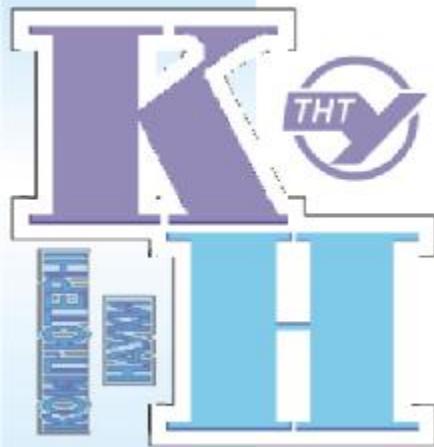


Література

Міністерство освіти і науки України
Тернопільський національний технічний університет ім. Івана Пуллюя
Кафедра комп'ютерних наук

Методичні вказівки до лабораторної роботи №3 з курсу

«Інтернет-технології опрацювання консолідованих
інформаційних ресурсів»



для студентів спеціальності 124 «Системний аналіз»

**Тема: «Дослідження алгоритмів роботи
відомих пошукових систем,
алгоритми ранжування WEB-сайтів»**

Навчально-методична

Тернопіль – 2017

Дослідження алгоритмів роботи відомих пошукових систем, алгоритми ранжування WEB-сайтів. Методичні вказівки до лабораторної роботи для студентів спеціальності 124 «Системний аналіз», кафедра КН ТНТУ ім. І. Пулюя, Тернопіль, 2017 р.

Укладачі:

ст. викл. кафедри КН ТНТУ ім. І. Пулюя Козбур Галина Володимирівна

ст. викл. кафедри АВ ТНТУ ім. І. Пулюя Козбур Ігор Романович

Відповідальний за випуск:

ст. викл. кафедри КН ТНТУ ім. І. Пулюя Козбур Галина Володимирівна

ЗМІСТ

Тема.....	4
Мета роботи.....	4
Короткі теоретичні відомості	
1. ЗАГАЛЬНІ ПОЛОЖЕННЯ ПРО ПОШУКОВІ СИСТЕМИ.....	5
1.1 Визначення пошукових систем	5
1.2 Класифікації пошукових систем	5
1.3 Історія розвитку пошукових систем.....	5
2. БУДОВА І ПРИНЦИПИ РОБОТИ ПОШУКОВИХ СИСТЕМ.....	6
2.1 Структура пошукової системи.....	6
2.2 Моделі пошуку пошукових систем	8
2.3 Оцінка якості пошуку.....	10
3. ПОШУК У МЕРЕЖІ	12
3.1 Обробка запиту пошуковою системою	12
3.2 Підготовка бази даних	14
3.3 Ранжування видачі результатів	15
3.4 Користувальницькі запити.....	16
3.5 Видача результатів	19
3.6 Сайт і пошукова система	21
Завдання.....	22
Хід роботи	22
1. Складання переліку пошукових систем.....	22
2. Класифікація пошукових систем.....	27
3. Регіональні пошукові системи.....	30
4. Визначення рейтингів пошукових систем	31
5. Проведення аналізу основних характеристик та наявних сервісів пошукових систем	41
6. Проведення аналізу можливостей веб-аналітики для заданих пошукових систем	43
ЛІТЕРАТУРА	46
ДОДАТОК 1. ОСНОВНІ ПОШУКОВІ СИСТЕМИ	47
ДОДАТОК 2. МОВА ПОШУКОВИХ ЗАПИТІВ	48
ДОДАТОК 3. GOOGLE ОПЕРАТОРИ В ПОШУКОВИХ ЗАПИТАХ	52
ДОДАТОК 4. YANDEX ОПЕРАТОРИ В ПОШУКОВИХ ЗАПИТАХ	54
Зразок титулки.....	59

Лабораторна робота №3

Тема: «Дослідження алгоритмів роботи відомих пошукових систем, алгоритми ранжування WEB-сайтів»

Мета роботи:

- дослідження принципів та алгоритмів роботи відомих пошукових систем. Алгоритми ранжування WEB-сайтів в пошукових системах;
- вивчення класифікації пошукових систем;
- складання рейтингу пошукових систем.

Звіт повинен містити:

- титульну сторінку, виконану за наведеним зразком;
- мету роботи;
- хід роботи (за пунктами 1 – 6);
- висновок.

Використовувані Інтернет-ресурси вказувати безпосередньо в ході роботи, де необхідно.

ЗВІТ ДО ЛАБОРАТОРНОЇ РОБОТИ ОФОРМИТИ В MS WORD ЗГІДНО З ВИМОГАМИ НОРМОКОНТРОЛЮ.

ПАПЕРОВИЙ ВАРИАНТ ЗВІТУ У ФОРМАТІ А4 ЗДАТИ ВИКЛАДАЧЕВІ ЗГІДНО З ГРАФІКОМ ЗДАЧІ ЛАБОРАТОРНИХ РОБІТ, ЕЛЕКТРОННИЙ У ФОРМАТІ lab3_ГРУПА_Прізвище.docx – ВІДПРАВИТИ В СКРИНЬКУ ЗАВДАНЬ ЕЛЕКТРОННОГО НАВЧАЛЬНОГО КУРСУ.

Короткі теоретичні відомості

1. ЗАГАЛЬНІ ПОЛОЖЕННЯ ПРО ПОШУКОВІ СИСТЕМИ

1.1 Визначення пошукових систем

Пошукова система – це програмно-апаратний комплекс, що здійснює пошук у мережі Інтернет необхідної інформації.

Пошукові системи розділяють на дві частини: front-end і back-end. Під front-end'ом мають на увазі веб-інтерфейс, через який здійснюється взаємодія користувача з back-end'ом. Back-end – це пошукова машина, що здійснює пошук у мережі Інтернет для повернення результатів пошуку їхньому користувачеві в front-end складову.

1.2 Класифікації пошукових систем

Розділяють кілька варіантів класифікації пошукових систем. Найчастіше використовують класифікацію по класу, але іноді застосовують класифікацію, засновану на призначенні пошукової системи.

а) класифікація пошукових систем по класу

Виділяють два основних класи пошукових систем:

- **каталоги** – це структурований набір посилань на веб- сайти з їхнім коротким описом. Вони мали широке поширення раніше, коли кількість ресурсів у мережі була порівняно невеликою.
- **показчики** (так називають пошукові машини, дослівно від Search Engines) – це пошукові системи, котрі реалізують пошук за ключовими словами.

б) класифікація пошукових систем по призначенню

За призначенням пошукові системи розділяють на:

- **тематичні** – цей вид пошукових систем працює з обмеженим контентом і виконує пошук інформації в інтернеті тільки серед певних ресурсів, наприклад, відповідно до релігійних або професійних переваг. Так, існує варіант пошуковика Google, орієнтований на ортодоксальних євреїв, він працює винятково з дозволеним для цієї релігії контентом.
- **спеціалізовані** – це пошукові системи для роботи з певною інформацією, наприклад, пошук по онлайн-магазинах, FTP-серверах та інші.
- **універсальні** – пошук у таких системах організований незалежно від виду контенту. Користувач може запитувати текстову інформацію, графічний, відео- або аудіо-контент. Безпосередньо пошук охоплює всі ресурси інтернету без обмежень. Всі провідні пошукові системи (Google, Яндекс, Yahoo!, Bing і т.д.) є універсальними.

1.3 Історія розвитку пошукових систем

Першою пошуковою системою був «Wandex», створений в 1993 році, що здійснював пошук по сайтах мережі за допомогою пошукового бота. В результаті отримували список відповідних до запиту ресурсів. Ранжування не

здійснювалось. Слід зазначити, що всього в інтернеті того часу було близько 620 веб-ресурсів, а сам пошук здійснювався тільки за певними словами.

Наступною примітною повнотекстовою пошуковою системою став WebCrawler, запущений в 1994 році, що, на відміну від своїх попередників, дозволяв користувачам шукати за будь-якими ключовими словами на будь-якій веб-сторінці – з тих пір це стало стандартом у всіх основних пошукових системах.

Першою революційною пошуковою системою став BackRub. Він уперше почав враховувати зовнішні посилання на сайт і ранжувати відповідно до цього. Її творцями були Ларрі Пейдж і Сергій Брин, які двома роками пізніше заснували Google. Попередники BackRub ранжували сайти по кількості відвідувань, що не завжди було правильним критерієм корисності ресурсу.

В 1996 році кириличним користувачам інтернету стало доступно морфологічне розширення до пошукової машини Altavista й оригінальні російські пошукові машини «Рамблер» й «Апорт», а в 1997 була відкрита пошукова машина Яндекс.

В 1998 році запустилася пошукова система MSN, що працювала на чужій пошуковій видачі.

Поступово із загальної маси схожих пошукових систем починають виділятись найуспішніші, починається низка злиттів і поглинань, з якої виділяються Google і Yahoo! Пошукові гіганти старанно обростають додатковими сервісами, щоб прив'язати користувача до себе.

У боротьбу вступає Microsoft зі своїм LiveSearch, що виходить в 2006 році, із цього моменту Microsoft сам формує свою видачу.

У тому ж році Yahoo! відмовляються від використання пошукової видачі Google і починають самостійно формувати пошукову видачу.

В 2009 році Microsoft анонсує нову пошукову систему Bing, що прийшла на заміну застарілому LiveSearch. У 2010 році компанія Microsoft підписала контракт із Yahoo! про те, що остання буде використовувати пошукову видачу системи Bing наступних 10 років.

За станом на 2016 рік частка пошукових систем на світовій арені була наступна:

1. Google – 83,87%;
2. Yahoo! – 6,20%;
3. Bing – 3,69%;
4. ASK Jeeves – 0,34%;
5. Baidu – 0,63% (Китайська національна пошукова система).

2. БУДОВА І ПРИНЦИПИ РОБОТИ ПОШУКОВИХ СИСТЕМ

2.1 Структура пошукової системи

2.1.1 Веб-сервер пошукової системи

Веб-сервер пошукової системи – це веб-сервер, що здійснює взаємодію між користувачем й іншими компонентами пошукової системи. Як правило, він забезпечує роботу веб-інтерфейсу (рис. 1) пошукової системи (сайту). Веб-

сервер також тримає на собі сторінку видачі результатів користувачеві у вигляді html-сторінки.

