, наприклад; вони також можуть мати деяку вбудовану аналітичну підтримку, де, наприклад, Cypher дозволяє знаходити найкоротші шляхи. З іншого боку, аналітика може сприяти процесу надсилання запитів з точки зору оптимізації, де, наприклад, аналіз підключення може запропонувати, як краще розподілити великий граф даних між кількома машинами для надсилання запитів, використовуючи, наприклад, мінімальні скорочення. Аналітику також використовували для ранжирування результатів запиту на великих графах, вибираючи найважливіші результати для представлення користувачеві. У деяких випадках використання ми можемо додатково поєднати запити та аналітичні процеси. [6]

### 3 ПОРІВНЯННЯ ПІДХОДІВ ДО АНАЛІЗУ ГРАФІВ ЗНАНЬ

Аналітика – це процес виявлення, інтерпретації та передачі значущих закономірностей, властивих (зазвичай великим) колекціям даних. Таким чином, аналітика графів - це застосування аналітичних процесів до (зазвичай великих) даних графів. Природа графів піддається певним типам аналітики, яка робить висновки про вузли та ребра на основі топології графа, тобто того, як з'єднані вузли графа. Таким чином, графоаналітика черпає багато своїх методів із суміжних областей, таких як теорія графів і мережевий аналіз, які використовувалися для вивчення графів, що представляють соціальні мережі, інтернет, інтернет-маршрутизацію, транспортні мережі, екосистеми, білок-білкові взаємодії, лінгвістичні випадки, і багато іншого. Повертаючись до області нашого поточного прикладу, рада з питань туризму може використовувати графову аналітику, щоб отримати інформацію про, наприклад: ключові транспортні вузли, які обслуговують багато туристичних визначних пам'яток (централізація); групування пам'яток, які відвідують одні і ті ж туристи (виявлення спільноти); атракціони, які можуть стати недоступними у разі ударів чи інших збоїв маршруту (з'єднання), або пари атракціонів, схожих одна на одну (схожість вузлів). З огляду на те, що така аналітика потребує складного, великомасштабного графа, для ілюстрації на малюнку 2.1 представлено стислий приклад графу знань сайту транспортного сполучення Чилі, спрямованого до популярних туристичних місць.

Для аналізу графів можна застосовувати широкий спектр методів:

- виявлення найважливіших вузлів графу;
- виявлення внутрішньо-зв'язаних підграфів;
- зв'язність;
- подібність вузлів графу;
- пошук шляху на графі.

Виявлення найважливіших вузлів графу має на меті ідентифікувати найважливіші (також центральні) вузли або ребра графа. Конкретні показники центральності вузла включають ступінь, проміжність, близькість, власний вектор, Page Rank, HITS, Katz тощо. Центральність між ними також можна застосувати до країв. Вимірювання центральності вузла дозволить, наприклад, передбачити транспортні вузли на малюнку 2.1, тоді як центральність краю дозволить нам знайти краю, від яких залежить багато найкоротших маршрутів для прогнозування трафіку.

Виявлення внутрішньо-зв'язаних підграфів має на меті фікувати спільноти на графі, тобто підграфи, які тісніше пов'язані внутрішідентиньо, ніж з рештою графа.

Алгоритми виявлення спільнот, такі як алгоритми мінімального скорочення, розповсюдження міток, модульність Лувена тощо, дозволяють виявити такі спільноти. Виявлення спільноти, застосоване до рисунка 2.1, може, наприклад, виявити спільноту ліворуч (що стосується півночі Чилі), праворуч (щодо півдня Чилі) і, можливо, також у центрі (що стосується міст з аеропортами)

Зв'язність має на меті оцінити, наскільки добре зв'язаний граф, виявляючи, наприклад, стійкість і (не)досяжність елементів графа. Конкретні методи включають вимірювання щільності графа або  $k$ -зв'язності, виявлення сильно зв'язаних компонентів і слабо зв'язаних компонентів, обчислення охоплюючих дерев або мінімальних розрізів тощо. У контексті рисунка 2.1 такий аналіз може сказати нам, що маршрути до льодовика Грей, вулкана Осорно та П'єдрас-Рохас є найбільш «крихким», він роз'єднується, якщо один із двох автобусних маршрутів виходить з ладу.