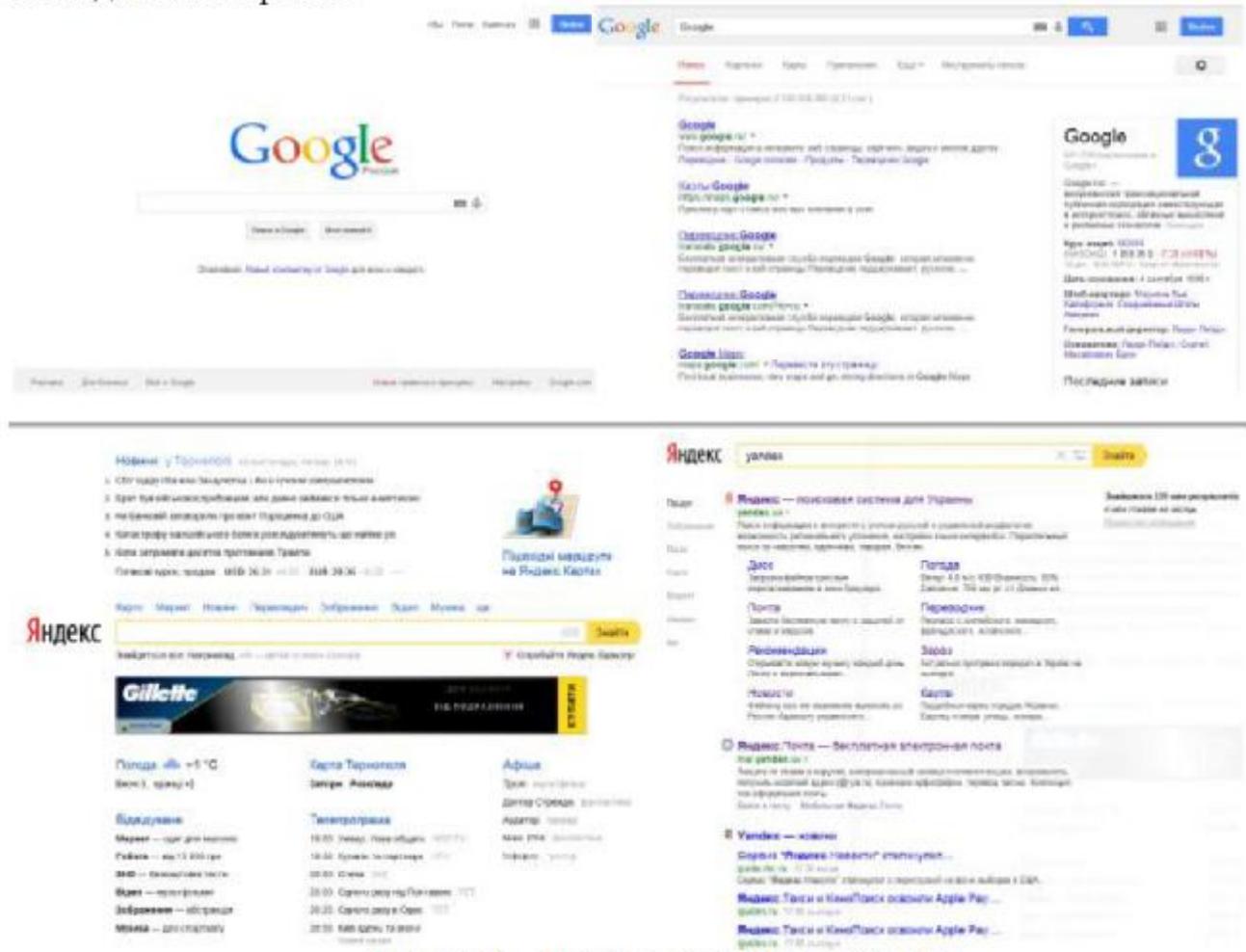


Рисунок 1 – Веб-інтерфейс Google, Yandex

2.1.2 Пошукові роботи

Пошуковий робот – це програма, що є складовою частиною пошукової системи й призначена для перебору сторінок Інтернету з метою занесення інформації про них у базу даних пошуковика. За принципом дій робот нагадує звичайний браузер. Він аналізує вміст сторінки, зберігає його в деякому спеціальному вигляді на сервері пошукової машини й відправляється по наступних посиланнях.

Пошукові роботи звичайно підрозділяються на «Планувальників» й «Павуків».

Планувальник (Crawler або «Плазун, Гусінь, Черв'як») – це програма, основна мета якої – витягати з ресурсу всі посилання на інші сторінки/ресурси й зберігати їх у своїй базі (процес індексування), одночасно створюючи шлях для «Павука». Після того як всі посилання будуть знайдені, «Планувальник» починає планомірний переход за цими посиланнями і весь процес розпочинається спочатку.

Павук (Spider) – це головна частина пошукового робота. Він переходить за спланованим «Планувальником» маршрутом й зберігає весь код сторінки в базі даних.

У різних пошукових системах пошукові роботи можуть відрізнятися, але принцип дії у всіх схожий – знайти контент і скачати його.

Існують підкатегорії пошукових роботів, які займаються пошуком картинок або переглядають швидкооновлювані сайти, інші перевіряють сайти на працездатність і т.д.

2.1.3 Індексатор

Індексатор – це програма, що здійснює обробку завантажених пошуковими роботами сторінок та виконує їхній аналіз. Як правило, індексатор розбиває сторінку на кілька частин й аналізує кожну з них. Індексатор призначений для додавання інформації про сайти в регулярно оновлювану базу даних (індекс) пошукової системи.

У результаті індексації створюється група індекс-файлів сторінки, у якому кожній пропозиції відповідає свій номер, а кожному слову – номер у пропозиції. Також у цих файлах містяться спеціальні теги, що допомагають правильно співвідносити елементи на сторінці. При цьому створюється так звана «область індексування» – це унікальний префікс, призначений для кожного проіндексованого сайта. Індекс-файл – це оптимальне рішення в пошукових системах, що допомагає заощаджувати багато дискового простору. Ще більшої економії можна досягти, використовуючи спеціальні алгоритми стиснення файлів.

Індексатори також вибудовують зворотний індекс – це коли слову відповідають всі б-ресурси. Індексатор, використовуючи зворотні і прямі індекси, здійснює ранжування залежно від релевантності ресурсу.

2.1.4 База даних

База даних – це сховище всіх даних, які пошукова система завантажує й аналізує. Іноді базу даних називають індексом пошукової системи, оскільки саме в ній зберігаються індекс-файли кожного ресурсу, обробленого системою. Вона постійно обновлюється й поповнюється службовими програмами, – пошуковими ботами й індексаторами.

2.2 Моделі пошуку пошукових систем

Модель пошуку – комплекс, на підставі якого виходить формула, що дозволяє програмі прийняти рішення: який документ уважати знайденим та як його ранжувати.

Основні моделі пошуку розділяють на три типи:

- теоретико-множинні;
- імовірнісні;
- алгебраїчні.

2.2.1 Теоретико-множинні моделі

В основі теоретико-множинних моделей лежить ідея про те, що зміст документів може бути виражений безліччю ключових слів, що представляють собою слова або словосполучення природної мови з текстів документів.

Булева модель пошуку і її розширення

У булевій моделі пошуку запит формується у вигляді логічного виразу, що використовує оператори булевої логіки: І, АБО, НІ. Модель орієнтована на пошук по тексту, де критерієм релевантності буде відповідність результату пошуковому запиту. Документ, метадані якого співпадають зі значеннями, заданими в запиті, буде вважатися релевантним.

У розширеній булевій моделі використовуються не чіткі бінарні величини, а відносні, які називають ваговими коефіцієнтами, значення яких лежать у діапазоні [0; 1]. Це надає даній моделі більшої гнучкості й дозволяє здійснювати ранжування відібраних документів.

Плюсами булевої моделі пошуку є простота реалізації, недоліком – складність для звичайного користувача, оскільки далеко не кожний зможе грамотно справлятися з операторами булевої логіки при формуванні свого запиту.

2.2.2 Імовірнісні моделі

Основна ідея таких моделей: найвища ефективність пошуку досягається тоді, коли документи ранжуються за спаданням ймовірності релевантності запиту. Для обчислення ймовірності використовується припущення, що певні ключові слова мають більшу чи меншу ймовірність релевантності для конкретної тематики.

Мережі виводу

Мережі виводу базуються на принципах так званої мережі Байєса, що побудована на чотирьох ключових вузлах. Перший вузол – це документи, вивчені користувачем у процесі попередніх пошуків. Другий вузол – це список ключових слів, що розміщені на сторінці. Третій вузол – це запити, що складаються із цих ключових слів. Четвертий вузол – це невідомі заздалегідь пошукові системі потреби користувача. Всі вузли першого й другого типів формуються заздалегідь для заданої колекції. Вузли третього типу і їхній зв'язок з вузлами першого й четвертого типу формуються щораз при відправленні запиту користувачем.

При одерженні запиту відбувається обчислення показника інформаційної потреби користувача, потім для кожного документа здійснюється імовірнісна оцінка релевантності запиту щодо потреб користувача. На підставі цієї оцінки відбувається ранжування ресурсів.

До плюсів даної моделі можна віднести:

- гарне теоретичне обґрунтування;
- при наявній інформації отримується краще ранжування відповідей.

До мінусів можна віднести:

- необхідність інформації про релевантність або її наближені оцінки;
- структура документа описується тільки ключовими словами;
- для отримання оптимальних результатів необхідне постійне машинне навчання на основі інформації про релевантності.

2.2.3 Алгебраїчні моделі

Векторно-просторові моделі

Векторні моделі є найпоширенішими на практиці моделями. Векторні моделі, на відміну від булевих, дозволяють легко ранжувати релевантні запити. Суть таких моделей зводиться до того, що документ у векторній моделі розглядається як невпорядкована множина позначень документа, де позначення – це всі слова, дати, назви й т.д., які формують документ.

Вага кожного позначення найчастіше визначається підрахунком кількості його повторень у тексті.

Всі позначення колекції релевантних документів упорядковуються. Для кожного документа складається вектор із всіх позначень, що є поданням даного документа у векторному просторі:

$$\mathbf{d}_j = (w_{1j}, w_{2j}, \dots, w_{nj}),$$

де \mathbf{d}_j – векторне подання j -го документа, w_{ij} – вага i -го терму в j -му документі, n – загальна кількість різних термів у всіх документах колекції.

Очевидно, що розмірність вектора й векторного простору однакова для всієї колекції.

Запит розкладається на позначення, потім складається вектор, після чого у видачі здійснюється ранжування колекції релевантних ресурсів відповідно до їх близькості до вектора запиту.

Переваги:

- оцінка вагових коефіцієнтів підвищує ефективність пошуку;
- метод дозволяє оцінити ступінь відповідності документа запиту;
- використання зручної для ранжування косинусної метрики.

Недоліки:

- немає достатнього теоретичного обґрунтування для побудови простору позначень;
- оскільки позначення не є незалежними одне від одного, то вектори не можуть бути повністю ортогональними.

Має переваги перед іншими моделями через простоту й добротність.

2.2.4 Моделі пошуку в існуючих пошукових системах

Насправді жодна з лідеруючих пошукових систем ні в кириличному сегменті інтернету, ні у всесвітній павутині, не користується лише однією з цих моделей пошуку. Яндекс й Google використовують синтез із всіх трьох перерахованих моделей спільно зі своїми наробками.

2.3 Оцінка якості пошуку

Коректна оцінка якості пошуку є одним із ключових факторів успішності пошукової системи. Оцінку якості пошукової видачі роблять на підставі безлічі критеріїв, а також за допомогою оцінки якості пошуку людьми спеціальної посади – **асесорами** (лат. *assessor* — засідатель, оцінщик).

2.3.1 Критерії оцінки якості роботи пошукової системи

- **Точність видачі** – відношення числа виданих релевантних документів до суми числа виданих релевантних і виданих нерелевантних документів.
- **Повнота видачі** або **чутливість** – відношення числа виданих релевантних документів до суми числа виданих релевантних і не виданих релевантних документів.
- **Втрати інформації** – відношення числа не виданих релевантних документів до суми числа виданих релевантних і не виданих релевантних документів.
- **Інформаційний шум** – відношення числа виданих нерелевантних документів до суми числа виданих релевантних і виданих нерелевантних документів.
- **Специфічність** – відношення числа невиданих нерелевантних документів до суми числа виданих нерелевантних і невиданих нерелевантних документів.

На практиці для порівняння пошукових систем використовуються усереднені графіки залежності повноти від точності. Щоб уникнути порівняння повнота–точність, використовуються однозначні оцінки. Однією з таких оцінок є Е-міра, а саме міра відношення їхньої значимості:

$$E(b) = 1 - \frac{(b^2 + 1) \cdot P \cdot R}{b^2 P + R},$$

де P – точність, R – повнота, b – відношення значимості повноти й точності.

2.3.2 Машинне навчання

Машинне навчання зараз активно використовується в Яндексі, називають таку систему Матрікснет. З її допомогою пошукова машина може враховувати набагато більше критеріїв релевантності, з помітно кращою якістю пошуку, що тільки укріплює позиції пошуковика в російському сегменті Інтернету.

Машинне навчання практикується й в Google. Програма називається Seti. За словами розроблювачів, поки що вони не готові повністю довірити ранжування програмі, тому що машинне навчання не здатне на 100% адекватно діяти в нестандартній ситуації. Також потрібно враховувати ширину охоплення пошукової системи Google, що незрівнянно більша порівняно з Яндексом.

Слід зазначити, що Google у співпраці з NASA в травні 2013 року запустив проект Quantum Artificial Intelligence Lab (QAIL), основним завданням якого є дослідження можливості використання квантового комп'ютера в машинному навчанні. Це дослідження здатне зробити революцію в машинному навчанні, використовуючи машину з нереальною, для сьогоднішніх суперкомп'ютерів, швидкістю здійснення складних обчислень.

2.3.3 Асесори пошукових систем

Асесори пошукових систем – працівники у штаті компанії-пошуковика. Вони виконують роботу по складанню навчальної вибірки для програм машинного навчання, а також для оцінки якості пошуку. Вони перші, хто тестиє нові зміни в алгоритмі пошуку й тільки після успішного проходження тестування асесорами, алгоритм проходить наступні стадії тестування.

3. ПОШУК У МЕРЕЖІ

3.1 Обробка запиту пошуковою системою

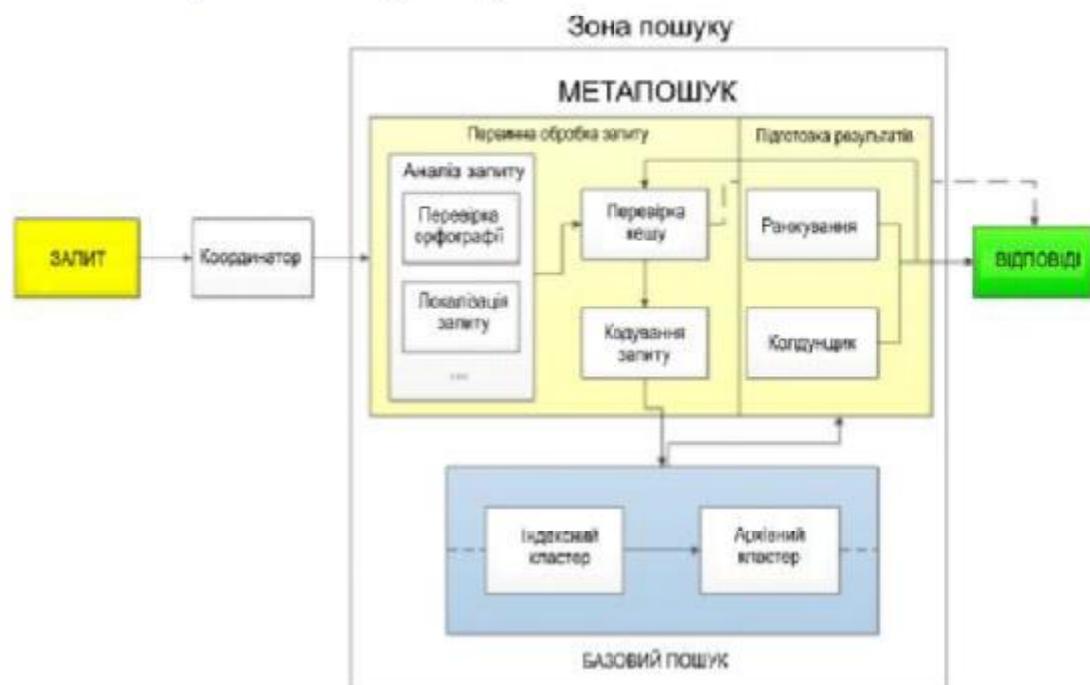


Рисунок 2 – Шлях від запиту до пошукової видачі

На рис.2 представлена дещо спрощена схема здійснення пошуку, на якій схематично зображені наступні основні етапи обробки запиту:

- **координатор** направляє запит на найближчий і менш завантажений сервер;
- **метапошук** аналізує запит, намагається зрозуміти його суть, перевіряє, чи не був цей запит уже оброблений раніше, ранжує знайдені на етапі базового пошуку результати;
- **базовий пошук** здійснює пошук релевантних результатів по базі даних.

Координатор не має складених локальних частин, а його функціонал одноманітний, тому він не представляє особливого інтересу з погляду логічної структури. Тому детально розглядати будемо тільки метапошук і базовий пошук.

3.1.1 Метапошук

Метапошук можна умовно розділити на два логічних відділи, у яких відбувається:

- первинна обробка запиту;
- підготовка результатів.

У відділі первинної обробки відбувається аналіз запиту і його інтерпретація. Також перевіряється, чи не був цей запит раніше оброблений системою. У випадку якщо такий запит уже оброблявся недавно, те виводяться його результати, інакше запит вважається новим. Такий запит спеціально кодують і передають у базовий пошук.

Під час аналізу запиту відбувається:

- визначення мови;
- аналіз запиту на помилки;
- запит розбирається за словами, аналізується зміст кожного слова і сукупний зміст, перебираються можливі трактування й відбирається найімовірніший варіант;
- запит класифікується;
- визначаються синонімічні варіанти цього ж запиту;
- визначається його географічне позиціонування.

Вище перераховані тільки основні процеси, що відбуваються під час аналізу.

Результати, отримані в результаті базового пошуку, відправляються в умовний відділ підготовки результатів. У ньому відбуваються дві дії:

- документи ранжуються за ступенем релевантності запиту спеціальним алгоритмом;
- спеціальна програма, іменована «колдунищю», «шідмішє» у пошукову видачу результати пошуку з інших сервісів, що відповідають запиту. Що саме буде додавати «колдунщик» – залежить від наявності результатів за цим пошуком, а також від особливостей самої пошукової системи.

«Колдунище» – спеціальний модуль Яндексу, що видає відповідь на ваш запит вище від результатів пошуку. Це може бути погода, час, переклад слова, розклад поїздів й інша інформація, яка, на думку пошуковика, вже є відповіддю на ваш запит і тому видається вище посилань на проіндексовані сайти (рис.3).

The screenshot shows the Yandex search interface. At the top, the search bar contains the query "который час в москве" with a note "(246 млн ответов)". To the right of the search bar are buttons for "Найти" (Search) and "Изменить" (Change). Below the search bar, there's a sidebar with links like "Поиск", "Картинки", "Видео", "Карты", "Новости", and "Словарь". The main content area displays the search results. The first result is a snippet from 24timezones.com with the heading "Часы Москва - местное время. Который час в г. Москва...". This snippet includes the text "Москва — 14:37" which is circled in red. To the right of this snippet, there are additional links: "Индекс ДД", "Дизайн", "Ищете", and "Новости". Below the main result, there are other links such as "Авто", "Адреса", "Афиша", "Вакансии", "Валютный калькулятор", "Видео", and "Википедия" on the left; and "Картинки", "Карты", "Конвертер", "Котировки", "Маркет", "Музыкальный плеер", and "Новости" on the right. At the bottom right, there are more links: "Отключение воды", "Погода", "Психологическая помощь", "Расписания", "Стихолюб", "Телепрограмма", and "IP-адрес пользователя".

Рисунок 3 – Модуль Яндексу «Колдунчию»

Функції пошуку Google

У Google, природно, є аналогічний функціонал, що дозволяє дізнатися погоду, час, різні міри, переклад слів і розклад кіносесансів, здійснювати пошук авіаквитків тощо (рис. 4). Google працює в цій площині дещо в іншому контексті, ніж Яндекс. Його вдосконалення в основному стосуються інтеграції з мобільними пристроями, соціальними мережами, розвитку голосового пошуку та інших інновацій, потреб, що відображають інтереси користувачів (наприклад AdWords, Froogle (google.com/shopping)).

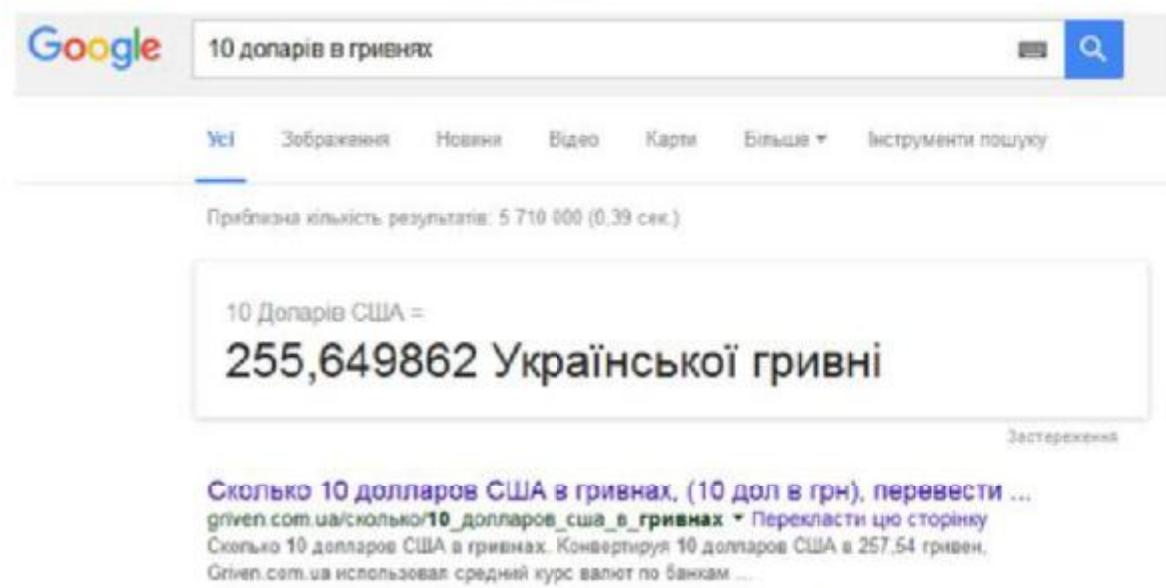


Рисунок 4 – Модуль Google

3.1.2 Базовий пошук

Базовий пошук здійснюється у випадку, якщо запит вважається новим. Базовий пошук ділиться на два етапи:

- пошук за індексною базою даних;
- пошук за архівною базою даних.

З індексної бази даних за допомогою пошукового алгоритму системи «витягають» ідентифікатори документів, а потім, за цими ідентифікаторами, «витягають» дані з архівної бази у вигляді короткого попереднього перегляду ресурсу в пошуковій видачі. В архівній базі даних також перебуває остання копія сторінки, яку обробляв пошуковик.

3.2 Підготовка бази даних

Індексація

Індексація веб-ресурсу пошуковою системою відбувається після завантаження контенту ресурсу пошуковими роботами. Робить індексацію відповідна програма – індексатор. Визначають два види індексу: прямий та інвертований (зворотній).

Процес складання інвертованого (зворотного) індексу відбувається приблизно в такий спосіб:

- завантажена сторінка очищається від мови розміток, тобто від тегів html, вставок коду JavaScript й інших службових інструкцій;
- складається прямий індекс. Реалізації можуть бути різні, але принцип залишається незмінним. Звичайно зі всіх слів, розділових знаків й т.д. створюють масив;
- текст сторінки, перетворений відповідно до правил машинної морфології, тобто приведений до початкової форми з відкиданням «не слів», вибудовується за абеткою. Слова, які не вдалося привести до первозданного виду, наприклад, слова з помилками або нові слова й т.д., залишаються без змін;
- складається інвертований індекс сторінки, тобто для кожного слова знаходяться всі його входження в оригінальний текст;
- зворотний (або інвертований) індекс кожного обробленого ресурсу використовується системою для складання глобального зворотного індексу, у якому кожному слову відповідають сайти й місця входження цього слова в текстах сайтів. Наявність у зворотному індексі позиції входження слова допомагає краще ранжувати ресурси у видачі, оскільки так можна визначати порядок слів на сторінці, що дозволяє коректно обробляти багатослівні запити.

3.3 Ранжування видачі результатів

3.3.1 Алгоритми ранжування сайтів

Алгоритми ранжування сайтів – це система формул, за допомогою яких пошукова система здійснює ранжування сайтів у пошуковій видачі. Повний опис алгоритмів ранжування пошукових систем та критеріїв, які вони враховують для розрахунку релевантності, перебуває у найсуворішому секреті.

3.3.2 Критерії релевантності

Критерії, які враховує пошукова система для розрахунку релевантності сторінки, перебувають, як уже було сказано, у найсуворішому секреті. Однак власникам сайтів удалось емпіричним шляхом виявити деякі закономірності, що впливають на позиції веб-ресурсів у пошуковій видачі. Серед них:

- індекс цитування (кількість зовнішніх посилань на сторінку. Чим авторитетніший сайт, з якого посилаються, тимвищий коефіцієнт релевантності та ранг сторінки);
- унікальність контенту;
- кількість відвідувачів;
- відсоток відмови (занадто швидкий відхід користувача зі сторінки);
- частота знаходження ключового слова на сторінці;
- частота відновлення ресурсу;
- як часто переходять на сайт зі сторінки пошукової видачі;

- наявність та кількість слів однієї тематики із ключовим словом на сторінці. Тобто, визначається, чи не є ресурс сміттєвим спам-сайтом із простим набором ключових слів;
- географія відвідувачів;
- вік сайту й релевантних сторінок;
- внутрішній код сторінки. Чим він ясніше й вірніше, тим вищий ранг;
- купівля або продаж посилань. Фактор, що знижує ранг у видачі.

Факторів ранжування в пошукових системах може бути дуже багато. Наприклад, розроблювачі Google пишуть, що їхні алгоритми враховують більше 200 критеріїв, а розроблювачі Яндекса рапортують про більш ніж 500 критеріїв, які враховує їхня система, побудована на машинному навчанні й імовірнісній моделі пошуку.

3.4 Користувальницькі запити

3.4.1 Загальні положення

Пошуковий запит – це слово, фраза, вираз, що вводить користувач у пошуковий рядок пошукової системи з метою одержання списку сайтів, у яких дане слово, фраза, вираз повторюється.