Подібність вузлів графу: прагне знайти вузли, схожі на інші вузли через те, як вони з'єднані в межах свого сусідства. Показники подібності вузлів можуть бути обчислені за допомогою структурної еквівалентності, випадкових блукань, дифузійних ядер тощо. Ці методи забезпечують розуміння того, що з'єднує вузли, і, згодом, у чому вони подібні. У контексті рисунка 2.1 такий

аналіз може сказати нам, що Калама та Аріка є подібними вузлами, оскільки обидва мають зворотні рейси до Сантьяго та зворотні автобуси до Сан-Педро.

Пошук шляху на графі спрямований на пошук шляхів у графі, як правило, між парами вузлів, наданих як вхідні дані. Існують різні технічні визначення, які обмежують набір дійсних шляхів між такими вузлами, включаючи прості шляхи, які не відвідують один і той самий вузол двічі, найкоротші шляхи, які відвідують найменшу кількість ребер, або звичайні запити шляхів, які обмежити мітки ребер, які може пройти шлях.

## 4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА ОТРИМАНИХ РЕЗУЛЬТАТІВ

### 4.1 Обґрунтування вибору інструментальних засобів для забезпечення аналізу просування сайту

На кожному етапі розробки сайтів використовуються різні інструментальні засоби для досягнення мети кожного з етапів.

Для проведення ефективної оптимізації сайту, який буде знаходитись високо в пошуковій видачі, необхідно обрати правильні інструментальні засоби щоб мати змогу своєчасно відстежувати результати роботи і приймати вірні рішення.

### 4.2 Експериментальна перевірка аналізу графу знань для пошукового просування

Апробація методу модернізації сайту з використанням графу знань буде проводитись на прикладі статті блогу на компанії, що займається проведенням тестів на COVID-19. Сайт належить американській компанії, тому текст написаний на англійській мові.

Для прикладу взято текст статті з іншого сайту з невірним використанням графу знань і через це з низьким пошуковим індексом.

Провівши інформаційний аналіз по ключовим запитам було виявлено найбільш актуальні по темі різної частотності, а саме: (antigen-test how gone, antigen-test what mean, covid-19 antigen-test, in antigen-test Miami). Чітко відстежується структура графу знань, що наглядно проілюстровано на рисунку 4.2.

## What is a COVID-19 Antigen Test?

A COVID-19 antigen test detects coronavirus proteins in the mouth and throat. This test determines whether a patient is currently infected with COVID-19.

Antigen testing is a suitable 'first line of defense' test against COVID-19, particularly in individuals who have COVID symptoms, because it is a relatively simple procedure, and is quicker than other approaches – like PCR testing; it typically takes around 15 minutes for a result.

It is less accurate than PCR testing, however, as a high viral load is required for the antigen test to detect SARS-CoV-2 virus (COVID-19). Some people in the early or late stages of infection may therefore return a false result, although this is not common.

## How is a COVID19 Antigen Test Taken?

COVID-19 antigen tests typically involve a swab of the patient's nose and throat (sample can also be saliva or blood). This sample is then analysed using the same underlying processes as a pregnancy test.

Results are typically available within 15 to 30 minutes of the sample being analysed, meaning a patient on site may have their results before they leave.

## Antigen Testing at Unilabs

Unilabs and International Guidelines consider PCR to be the "Gold standard" for diagnosing current COVID-19 infections, due to its higher sensitivity and accuracy. PCR testing can also reliably detect COVID-19 in people who have no symptoms or signs, which is true in around 1/3 of infected people.

However we recognise there are specific situations when the speed and simplicity benefits of antigen testing make it an attractive option; for example, the mass-testing of schoolchildren and their families, or last-minute testing of travellers.

Unilabs works with a wide variety of businesses and groups to deliver bespoke antigen testing programmes at scale.