За **популярністю** запити поділяють на:

- низькочастотні (НЧ) – запити, що найрідше вводяться користувачами в пошуковий рядок за певною тематикою;
- середньочастотні (СЧ) – запити, створювані із середньою частотою;
- високочастотні (ВЧ) – найчастіше створювані запити.

Запити також розрізняють за **конкурентністю**, тобто за кількістю сайтів, які відповідають на даний запит:

- низькоконкурентні – на запит відповідає мала кількість сайтів;
- висококонкурентні – на запит відповідає велика кількість сайтів. Це найскладніші запити, оскільки безліч ресурсів праґнуть в результаті запиту виявитися в «топі» пошукової видачі.

3.4.2 Класифікація за видом

Розрізняють такі види користувальницьких запитів (рис.5):

- транзакційні (TRNS);
- навігаційні (NVGT);
- інформаційні (INFO).

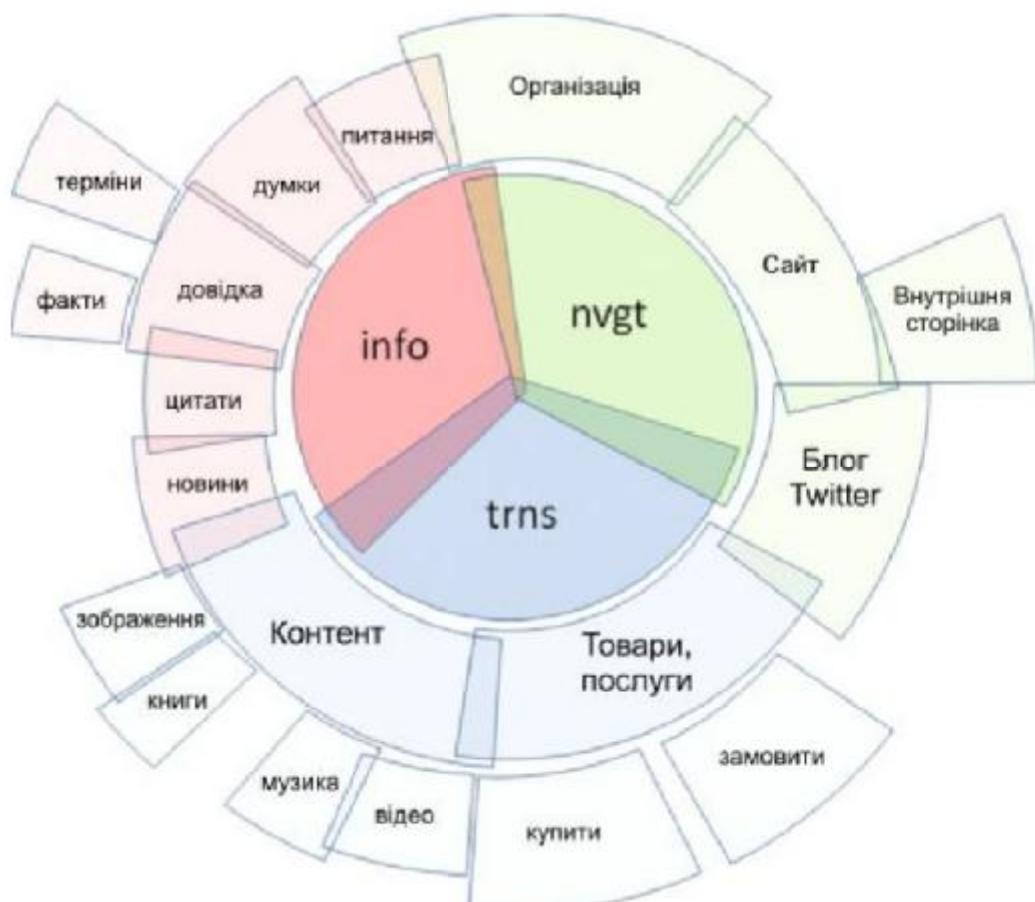


Рисунок 5 – Класифікація запитів

Транзакційні запити (TRNS) формуються користувачами з метою зробити яку-небудь подальшу дію, тобто що-небудь купити, продати, скачати й т.д. Наприклад: «купити кондиціонер», «скачати фільм такий-то» і т.д.

Навігаційні запити (NVGT) формуються з метою знайти конкретне місце у світовій мережі Інтернет – сайт компанії, що цікавить, тематичний форум або дошку оголошень. Наприклад: «сайт уряду», «офіційний сайт компанії Тернопільгаз» і т.д.

Інформаційні запити (INFO) – найбільш широкий клас запитів, до якого відносять все інше: новини, довідкові запити по фактах і термінах, пошук думок і відгуків, відповіді на питання тощо. Наприклад: «найвища гора у світі», «рейтинг фільмів 2016» і т.д.

3.4.3 Обробка запитів

3.4.3.1 Парсер

Парсер запитів - спеціальна програма для первинної обробки запиту, яка містить чергу агентів, кожен з яких робить яку-небудь елементарну дію над запитом, починаючи від розбивки його на слова чи інші регулярні змістовні типи (як, наприклад, URL-и і телефонні номери), і закінчуючи складними класифікаторами типу детектора штатів.

Кожен запит проганяється через цю чергу, причому результати роботи кожного агента впливають на подальший логічний шлях запиту. Наприклад, запит, класифікований як навігаційний, не потрапить на обробку детектора

цитат. Сам запит тут представляється як дошка оголошень, на якій кожен агент може залишити своє повідомлення, прочитати повідомлення інших агентів і за необхідністю змінити або скасувати їх. На завершення запит передається додатковим програмам і остаточно відправляється на подальшу обробку, де всі записані агентами додаткові параметри використовуються як фактори ранжування.

3.4.3.2 Словники

Словники допомагають парсерам обробляти запити (зрозуміти їх зміст). Це спеціальні файли, що містять згруповані масиви рядків, кожний з яких має певний сенс. Кожний із цих груп слів зіставляється маркер. Його наявність означає, що у випадку знаходження в запиті відповідного фрагмента, всьому запиту буде поставлена мітка, котра вказує на предмет цього запиту, наприклад «кіно», «музика».

У цих же словниках задаються синоніми слів за даною тематикою. Якщо, приміром, у запиті є слово «саундтрек», запит буде розширений словосполученнями «звукова доріжка», «саунд трек» й абревіатурою «OST».

Словники поділяють на два класи:

- частотні вирази-маркери, що характеризують певну тематику або намір користувача. Такі словники складаються вручну;
- об'єкти (фільми, рецепти, знамениті особистості, географічні назви й т.д.).

Таких словників існує кілька десятків. Розмір деяких з них – сотні тисяч, навіть мільйони подібних рядків. На їхній основі відбувається первинна обробка запиту, тобто визначення словника, до якого відносяться слова із запиту.

3.4.3.3 Маркери

Маркери, сформовані парсерами і словниками, допомагають пошуковій системі зрозуміти зміст запиту. З допомогою парсерів запит розбирається на слова й словосполучення. Ці слова, словосполучення, а також весь запит цілком пропускаються через словники. У результаті одержують ряд відповідностей, що визначають, до якої тематики відносяться частини запиту або запит цілком.

У випадку входження одного маркера в інший верхній маркер «поїдає» внутрішній. Наприклад, запит «пірати Карибського моря» буде мати три маркери: «пірати» – як назва, професія, «Карибського моря» – як ще одна назва, географія, і «пірати Карибського моря» – як фільм. У результаті запит буде визначений, у більшості випадків, як фільм «Пірати Карибського моря».



Рисунок 6 – Маркери

3.4.4 Персоналізація запитів

Персоналізація – це один зі способів видати користувачам саме ті результати, на які вони очікують.

Наприклад, відповідь на запит “Північне сяйво” для різних людей буде різною. Мандрівниківі покажуть відповідь про природне явище; жителю міста, що цікавиться покупками, – торговий центр; аматорові кіно – посилання на інформацію про фільм.

Здійснюють персоналізацію на підставі історії пошуку та місця розташування користувача. Деякі пошукові системи додають категорії, що впливають на персоналізацію. Активно використовуються файли Cookies попередніх сесій у браузері.

Методи збору пошукової історії в пошукових системах схожі:

- використання спеціальних розширень для браузерів (або їх вбудовування у власні браузери пошуковиків, наприклад, GOOGLE – в GoogleChrom, YANDEX – в YandexBrowser, BING – в IE Explorer, Edge, MailRU – в Amigo);
- прив'язка історії до аккаунтів користувачів в інших ресурсах пошуковика, таких як: пошта, соціальна мережа й т.д.

Визначення місця розташування користувача допомагає правильно ранжувати результати видачі. Наприклад, жителеві Тернополя за запитом «викликати електрика» будуть показані сайти, релевантні запиту «викликати електрика Тернопіль» (або «... у Тернополі», або «Тернопіль, викликати електрика», і т.п.)

3.5 Видача результатів

Коли пошук уже зроблений і його результати відсортовані за релевантністю, вони повертаються у веб-інтерфейс пошукової системи.

Пошукова видача складається з посилань на релевантні ресурси й сніпети, також її результати доповнюються спеціальними роботами («колдунищами»).

3.5.1 Сніпети

Сніпети – це невеликий шматок текстової інформації, що виводиться поруч із посиланням у пошуковій видачі. Інакше кажучи, цей короткий опис сторінки сайту, релевантної пошуковому запиту.

У якості сніпетів використовуються звичайно шматки тексту із цього документа. Цей спосіб покликаний надати користувачеві можливість скласти уявлення про вміст сторінки, не переходячи на неї.

The screenshot shows the Yandex search interface with the query 'Эверест' entered in the search bar. The results are organized into two main sections: 'Сниппеты' (Snippets) on the left and 'Объектный ответ' (Object Answer) on the right.

Сниппеты:

- Джомолукма — Википедия**: A snippet from Wikipedia about Mount Everest's name origin.
- Эверест (2015) смотреть онлайн бесплатно**: A snippet for the movie 'Everest' with a link to the trailer.
- Гора Эверест — самая высокая точка нашей планеты...**: A snippet about Mount Everest's height and record.
- Эверест (Everest, 2015). Кинопоиск.Ру**: A snippet for the movie 'Everest' from Kinopoisk.Ru.
- Эверест — 15 тыс. видео**: A snippet for video results related to Mount Everest.

Объектный ответ:

- Эверест**: An object answer card showing the peak of Mount Everest, its height (8,848 m), and first ascent date (1953).
- Смотрите также**: Recommendations for other peaks: Лайсан, Базовый лагерь Джомолукма, and Чогори.

Annotations on the page include:

- A yellow box highlights the 'Сниппеты' section.
- A yellow box highlights the 'Объектный ответ' section.
- A yellow arrow points from the 'Сниппеты' section to the 'Смотрите также' section.
- A yellow arrow points from the 'Сниппеты' section to the 'Кодирующий' section.
- A yellow arrow points from the 'Сниппеты' section to the 'Параллельный поиск' section.
- A yellow arrow points from the 'Сниппеты' section to the 'Другие поисковые системы' section.
- A yellow arrow points from the 'Сниппеты' section to the 'Кодирующий' section.

Приблизна кількість результатів: 2 220 000 (0.48 сек.)

Google сніппет

Сніппет — Вікіпедія

<https://uk.wikipedia.org/wiki/Сніппет>

Сніппет (англ. snippet — фрагмент, уривок) в практиці програмування — невеликий фрагмент вихідного коду або тексту, придатний для повторного ...

Огляд · Приклади · Можливості програмних сніппетів

Сніппет — Вікіпедія

<https://ru.wikipedia.org/wiki/Сніппет> · Перекласти цю сторінку

Сніппет (англ. snippet — отрывок, фрагмент) — фрагмент исходного текста или кода программы, применяемый в: поисковых системах, текстовых ...

Обзор · Возможности программных сніппетов

Рисунок 7 – Структура одного результату на сторінці видачі

Сніппет формується автоматично, а от які саме фрагменти тексту будуть використовуватись в ньому – вирішує алгоритм, причому для різних запитів в одній тій же веб-сторінки будуть різні сніппети. Сніппет вибирається із прямого індексу пошукової системи.

У додаткові опції видачі іноді додають перегляд збереженої копії інтернет-сторінки, що вже обробив пошуковик.

3.6 Сайт і пошукова система

3.6.1 Ключові слова

Ключове слово – це слово в тексті, здатне в сукупності з іншими ключовими словами відображати зміст або мету сайту. У WEB використовується головним чином для пошуку.

Набір ключових слів документа називають пошуковим образом документа. Набір ключових слів близький до анотації, плану й конспекту, які теж представляють документ із меншою деталізацією, але такий набір позбавлений синтаксичної структури.

3.6.2 Семантичне ядро

Семантичне ядро – це набір ключових слів усього ресурсу, які найточніше характеризують тематику, товар або послугу, пропоновані сайтом.

Семантичне ядро складається з тих слів, які змогла виявити при скануванні документа пошукова система. Від того, наскільки грамотно ядро складене розроблювачами сайту, залежить релевантність сайту стосовно запитів.

Завдання

В лабораторній роботі потрібно провести дослідження та виконати порівняльний аналіз пошукових серверів, згідно класифікації, – **directory, search engines**. Для цього виконати такі пункти:

1. Скласти максимально повний список світових пошукових систем. Окремо виділити пошукові системи типу «каталог» (directory).
2. Виконати класифікацію пошукових систем: привести перелік тематичних та спеціалізованих пошукових систем, та визначити їх позиції в світовому та регіональному індексі.
3. Скласти максимально повний список регіональних (згідно варіанту) пошукових систем.
4. Визначити загальні та регіональні (згідно варіанту) рейтинги пошукових систем, котрі входять у першу п'ятірку рейтингу (TOP Ranking). Перевірити доступними засобами позиції пошукових систем та їх відповідність рейтингової оцінці.
5. Провести аналіз основних характеристик та наявних сервісів пошукових систем (згідно варіанту), та привести їх порівняльну характеристику.
6. Провести аналіз можливостей веб-аналітики (або обґрунтувати їх відсутність) для заданих пошукових систем (WebAnalytics, webmaster.yandex.ru, metrika.yandex.ru та ін.)

Застереження. У зв'язку з блокуванням доступу до російських ресурсів строком на три роки (згідно з указом Президента України №133/2017) використання у лабораторній роботі російських пошукових сервісів YANDEX, Mail.ru не є обов'язковим.

Хід роботи

1. Складання переліку пошукових систем

Скласти максимально повний список світових пошукових систем. Окремо виділити пошукові системи типу «каталог» (directory).

Для цього потрібно здійснити пошук, використовуючи наступні ключові фрази (використовувати англомовну та кириличну термінологію):

- **Web Search engines full index (web search engines full list)**
- **Пошукові системи інтернет повний список**
- **Поисковые системы интернет полный список**

В якості засобів пошуку пропонується вибрати наступні пошукові системи: Google.com, Yandex.ua, Nigma.ru (метапошуковий сервер, котрий здійснює видачу зі сторонніх пошукових серверів Google, Yandex, Yahoo, Bing, AltaVista, Rambler).

Для отримання підсумкового результату вибираємо перших 10 результатів пошукової видачі для кожного сервера і для кожного варіанту запиту (англійська, українська, російська мови).

Результати проілюструвати скріншотами кожного пошукового сервера та кожного пошукового запиту (9 скріншотів, рис.8).

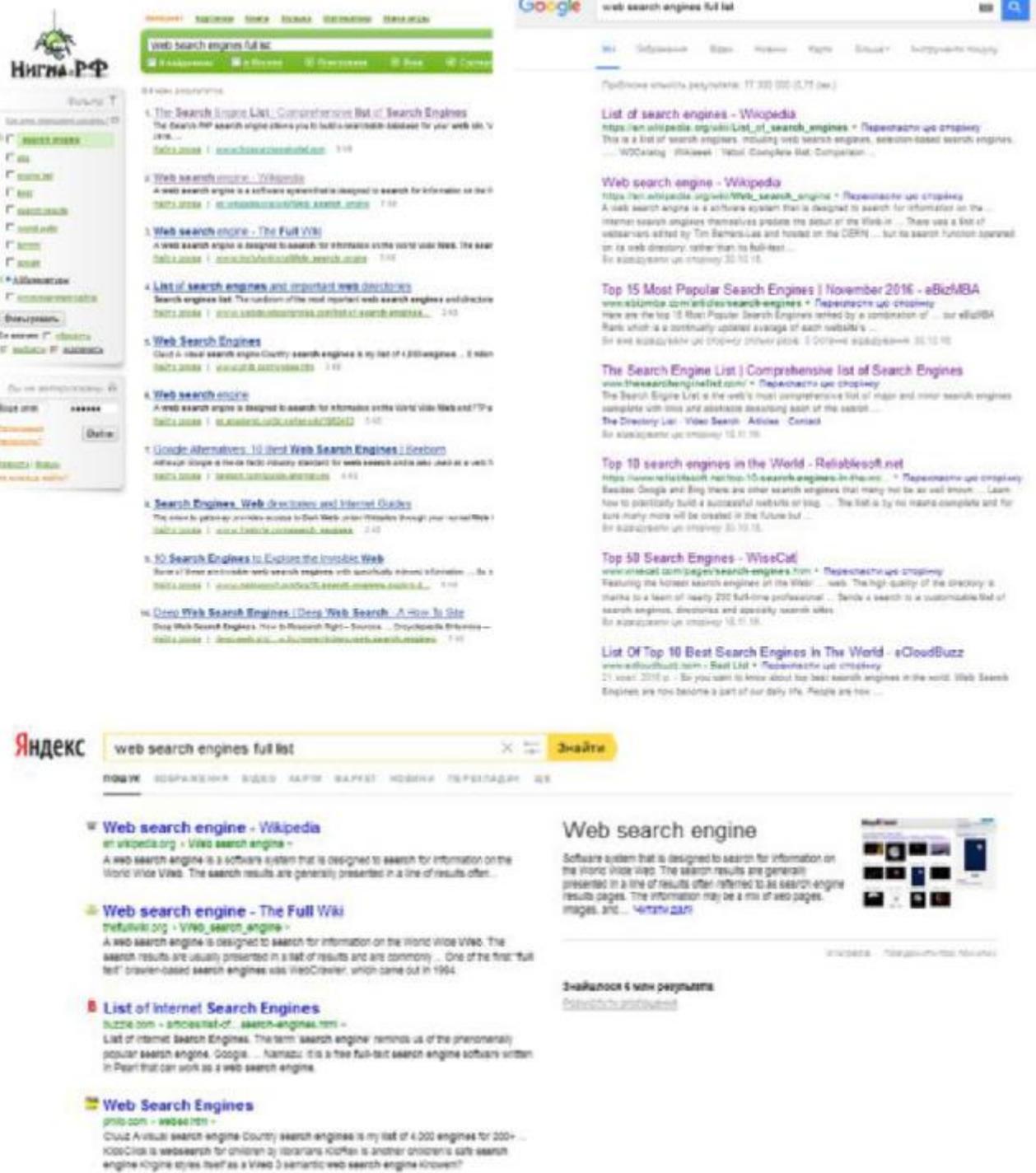


Рисунок 8 – Скріншоти пошуків Nigma.ru, Google.ua, Yandex.ua

Результати пошуків звести у таблиці з відмітками по релевантності кожної з 10-ти позицій видачі (табл.1). Релевантністьожної з позицій видачі оцінюємо наступним чином:

- «0» – результат пошуку у видачі повністю не релевантний (не відповідає пошуковому запиту);
 - «1» – результат пошуку у видачі частково релевантний (частково відповідає пошуковому запиту);
 - «2» – результат пошуку у видачі повністю релевантний (у повній мірі відповідає пошуковому запиту).

Таблиця 1 – Результати пошуку повного переліку пошукових серверів

Запит	№ видачі	Google.com		Yandex.ru		Nigma.ru	
		Позиція результуту видачі	Релевантність	Позиція результуту видачі	Релевантність	Позиція результуту видачі	Релевантність
web search engines full list	1	List of search engines - Wikipedia https://en.wikipedia.org/wiki/List_of_search_engines	2	Web search engine - Wikipedia en.wikipedia.org/w/index.php?title=Web_search_engine&oldid=9830000	2	The Search Engine List Comprehensive list of Search Engines http://www.thesearchenginelist.com/	2
	2	Результат видачі 2	1		2		2
	...						
	10	Результат видачі 10					
	Σ		сума		сума		сума
пошукові системи інтернет повний список	1	Огляд Пошукових систем - Інтернет компанія Новикова http://novikov.com.ua/service/node/6299	1	Пошук інформації в Інтернет. Пошукові системи. Повні... edufuture.biz/index.php?...Інтернет...системи...Повні...	0	Пошукові системи. Інтернет. Короткий опис і перелік найпопулярніших... newss.com.ua/specproekt/kluch_doi_informacij_posh...	1
	2		1		1		1
	...						
	10		1		1		0
	Σ		сума		сума		сума
	1						
	2						
	...						
	10						
	Σ						
Сума Релев.	Σ		XX		XX		XX

Провести аналіз пошукової видачі пошукових систем Google.com Yandex.ru Nigma.ru, за отриманими балами відзначити найбільш релевантну пошукову видачу, окремо за запитами різними мовами та сумарну.

Після проведеного пошуку необхідно обрати найвідповідніший результат з максимальною кількістю приведених у списку пошукових систем. Схвалюється, коли у списку є поділ систем за їх типами та приведена коротка анотація.

У звіті необхідно привести частковий скріншот обраного джерела з вказаним гіперпосиланням і джерелом пошукової видачі; скопіювати результати з джерела у звіт (рис.9).

The screenshot shows two search results pages. On the left, 'The Search Engine List' homepage displays a table of various search engines categorized by type. On the right, a search results page for 'Nigma.ru' is shown, listing several services related to business and trade.

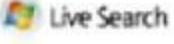
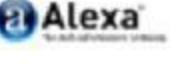
Type	Search Engine	Description
	Google	Google - The world's most popular search engine.
	bing	Bing Search: Microsoft's entry into the burgeoning search engine market. Better late than never.
	YAHOO! SEARCH	Yahoo! Search: The 2nd largest search engine on the web (as defined by a September 2007 Nielsen ratings report).
	altavista	Altavista: Launched in 1995, built by researchers at Digital Equipment Corporation's Western Research Laboratory. From 1996 powered Yahoo! Search, since 2003 - Yahoo technology powers Altavista.
	cuil	Cuil: Cuil was a search engine website (pronounced as Cool) developed by a team of ex-Googleers and others from Altavista and IBM. Cuil, named as the 'Google killer' was launched in July, 2009 and claimed to be world's largest search engine, indexing three times as many pages as Google and ten times that of MS. Now defunct.
	excite	Excite: Now an Internet portal, was once one of the most recognized brands on the Internet. One of the famous 90's dotcoms.
	Go.com	Go.com: The Walt Disney Group's search engine is now also an entire portal. Family-friendly!
All-Purpose Search	HotBot	HotBot was one of the early Internet search engines (since 1996) launched by Wired Magazine. Now, just a front end for Ask.com and AOL.

SEARCH RESULTS FOR "Nigma.ru"

Rank	Service	Description
1	Google Book Search	book searching portal with more than 30 search engines in its archive, the site searches hundreds of digital libraries and also occurs the net for network books.
2	Google	Google Book Search: The power of Google to find books. Google's entry will not let you see full text if the copyright is still active in your jurisdiction.
3	Alibaba.com	Alibaba.com - Claim to be the world's largest database of suppliers. Based in China, it is a marketplace of export and import, offers search, company directory, catalog, trade leads and more.
4	BANKERSAlmanac.com	BankersAlmanac.com provides intelligent reference data solutions to the banking industry for placements, due diligences, risk assessments and financial research.
5	IBI BUSINESS.COM	IBI BUSINESS.COM contains more than 400,000 listings within about 65,000 categories. Search results are provided by four types of search links.
6	Dun & Bradstreet	Dun & Bradstreet Company, maintains a database of over 23 million companies. Some information is provided free, other information is available to paid subscribers. Good for company stock information.
7	KOMPASS	Kompass: 2.3M companies in 70 countries categorized by 57,000 product & service keywords, 865,000 trade names and 4,000 association names. A paste for immediate sharing.
8	LexisNexis	LexisNexis: LexisNexis claims to be the world's largest collection of public records, unpublished opinions, forms, legal, news, and business information". Searchable archive of newspapers, public records & more.
9	ThomasNet.com	ThomasNet: Powered by the Thomas Register of American Manufacturers (The big Green Book) published since 1865. Catalogs over 630,000 american companies in 27,000 categories.
10	eMail-Search.org	eMail-Searching: a meta-portal with a number of tools for searching email addresses. Find current, former email addresses, extract them from the web.