### Рисунок 4.1 Приклад статті взятий за основу

Пишучи, наприклад, про «antigen», можете матися на увазі не тільки тест на антитіла до COVID-19 а ще різних речей. Це може бути тест з загальної біології або музичний виконавець Antigen Shift. Залежно від продуктивності на веб-сторінці сканери Google дізнаються, що саме представляє ваша веб-сторінка приблизно. Google може точно аналізувати вміст, оптимізувавши граф знань.

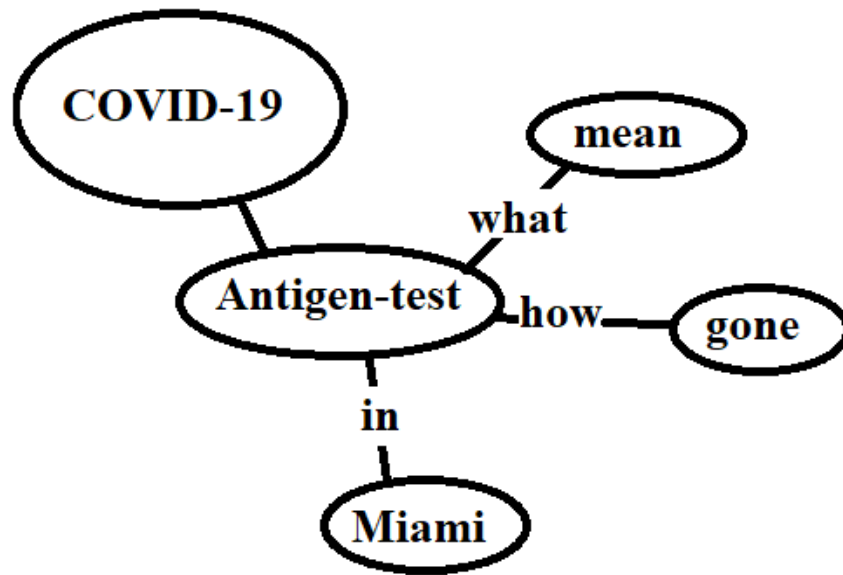


Рисунок 4.2 Граф знань

Звичайно, ключові слова повинні з'явитися у вступі, визначенні та висновку. Далі ключові слова було впроваджено в статтю в прямому входженні в текст. Термін застосовується щодо ключових слів, що входять до складу текстового контенту веб-сторінок, які оптимізуються під просування пошукових систем. Пряме входження ключового запиту має на увазі вживання його в незмінному вигляді, тобто без відмінювання, розведення, перестановки слів місцями і т. д., але з можливістю поділу слів у ньому деякими розділовими знаками (кома, двокрапка, тире).

Перевіривши текст у сервісі аналізу тексту ми бачимо на рисунку 4.4, що після переробки тексту можемо спостерігати наступні показники:

- ключова щільність 52 відсотка;
- вода 31 відсоток;
- унікальність 100 відсотків.

The screenshot displays a text analysis tool interface with three main panels at the top:

- Check for plagiarism:** Shows a unique score of 100.00%. It includes links for 'Get the link', 'Record the text originality', and 'Get Originality Banner', along with a 'More information' button.
- Check Grammar and Spelling:** Reports 7 mistakes found in the text, listing 'COVID 19', 'but', and 'Covid 19'. It also has a 'More information' button.
- Word Count:** Provides statistics: Total symbols: 1845, Keyword density: 52%, Without spaces: 1525, Watery: 31%, and Total words: 324. It includes a 'More information' button.

Below these panels is a text preview window with a 'Highlighted: Matched fragments' indicator. The text preview shows the beginning of an article about COVID-19 antigen tests. To the right of the text preview is a 'Text versions' section showing a version from '5 minutes ago (UTC -05:00)' with a summary of its statistics: Unique: 100%, Grammar and S...: 7, Total symbols: 1845, Keyword density: 52%, Without spaces: 1525, Watery: 31%, and Total words: 324.

Рисунок 4.3 Аналіз тексту

Виходячи з даного аналізу ми можемо зробити висновок, що написаний текст є оптимізований та повністю відповідає заданим раніше критеріям.

Також необхідно оптимізуємо розмітку сайту додавши відповідні семантичні теги як на рисунку 4.3. Це допоможе пошуковій системі визначити які з елементів є більш вагомими за інші і з яких елементів потрібно в першу чергу брати інформація для пошукової видачі.

<H1> What is a COVID 19 Antigen Test?
<H2> What Is the Covid 19 Antigen Test? What Does The Result Mean?
<H2> How is a COVID19 Antigen Test Taken?
<H2> Who Should Get a COVID-19 Antigen Test?

Рисунок 4.4 Елементи мікророзмітки які ми додали до тексту

Після виконання усіх необхідних пунктів, отриманий результат можна спостерігати на рисунку 4.5.

# What is a COVID 19 Antigen Test? What Does The Result Mean?

Covid has been present in our lives for 3 years already but some things still need some clarification. This article is created to answer all your questions about the Covid-19 Antigen test.

## What Is the Covid 19 Antigen Test?

A Covid-19 antigen test detects the presence of antigens within the human body. A positive Covid-19 antigen test result means that the person is infected with Covid at the moment of getting tested. Antigens are the structures made of Coronavirus proteins that appear in your body to fight Covid-19. It's important to mention that it is less reliable than PCR testing. The reason for it is that the antigen test results require a large viral load to identify the SARS-CoV-2 virus. As a result, some patients in the early or late stages of illness may get false results.



## How to Get An Antigen Covid 19 Test?

An antigen Covid-19 test is taken through the nasal and throat swab, however, some may be conducted through saliva or blood. One hour before the test it is recommended not to smoke, eat, or drink beverages to keep the natural environment and reduce the risks of getting falsely negative results.

## How To Make Sure That The Nasal Swab Is Done Correctly?

The medical worker must follow simple steps. The swab must be placed vertically into each nostril over 2 to 3 cm (0.8-1.1 inch). Then, gently shake the swab horizontally and enter it until you encounter some resistance. After it, for around 15 seconds, make a spinning movement within each nostril for at least 5. After collecting the sample, place the swab straight into the extraction tube by swirling it in the solution 5 to 10 times while compressing the tube. Squeeze the tube tightly to extract as much liquid as possible from the swab.

### Рисунок 4.5 Фінальна стаття

Після публікації статті необхідно перевірити отримані результати. Сайт був зроблений з використанням CMS, а саме WordPress, в який вже вбудовані певні інструменти для слідкування за трафіком. На рисунку 4.6 видно, що ця процедура з переробкою тексту принесла понад 16 одиниць живого трафіку на сторінку, що підняло загальний показник трафіку сайту на 15 відсотків.

URL	Status	Traffic
<a href="https://getresulttoday.com/locations/orlando/">https://getresulttoday.com/locations/orlando/</a>		61 54.5%
<a href="https://getresulttoday.com/what-is-a-covid-19-antigen-test/">https://getresulttoday.com/what-is-a-covid-19-antigen-test/</a>		16 14.3%
<a href="https://getresulttoday.com/locations/hialeah/">https://getresulttoday.com/locations/hialeah/</a>		12 10.7%
<a href="https://getresulttoday.com/locations/sunrise/">https://getresulttoday.com/locations/sunrise/</a>		9 8.0%
<a href="https://getresulttoday.com/locations/pompano-beach/">https://getresulttoday.com/locations/pompano-beach/</a>		4 3.6%
<a href="https://getresulttoday.com/locations/kendall/">https://getresulttoday.com/locations/kendall/</a>		4 3.6%
<a href="https://getresulttoday.com/locations/coral-gables/">https://getresulttoday.com/locations/coral-gables/</a>		4 3.6%
<a href="https://getresulttoday.com/locations/plantation/">https://getresulttoday.com/locations/plantation/</a>		1 0.9%
<a href="https://getresulttoday.com/locations/">https://getresulttoday.com/locations/</a>		1 0.9%
<a href="https://getresulttoday.com/biofire-pcr-respiratory-panel-1-5-hours-results/">https://getresulttoday.com/biofire-pcr-respiratory-panel-1-5-hours-results/</a>		0
<a href="https://getresulttoday.com/rt-pcr-covid-test-for-travel/">https://getresulttoday.com/rt-pcr-covid-test-for-travel/</a>		0

Рисунок 4.6 Аналіз нового трафіку

Якщо дивитися статистику по сторінці детальніше на рис 4.5, можна побачити, що даний варіант роботи з текстом приніс більше 45 ключових запитів по сторінці з яких головний високочастотний запит має 4 позицію в пошуку Google. Тим самим зробивши простий інформаційний малотрафіковий текст SEO оптимізованим.