Рисунок 9 – Скріншот обраного результату пошуку (Nigma.ru – перша позиція пошукової видачі), <http://www.thesearchenginelist.com/>

Таблиця 2 – Результат пошуку повного переліку пошукових серверів
<http://www.thesearchenginelist.com/>

Type	Search Engine	Description
All-Purpose Search Engines	 https://www.google.com	Google - The world's most popular search engine.
	 http://www.bing.com/	Bing Search: Microsoft's entry into the burgeoning search engine market. Better late than never.
	 https://www.yahoo.com/	Yahoo! Search: The 2nd largest search engine on the web (as defined by a September 2007 Nielsen Netratings report .
	 https://search.yahoo.com/?fr=altavista	AltaVista: Launched in 1995, built by researchers at Digital Equipment Corporation's Western Research Laboratory. From 1996 powered Yahoo! Search, since 2003 - Yahoo technology powers AltaVista.
	 Закритий 23 вересня 2010 http://www.cuil.com/	Cuil: Cuil was a search engine website (pronounced as Cool) developed by a team of ex-Googlers and others from Altavista and IBM. Cuil, termed as the 'Google Killer' was launched in July, 2008 and claimed to be world's largest search engine, indexing three times as many pages as Google and ten times that of MS. Now defunct.
	 http://www.excite.com/	Excite: Now an Internet portal, was once one of the most recognized brands on the Internet. One of the famous 90's dotcoms.
	 http://www.go.com/	Go.com: The Walt Disney Group's search engine is now also an entire portal. Family-friendly!
	 http://www.hotbot.com/	HotBot was one of the early Internet search engines (since 1996) launched by Wired Magazine. Now, just a front end for Ask.com and MSN.
	 alltheweb	AllTheWeb: Search tool owned by Yahoo and using its database, but presenting results differently.
	 galaxy.	Galaxy: More of a directory than a search engine. Launched in 1994, Galaxy was the first searchable Internet directory. Part of the Einet division at the MCC Research Consortium at the University of Texas, Austin
Accounting	 AOL Search	search.aol: Now powered by Google. It is now official.
	 Live Search	Live Search (formerly Windows Live Search and MSN Search) Microsoft's web search engine, designed to compete with Google and Yahoo!. Included as part of the Internet Explorer web browser.
	 LYCOS.	Lycos: Initial focus was broadband entertainment content, still a top 5 Internet portal and the 13th largest online property according to Media Metrix.
	 GIGABLAST	GigaBlast was developed by an ex-programmer from Infoseek. Gigablast supports nested boolean search logic using parenthesis and infix notation. A unique search engine, it indexes over 10 billion web pages.
	 Alexa	Alexa Internet: A subsidiary of Amazon known more for providing website traffic information. Search was provided by Google, then Live Search, now in-house applications run their own search.
Accounting	IFAC.com	IFAC.com: For resources and information on Ifrs and Accounting.

2. Класифікація пошукових систем

Після визначення загального переліку пошукових систем у лабораторній роботі потрібно привести перелік пошукових систем, згідно класифікації, – **directory, search engines, metasearch engines, semantic search engines, special-purpose search engines.**

Класифікація пошукових систем не стандартизована і може представлятись у достатньо довільній формі. Обов'язковими в класифікації є вищезгадані типи:

- Directory (Каталоги);
- All-purpose Search engines (Універсальні пошукові системи);
- Metasearch engines (Метапошукові системи);
- Semantic search engines (Семантичні пошукові системи);
- Special-purpose search engines (Пошукові системи спеціального призначення).

Слід зауважити, що універсальні пошукові системи можуть містити і функцію пошукових каталогів (наприклад Yandex – <https://yandex.ua/vaca/>), володіти функціями пошукових систем спеціального призначення, пошук зображень (наприклад Google – <https://images.google.com.ua/imghp>), пошук по географічних картах (наприклад Bing – <https://www.bing.com/mapspreview>), і т.д.

Критерії та типи класифікації, які стосуються пошукових систем спеціального призначення, є доволі розмитими та відносяться до вузькоспеціалізованих порталів.

Класифікацію пошукових серверів та систем спеціального призначення привести окремо з указанням критеріїв класифікації, назв та гіперпосилань ресурсів (табл.3).

Таблиця 3 – Класифікація пошукових серверів та систем спеціального призначення

Критерій класифікації	№№ з.п.	Назви та гіперпосилання ресурсів.
All-purpose Search engines (Універсальні пошукові системи)	1	Яндекс YANDEX https://yandex.ua/ Yandex is one of the largest internet companies in Europe, operating Russia's most popular search engine. We provide user-centric products and services based on the latest innovations in information retrieval, machine learning and machine intelligence to a worldwide customer audience on all digital platforms and devices. Headquartered in Moscow, we have development and sales offices in 17 locations across nine countries
	2	
	3	
	4	
	5	
Directory (Каталоги)	1	 DMOZ http://www.dmoz.org/ DMOZ is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a passionate, global community of volunteer editors. It was historically known as the Open Directory Project (ODP).
	2	
	3	
	4	
	5	
Metasearch engines (Метапошукові системи)	1	
	2	
	3	
	4	
	5	
Semantic search engines (Семантичні пошукові системи)	1	 Evi https://www.evi.com/ Evi is our best selling mobile app that can answer questions about books, music, films, conversions, history, people, places and much more.
	2	
	3	
	4	
	5	
Special-purpose search engines (Пошукові системи спеціального призначення)	зображення	
	відео	
	бізнес	
	торгівля	Логотипи систем обов'язкові
		Анотацію пошукових систем та сервісів вказувати оригінальну, з сайтів, розділ «About»
	І тд. усі	↓
	знайдені	↓
	типи пошуку	↓
		Мінімальна кількість позицій у класифікаторі – 5

Для визначення критеріїв класифікації пошукових сервісів і систем можна скористатись пошуком, ресурсами WIKI, ресурсами пошукових каталогів (DMOZ, наприклад, у відповідних рубриках – [Computers\ Internet\ Searching\ Search Engines](#)).

У випадку, коли загальний перелік пошукових серверів (табл. 4) містив поділ за класифікаційними критеріями, достатньо їх перерахувати з вказанням короткої анотації класифікатора.

Джерело, вказане в методичних вказівках, не використовувати.

Таблиця 4 - Критерії класифікації пошукових систем

№	Критерій класифікації	Коротка анотація пошукової системи чи сервісу
1.	All-Purpose Search Engines	<i>Універсальні пошукові системи</i>
2.	Accounting	
3.	Semantic	
4.	Bit Torrent	
5.	Blog	
6.	Books	
7.	Business	
8.	Email	
9.	Enterprise	
10.	Forum	
11.	Games	
12.	Human Search	
13.	International	
14.	Job	
15.	Legal	<i>Пошукові системи юристуденції та законодавства</i>
16.	Maps	
17.	Medical	
18.	MetaSearch	
19.	MultiMedia	
20.	News	
21.	Open Source	
22.	People	
23.	Question & Answer	
24.	Real Estate	
25.	School	
26.	Scientific	
27.	Shopping	
28.	Source Code	
29.	Usenet	
30.	Visual Search Engines	

3. Регіональні пошукові системи

Склади максимально повний список українських та інших регіональних пошукових систем (згідно з варіантом, табл.5).

Таблиця 5 – Розподіл регіонів за варіантами

№ варіанта	Регіон №1	Регіон №2
1.	Argentina	Armenia
2.	Australia	Austria
3.	Azerbaijan	Belarus
4.	Belgium	Brazil
5.	Bulgaria	Canada
6.	Chile	China
7.	Costa Rica	Croatia
8.	Czech Republic	Denmark
9.	Egypt	Estonia
10.	Finland	France
11.	Georgia	Germany
12.	Greece	Hong Kong
13.	Hungary	Iceland
14.	India	Ireland
15.	Israel	Italy
16.	Japan	Latvia
17.	Lithuania	Luxembourg
18.	Macedonia	Mexico
19.	Moldova	Netherlands
20.	New Zealand	Norway
21.	Poland	Portugal
22.	Romania	Russia
23.	Serbia	Singapore
24.	Slovakia	Slovenia
25.	South Africa	South Korea
26.	Spain	Sweden
27.	Switzerland	Taiwan
28.	Tajikistan	Thailand
29.	Turkey	Bosnia and Herzegovina
30.	United Kingdom	United States

Результати звести у таблиці з вказанням типу пошукової системи, наявності в ній каталогів та додаткових сервісів (табл.6).

Для визначення переліку регіональних пошукових сервісів і систем можна скористатись пошуком, ресурсами WIKI, ресурсами пошукових каталогів DMOZ та інших (наприклад, у відповідних рубриках – <http://www.dmoz.org/Regional>), та ресурсами сайтів WEB-аналітики (<http://www.alexa.com/topsites/countries>, <http://gs.statcounter.com/>).

Таблиця 6 –Регіональні пошукові системи та сервіси (Україна)

№ з.п.	Логотип, назва та гіперпосилання ресурсів,	Коротка анотація	Поз. Укр.	Поз. світ
1	 http://meta.ua/	META — український інтернет-портал. Використовує пошукову систему власної розробки з українською, російською та англійською мовами пошуку. Зона пошуку — українські сайти та сайти, що стосуються України.	67 Alexa	6,217 Alexa
2				
3		Логотипи систем обов'язкові		
4		Анотацію пошукових систем та сервісів вказувати оригінальну, з сайтів, розділ «About»		
5				

4. Визначення рейтингів пошукових систем

Визначити загальні та регіональні (згідно з варіантом) рейтинги пошукових систем, котрі входять у першу п'ятірку рейтингу (TOP Ranking). Перевірити доступними засобами позиції пошукових систем та їх відповідність рейтинговій оцінці.

Для визначення рейтингів пошукових сервісів і систем можна скористатись пошуком, ресурсами WIKI, ресурсами пошукових каталогів DMOZ та інших. Рейтинги пошукових систем потрібно скласти для Internet у загальному по цілому світу (world ranking), окремо для українського сегменту та для регіональних сегментів Інтернету (згідно з варіантом). В цілому у лабораторній роботі **потрібно скласти 4 рейтинги TOP 5 Ranking пошукових систем, – світу, України, країна 1 згідно варіанту, країна 2 - згідно з варіантом.**

4.1 Визначення рейтингів пошукових систем по ключових фразах пошуку

Для визначення загальносвітового рейтингу потрібно здійснити пошук, використовуючи наступні ключові фрази (використовувати англомовну та кириличну термінологію):

- **TOP 5 ranking Search engines in a World**
- **Рейтинг ТОП 5 пошукових систем Інтернет у світі**
- **Рейтинг ТОП 5 поисковых систем Интернет в мире**

В якості засобів пошуку пропонується вибрати наступні пошукові системи: Google.com, Yandex.ua, Nigma.ru (метапошуковий сервер, котрий здійснює видачу зі сторонніх пошукових серверів Google, Yandex, Yahoo, Bing, AltaVista, Rambler).

Результати видачі пошуку за кожною з пошукових фраз та в кожній з пошукових систем звести у табличну форму.

Кожен результат оцінити балами від 1 до 5 у кожній пошуковій видачі, відбираючи лише актуальні (за останні два роки) та ті, у яких вказано джерело оцінювання рейтингу. Неактуальні результати та ті, для яких не вказане джерело оцінювання, оцінити нулями.

Виставити усереднену оцінку по кожній серії пошукової видачі для кожної пошукової фрази і для кожного пошукового сервера:

СЕРЕДНЯ ОЦІНКА = \sum балів / кількість зарахованих результатів

**Таблиця 7 – Результати пошуку Google за пошуковою фразою
TOP 5 ranking Search engines in a World**

№	Результат пошукової видачі	Джерело рейтингу	Рейтинг TOP 5	
1	Top 10 Best Search Engines In The World - eCloudBuzz <i>www.ecloudbuzz.com > Best List</i> Перекласти цю сторінку <p>21 жовт. 2016 р. - If you want to know about <i>Best Search Engines In The World</i> then you should read this article. ... Some of queries includes improving alexa rank, speed up ... 5. AOL. AOL.com is also among the <i>top search engines in the world</i>.</p>	Alexa-rank Speed-up-Wordpress	Google	5
			Bing	4
			Yahoo	3
			Baidu	2
			AOL	1
2	Top 15 Most Popular Search Engines November 2016 - eBizMBA <i>www.ebizmba.com/articles/search-engines</i> Перекласти цю сторінку <p>Here are the <i>top 15 Most Popular Search Engines ranked</i> by a combination of ... which is a continually updated average of each website's Alexa <i>Global Traffic Rank</i>, ... 15 - eBizMBA <i>Rank</i> 400,000,000 - Estimated Unique Monthly Visitors 5 ...</p>	eBizMBA Rank Compete Rank Quantcast Rank Alexa Rank	Google	5
			Bing	4
			Yahoo	3
			Ask	2
			AOL	1
3	What are the top 10 most popular search engines? Search Engine ... <i>https://searchenginewatch.com/.../what-are-the-top-10-most-...</i> Перекласти цю сторінку <p>8 серп. 2016 р. - The second list is a <i>global overview</i> of most popular <i>search engines</i>, according to ... Bing has a great linkfromdomain:[site name] feature that highlights the <i>best ranked outgoing links</i> from that site, helping you ... 5) AOL Search.</p>	Net Market Share eBizMBA Rank	Google	5
			Bing	4
			Yahoo	3
			Ask	2
			AOL	1
4	Top 10 search engines in the World - Reliablesoft.net <i>https://www.reliablesoft.net/top-10-search-engines-in-the-wo...</i> Перекласти цю сторінку <p>Find out which are the <i>top 10 search engines</i> in the <i>World today!</i> ... chacha.com is the 8th most popular <i>search engine</i> with a <i>ranking position</i> of 297 in the US.</p>	comScore qSearch searchengineland.com netmarketshare Неактуально (2012)	Google	0
			Bing	0
			Yahoo	0
			Ask	0
			AOL	0
5	Top 10 Most Popular And Best Internet Search Engines in the World <i>www.abcnewspoint.com/top-10-most-popular-and-best-inter...</i> Перекласти цю сторінку <p>9 серп. 2015 р. - Yandex Search is one <i>best internet search engine</i> in the <i>world</i> which is owned by Russian based Internet ... share in Russia and considered as <i>top ranked search engine</i> in the country with respect to information. ... 5: Ask.com.</p>	Джерело не вказане	Google	0
			Bing	0
			Baidu	0
			Yahoo	0
			Ask	0

Невідповідні результати (позиції 4-5 у табл. 7) можна замінити **наступними релевантними у пошуковій видачі, вказуючи їх відповідний порядковий номер.**

Зводимо усереднені рейтингові результати в таблицю за зразком табл. 7.

Таблиця 8 – Усереднений рейтинг пошукових систем згідно пошуку Google за пошуковою фразою «TOP 5 ranking Search engines in a World»

№	Джерела рейтингу	Рейтинг TOP 5	
1	Alexa-rank	Google	15/3=5
	Speed-up-Wordpress	Bing	12/3=4
	eBizMBA Rank	Yahoo	9/3=3
	Compete Rank	Ask	4/3=1,33
	Quantcast Rank	AOL	3/3=1
	Alexa Rank	Baidu	2/3=0,66
	<u>Net Market Share</u>		
	eBizMBA Rank		
	Alexa-rank - 2		
	eBizMBA Rank -2		

Таблиці, аналогічні до табл.7, 8, скласти для варіантів пошуку у Google, Yandex та Nigma для всіх трьох варіантів ключових пошукових фраз (всього 9 таблиць).

Результати усередненого рейтингового оцінювання звести у підсумкову таблицю рейтингу пошукових систем (за зразком табл.8), у **стовпчику отриманих балів виставляється сумарний бал з 9-ти таблиць**.

Зверніть увагу:

- ✓ кількість оцінюваних пошукових систем може бути більшою від 5;
- ✓ необхідність пошуку в різних пошукових системах полягає в тому, що залежно від мови пошукового запиту та самого пошукового сервера, результати пошукової видачі можуть ранжуватись по різному і сильно відрізнятись у деяких позиціях;
- ✓ вказування джерела рейтингової оцінки дозволить обрати перелік систем WEB-аналітики та статистики, якими можна в подальшому скористатись для визначення позицій та рейтингів. Критерієм є максимальна кількість посилань на джерело при рейтинговому оцінюванні.

При визначенні регіональних рейтингів пошукових систем України та країн, призначених за варіантами, використовуючи метод пошуку за ключовою фразою через пошукові сервери, можна отримати багато результатів у пошуковій видачі, які будуть нерелевантні або неактуальні. В такому випадку потрібно або уточнювати ключові фрази, або вибирати з пошукової видачі лише релевантні позиції.

Слід зауважити, що в кириличному сегменті видачі пошукових результатів дуже часто прив'язка до джерела WEB-статистики і аналітики відсутня. В англомовному сегменті посилання на джерела, як правило, є присутні.

4.2 Визначення загальних та регіональних рейтингів пошукових систем за допомогою систем WEB-статистики і аналітики

Наступним етапом визначення рейтингових оцінок пошукових систем є їх перевірка в системах WEB-статистики та аналітики.

Для цього можна скористатись результатами попереднього пошуку та використати джерела WEB-статистики та аналітики, котрі найчастіше приводились при визначенні загального рейтингу.

Для виконання даної лабораторної роботи пропонуються декілька авторитетних джерел. При виборі переліку засобів WEB-статистики і аналітики слід враховувати методи збору інформації та регіональні особливості сегментів Інтернету, де виконується збір статистики.

Наприклад, системи StatCounter та Alexa більше орієнтовані на західний сегмент Інтернету та США. Тому результати за такими пошуковиками, як Yandex, Meta, BigMirNet, Rambler, орієнтованих на кириличну аудиторію, будуть випадати із загальної статистики. Навпаки, місцеві ресурси WEB-статистики окремих регіонів не завжди об'єктивно оцінюють величину трафіку західних пошукових систем, особливо тих, які менші за розміром від Google, Bing, Yahoo, AOL, хоча їх частка в загальному рейтингу є відчутною.

4.2.1. StatCounter

<http://gs.statcounter.com/> (http://gs.statcounter.com/#all-search_engine-ww-yearly-2008-2016)



Рисунок 10 – Сервіс StatCounter, оцінювання статистики пошукових систем

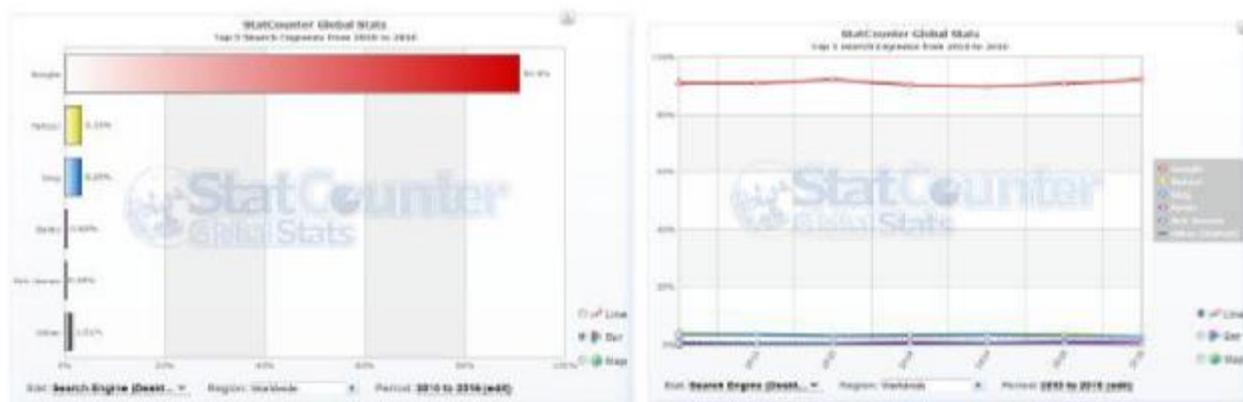


Рисунок 11 – Сервіс StatCounter, при виборі оцінювання статистики пошукових систем за період 2010-2016 р.р.



Рисунок 12 – Сервіс StatCounter, при виборі оцінювання статистики пошукових систем за період 3 квартал 2016 р., регіон – WorldWide

При використанні сервісу StatCounter для визначення загальних та регіональних рейтингів потрібно встановити такі параметри відображення рейтингової статистики: **Stat**, **Region**, **Period**.

Параметр **Stat** встановити у позицію **Search Engines**, параметр **Region** почергово встановити у позицію **WorldWide**, **Ukraine**, **Country 1**, **Country 2** (згідно варіанту). Параметр **Period** визначає період, за який буде відображена статистика рейтингової оцінки.

Результат оцінювання рейтингу пошукових систем за регіонами згідно варіанту (**Country 1**, **Country 2**) проілюструвати скріншотами по аналогії з рисунками 11, 12.

Отримати результат у форматі [csv spreadsheet](#), використовуючи команду Download Data:

Search Engine, "Market Share Perc. (Oct to Nov 2016)"	
Google	92.87
bing	2.68
Yahoo!	2.28
Baidu	0.84
YANDEX RU	0.32
YANDEX	0.29
Haosou	0.11
DuckDuckGo	0.11
AOL	0.07
Shenma	0.07
Naver	0.06
Mail.ru	0.05
Soguo Search	0.05
Ask Jeeves	0.04
MSN	0.04
Seznam	0.03
Daum	0.02
t-online	0.01
Other	0.06

Результати рейтингової статистичної оцінки занести у таблицю за зразком 9.

Таблиця 9 – Результати рейтингової статистичної оцінки пошукових систем сервісом StatCounter

WorldWide	Ukraine	Country 1	Country 2
Google	92.87		
bing	2.68		
Yahoo!	2.28		
Baidu	0.84		
YANDEX RU	0.32		
YANDEX	0.29		
Haosou	0.11		
DuckDuckGo	0.11		
AOL	0.07		
Shenma	0.07		
Naver	0.06		
Mail.ru	0.05		
Soguo Search	0.05		
Ask Jeeves	0.04		
MSN	0.04		
Seznam	0.03		
Daum	0.02		
t-online	0.01		
Other	0.06		

Порівняти отриманий результат з підсумковою табл. 8 ТОП-5 (п.4.1), зробити висновки (при наявності відхилень у рейтинговому оцінюванні пошукових систем).

4.2.2. Рейтингові оцінювання за іншими джерелами веб-аналітики

Аналогічні рейтингові оцінювання пошукових машин можна здійснити за такими запропонованими джерелами веб-аналітики:

- 1) <http://www.freewebsubmission.com/>

The screenshot shows a table of search engine and directory rankings. The columns include the rank, name of the site, its position in the search results, the number of pages indexed, and the number of backlinks. The table lists 50 entries, starting with Google at rank 1.

1.	Google	1	100	26.	SWD	46,924 56
2.	Bing	16	97	27.	24/7 Web Directory	47,207 39
3.	Yandex	1,495	78	28.	Info Tiger	48,364 53
4.	Open Directory	7,303	94	29.	AceWebDirectory	49,219 36
5.	SimilarSites	17,560	57	30.	W3 Catalog	49,940 35
6.	ScrubTheWeb	18,221	75	31.	Online Society	50,151 38
7.	EntireWeb	18,944	64	32.	Add Business	51,012 46
8.	ASR	25,026	73	33.	Pegasus Directory	55,160 56
9.	SimilarSiteSearch	27,316	47	34.	Directory Free	56,820 42
10.	Amfibi	30,819	1	35.	1Abc Directory	58,300 57
11.	ExactSeek	31,812	70	36.	IS	58,826 45
12.	SoMuch	33,610	59	37.	OBLN	59,104 32
13.	GigaBlast	34,670	70	38.	oneMission	62,731 44
14.	Viesearch	37,811	56	39.	ESD	63,916 45
15.	A1WebDirectory	38,696	49	40.	LinkRoo	64,439 47
16.	Anpoox	40,591	55	41.	Cipinet	64,545 52
17.	SecretSELabs	41,096	43	42.	BusinessSeek	65,358 50
18.	SPD	43,573	54	43.	MasterMOZ	67,441 43
19.	Ananar	44,407	35	44.	Botid	67,473 53
20.	Sites Web Directory	44,981	48	45.	SimilarPages	70,831 35
21.	ABC Directory	44,983	48	46.	HotvsNot	71,151 67
22.	1WebsDirectory	46,091	41	47.	Link Centre	71,168 54
23.	Gain Web	46,293	42	48.	OneMillionDirectory	73,623 36
24.	SonicRun	46,641	62	49.	Directory SEO	74,770 41
25.	9Sites	46,709	46	50.	LinkPedia	75,113 49

The screenshot shows the homepage of FreeWebSubmission.com. At the top, there's a red circular logo with 'FWS' and the text 'FreeWebSubmission.com'. To the right is a graphic of a rocket launching. Below the logo, there are links for 'Home', 'About Us', 'Privacy', 'Link to FWS', and 'Contact Us'. There are also social media icons for Facebook and Google+. A red button on the right says 'CLICK HERE' and 'SUBMIT YOUR SITE'. Below this, a dark box contains the text 'THE SUBMISSION MONSTER COMPLETE PACKAGE' and 'Receive Monthly Submission, a Featured Listing, 10 SEO tools, Guaranteed Traffic, and Rapid Paid Inclusion across our network of search engines.' A list of search engines follows:

Add Page	MillionShort
All Business Directory	NetSmart
AllFreeThings	NewWebDirectory
Amidalia	P-Series
aQualityNet	Polygal
Bamboo	Ripley
Bonito	SeoDI
Bubbleplex	Skaffle
Daduru	Skyobe
Dotuu	SpiffySearch
DirectoryHT	Subject
Finest 4	TagoHub
Gimpix	100Directory
InnerSeek	Thales Directory
Italist	URLGatbe
Linkaron	WiseWeb
LinkDir	Wondex
LinkDirka	Wotbox
Link1 Directory	01WebDirectory

2) <https://moz.com/top500>

The Moz Top 500

Moz's list of the top 500 domains and pages on the web. Last updated October 26, 2016.