SERP titles ⓘ    Changes: Charts ▾    API    Export

Keywords	Change	Top keyword	Volume	Position
48		monkeypox orlando	3.1K	11 ▲1
47		covid-19 antigen test	300	4
8		covid testing hialeah	600	16 ▼2
10		alliance health sunrise	90	4
12		pcr test pompano beach	70	12
5		alliance health kendall covid testing kendall	30 500	4 17
7		covid testing coral gables	250	16
6		covid testing plantation	250	22
4		alliance health glen cove	450	22
4		what is a biofire test	60	37
6		pcr test miami fast results florida pcr test for travel	30 50	63 58

Рисунок 4.7 Аналіз пошукової видачі

Статті у блозі це лише другорядний контент, який, при правильному використанні залучає додатковий трафік на основну сторінку. Саме тому основним результатом цієї роботи буде підвищення статусу самого сайту у пошуковій видачі. Для перевірки цього був використаний онлайн сервіс Ahrefs.



Рисунок 4.8 аналіз рейтингу сайту

При перегляді статистики на Рис. 4.7 можна побачити, що покращення SEO показників контент складової сайту, а саме правильне використання графу знань призвели до збільшення загального рейтингу сайту.

## ВИСНОВКИ

Кваліфікаційну роботу виконано згідно з методичними вказівками до виконання та захисту кваліфікаційної роботи другого (магістерського) рівня вищої освіти [1].

Метою цього роботи було вивчення практики SEO в сучасному бізнес-середовищі та того, як фахівці з маркетингу використовують його для впливу на графи знань. Дослідження спрямоване на отримання більш глибоких і широких знань про різні методи SEO, а саме про те, чи можуть вони бути використані в створенні графів знань.

Об'єктом дослідження є процес просування сайтів в пошуковій системі з використанням графу знань

Предметом дослідження є методи просування сайтів в пошуковій системі з використанням графу знань

Для виконання поставленої мети в роботі:

- проведено аналіз особливостей семантичного пошуку;
- досліджено методи пошукового просування;
- проаналізовано граф знань;
- порівняно підходи до аналізу графів;
- експериментально перевірено аналіз графу знань для пошукового просування.

## 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. Крітцінгер, В. та Вайдеман, М., 2013. Пошукова оптимізація та маркетингові стратегії з оплатою за клік. *Journal of Organizational Computing and Electronic Commerce*, [online] 23(3), pp.273-286. [Доступно 10 листопада 2020 р.].
3. Озкан, Б., Озчейлан, Е., Кабак, М. и Дагдевірен, М., 2019. Оцінювання веб-сайтів академічних кафедр за критеріями SEO: нерішучий нечіткий лінгвістичний MCDM-підхід. *Artificial Intelligence Review*, [online] 53(2), pp.875-905. Available at: <<https://link.springer.com/article/10.1007/s10462-019-09681-z>> [Дата звернення: 10 листопада 2020 р.].
4. Пох'ялайнен, Р., 2019. Переваги пошукової оптимізації в Google для бізнесу. [Онлайн] *Jultika.oulu.fi*. Available at: <<http://jultika.oulu.fi/files/nbnfioulu-201910112963.pdf>> [Дата звернення: 14 жовтня 2020 р.].
5. Гао, В. та Лі, Х., 2018. Створення присутності на веб-сайті інтернет-магазину: роль якості веб-сайту. *Поведінка та інформаційні технології*, с.28-41.  
<<https://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=5&sid=bed9ae7e-9e83-432e-bd35-36ef5d40f780%40pdc-v-sessmgr02>> [Доступно 14 листопада 2020 р.].
6. Чалая О.В. Информационный подход к обеспечению конкурентоспособности и качества продукции машиностроительного предприятия. *Вісник Харківського національного економічного*

- університету. 2000. №1(13). С. 73–75.
7. Чалая О.В. Структуризация описания процессов при управлении качеством продукции. *Вісник економіки транспорту і промисловості: Зб. наук. праць*. Харків: УкрДАЗТ, 2003. Вип. 4. С. 151–154.
  8. Чалая О.В. Процессно-ориентированный подход к формированию показателей качества. *Вісник економіки транспорту і промисловості: Зб. наук. праць*. Харків: УкрДАЗТ, 2004. Вип. 5–6. С. 289–292.
  9. Чала О.В. Визначення ресурсів до процесів системи управління якістю промислового підприємства. *Економіка: проблеми теорії та практики*. Дніпропетровськ: ДНУ, 2005. Вип. 206. С. 1139–1144.
  10. Чала О.В. Облік витрат на якість при процесному підході до управління підприємством. *Економіка: проблеми теорії та практики. праць*. Дніпропетровськ: ДНУ, 2005. Вип. 207. С. 1403–1407.
  11. Чала О.В. Методичні рекомендації до побудови та аналізу системи управління якістю промислового підприємства. *Економіка: проблеми теорії та практики*. Дніпропетровськ: ДНУ, 2006. Вип. 215. С. 1162–1167.
  12. Чала О.В. Поліпшення процесів системи управління якістю на основі коригувальних і попереджувальних дій. *Вісник економіки транспорту і промисловості*. Харків: УкрДАЗТ, 2006. Вип. 15–16. С. 118–121.
  13. Чала О.В. Удосконалення механізму управління якістю продукції на основі процесного підходу до управління підприємством. *Вісник економіки транспорту і промисловості*. Харків: УкрДАЗТ, 2007. Вип. 18. С. 182–183.
  14. Чала О.В., Чалий С. Ф., Маковоз О.В. Моніторинг процесів в системі управління якістю. *Вісник економіки транспорту і промисловості*. 2009. Вип. 26. С. 213-216.
  15. Chalyi S., Bogatov Ie. Method of constructing an attribute description of the business process "as is" in the process approach to enterprise management *EUREKA: Physics and Engineering*. 2018. Vol. 6. P. 35-40.
  16. Chalyi S., Pribylnova I. The method of constructing recommendations online on

- the temporal dynamics of user interests using multilayer graph. EUREKA: Physics and Engineering. 2019. Vol. 3. P. 13-19. **(SCOPUS)**
17. Чалий С.Ф., Прибильнова І.Б. Ситуаційна модель користувачького вибору в рекомендаційній системі.// Системи управління, навігації та зв'язку. 2019. Вип. 2(54). С.159-163. doi:<https://doi.org/10.26906/SUNZ.2019.2.159>.
18. Chalyi S., Leshchynskiy V., Leshchynska I. Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. EUREKA: Physics and Engineering Vol. 4. P. 34-40. **(SCOPUS)**
19. Chalyi S., Leshchynskiy V., Leshchynska I. Designing explanations in the recommender systems based on the principle of a black box. Сучасні інформаційні системи. 2019. Т. 3, № 2 С. 47-51.
20. Chalyi S., Leshchynskiy V. Knowledge Representation in the Recommendation System Based on the White Box Principle Сучасні інформаційні системи. 2019. Т. 3, № 3. С 82-86.
21. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання пояснень щодо рекомендованого переліку об'єктів з урахуванням темпорального аспекту вибору користувача. Системи управління, навігації та зв'язку, 2019. Том 6 № 58. С. 97-101.
22. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Концепція формування пояснень в рекомендаційних системах за принципом білого ящика. Системи управління, навігації та зв'язку. Збірник наукових праць. Полтава: ПНТУ, 2019. Т. 3 (55). С. 156-160. doi:<https://doi.org/10.26906/SUNZ.2019.3.156>.
23. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Доповнення вхідних даних рекомендаційної системи в ситуації циклічного холодного старту з використанням темпоральних обмежень типу «NEXT». Системи управління, навігації та зв'язку, 2019. Вип. 4(56). С. 105-109.