The table below contains a list of the top 500 registered domains (+ example.com) ranked by the number of linking root domains. This data is sourced from the Mozscape web index of 12 Billion domains and 147 Billion pages.

[Download the Top 500 Domains as a CSV](#)

Top 500 Domains		Top 500 Pages				
Rank	Root Domain	Linking Root Domains	External Links	Domain mozRank	Domain mozTrust	Change
1	Facebook.com	13,960,075	2,147,483,647	9.67	9.42	0
2	Twitter.com	8,882,677	2,147,483,647	9.67	9.47	0
3	Google.com	9,635,974	2,147,483,647	9.46	9.34	0
4	YouTube.com	5,358,727	2,027,026,599	9.24	9.14	0
5	Wordpress.org	9,549,060	329,714,578	9.07	8.80	0
6	LinkedIn.com	3,372,272	1,006,723,037	9.03	8.92	0
7	Instagram.com	2,404,961	1,208,565,312	9.01	8.82	+1
8	Wikipedia.org	2,205,696	273,877,952	8.74	8.64	-1
9	Pinterest.com	2,197,101	658,179,448	8.83	8.67	0
10	Adobe.com	1,809,414	240,650,559	9.00	9.10	+1
11	Blogspot.com	1,849,546	1,042,376,654	8.53	8.51	-1
12	Hugedomains.com	1,633,463	30,241,135	8.31	7.80	+3
13	Wordpress.com	1,620,116	301,381,838	8.61	8.57	-1

3) <https://botw.org/top/Computers/Internet/Statistics/>

The screenshot shows the Best Web homepage with a navigation bar at the top. The main content area is titled "Internet Statistics". It features a "Free 30 Day Advertising Trial" banner and a "TRUSTED" badge. Below the title, there's a list of links to various resources:

- [Alexa](#): Provides traffic details on top sites and offers a list of the top 500 most trafficked sites.
- [Cooperative Association for Internet Data Analysis](#): Explores Internet infrastructure through traffic performance and flow characterization, simulation, analysis, and visualization, with an extensive set of network tools and utilities.
- [Cyber-Geography](#): Collection of cyberspace visualizations, with a cyberatlas of census, geographic, topological, and other types of Internet maps.
- [Dataopedia](#): Statistics and other information about the Internet. Presents the most popular and most shared websites as well as the most recently uploaded ones.
- [Free Counter](#): Provides a reliable hit counter with real time web site statistics and customizable images.
- [Internet Measurement Infrastructure - CAIDA](#): Summaries of current measurement projects offering public reports.
- [Internet Traffic Report](#): Monitors the worldwide flow of Internet data by continent, provider/router and location, tracking response time and bandwidth, then assigning an overindex rating (Index = 100).
- [Internet World Stats](#): Site dedicated to providing information on world Internet usage, population statistics, and internet market research data.
- [Keynote](#): Generates information about network availability, latency, and packet loss between major Internet backbones in the United States.
- [National Broadband Map \(NBM\)](#): United States government initiative aimed at providing information on broadband coverage in a particular area. Also presents broadband coverage maps.
- [Nielsen NetRatings](#): Provides a variety of business solutions and research on web traffic and site optimization.
- [One Hour Per Second](#): Interesting statistic on the number of videos created and uploaded to YouTube.

4) інші джерела:

[Top / Computers / Internet / Statistics](#)

[Alexa](#)

Provides traffic details on top sites and offers a list of the top 500 most trafficked sites.

[Cooperative Association for Internet Data Analysis](#)

Explores Internet infrastructure through traffic performance and flow characterization, simulation, analysis, and visualization, with an extensive set of network tools and utilities.

[Cyber-Geography](#)

Collection of cyberspace visualizations, with a cyberatlas of census, geographic, topological, and other types of Internet maps.

[Dataopedia](#)

Statistics and other information about the Internet. Presents the most popular and most shared websites as well as the most recently uploaded ones.

[Free Counter](#)

Provides a reliable hit counter with real time web site statistics and customizable images.

[Internet Measurement Infrastructure – CAIDA](#)

Summaries of current measurement projects offering public reports.

Internet Traffic Report

Monitors the worldwide flow of Internet data by continent, provider/router and location, tracking response time and packet loss, then assigning an overall index rating (best = 100).

Internet World Stats

Site dedicated to providing information on world Internet usage, population statistics, and Internet market research data.

Keynote

Generates information about network availability, latency, and packet loss between major Internet backbones in the United States.

National Broadband Map (NBM)

United States government initiative aimed at providing information on broadband coverage in a particular area. Also presents broadband coverage maps.

Nielsen NetRatings

Provides a variety of business solutions and research on web traffic and site optimization.

One Hour Per Second

Interesting statistic on the number of videos created and uploaded to Youtube.

StatCounter Global Stats

Provides statistics on web browser usage, browser versions, search engines, operating systems, screen resolutions, and social media.

statMetrix

Supplies monthly reports on global Internet usage trends useful for web developers and online businesses.

ZookNIC Internet Intelligence

Provides domain name data and analysis, and academic research.

<http://www.bigmir.net>

<http://www.liveinternet.ru>

<http://www.netapplications.com/>

<http://marketshare.hitslink.com/>

<https://www.similarweb.com/top-websites>

<http://www.alexa.com/toolbar>

<http://www.alexa.com/siteinfo>

5. Провести аналіз основних характеристик та наявних сервісів пошукових систем (згідно з варіантом) та привести їх порівняльну характеристику.

Пошукова машина (search engine), на теперішній час, – це не тільки засіб пошуку, а ще набір додаткових сервісів, які вона надає користувачу. Тому для загальної оцінки якості пошукових машин потрібно долучити оцінювання пропонованих користувачеві додаткових сервісів. Потрібно оцінити їх кількість та призначення.

У цю оцінку потрібно залучити наявність та зручність додаткових служб пошукових машин, – поштовий сервіс, наявність мережевих дисків та інших хмарних технологій, наявність засобів комунікації тощо.

Як приклад можна представити пошукову машину google.com/

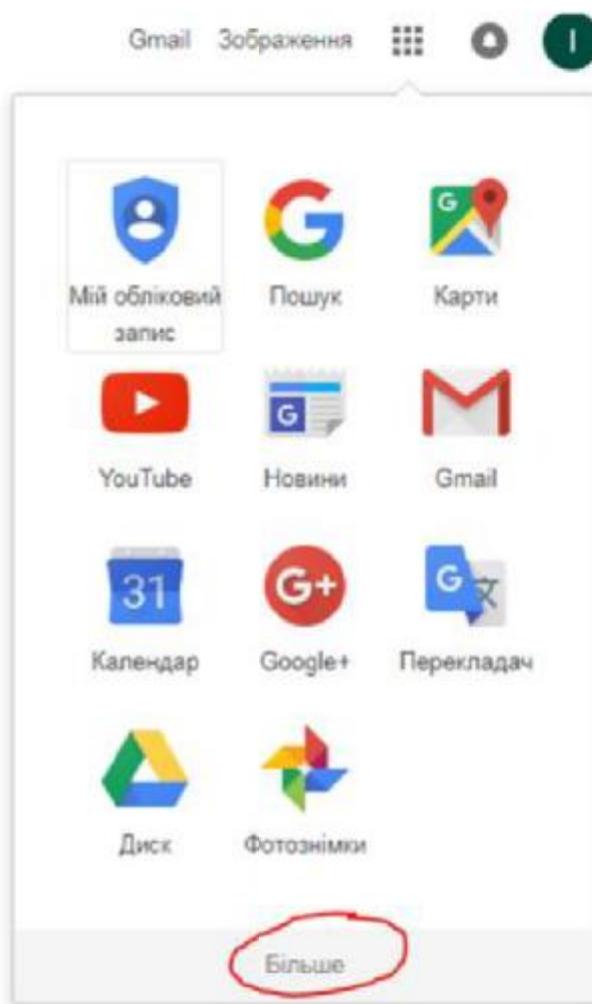


Рисунок 13 - Вкладка служб google.com на стартовій сторінці власного web-браузера Google CHROME

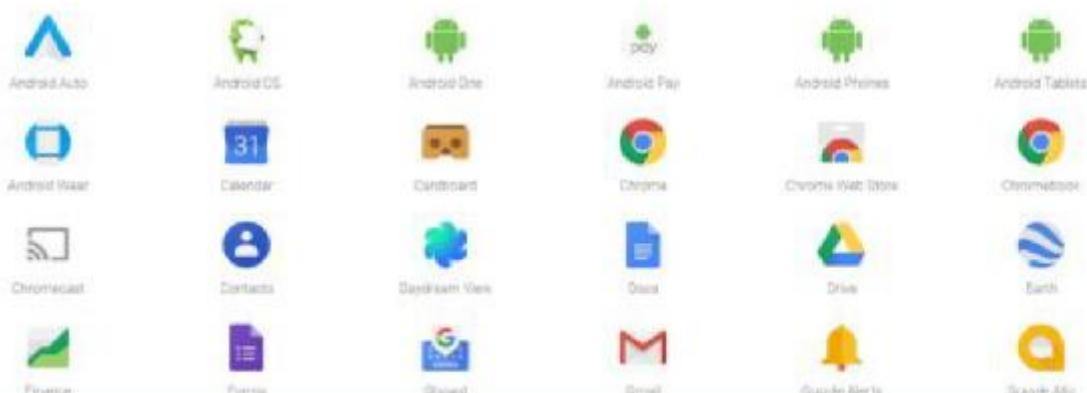
[See all products](#)

- ▼ For all
- ▼ For business
- ▼ For developers

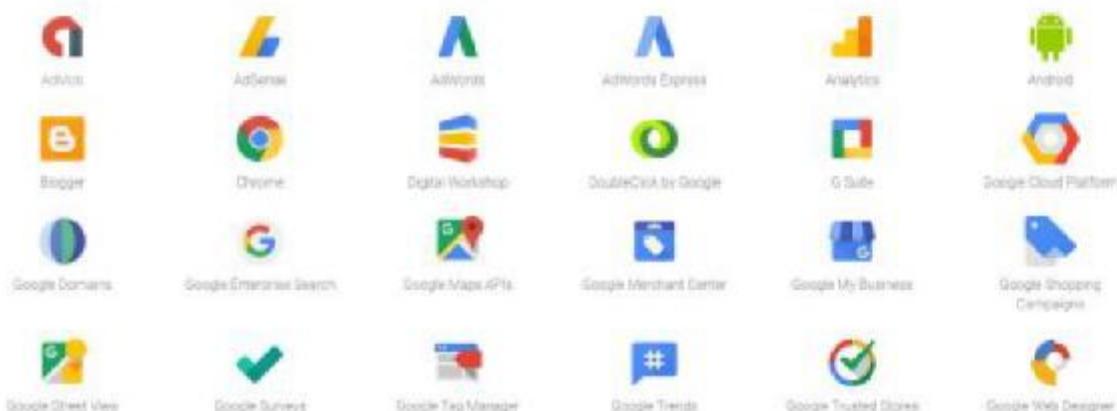
[BACK TO TOP](#)

[See all products](#)

- ▲ For all



- ▲ For business



- ▲ For developers

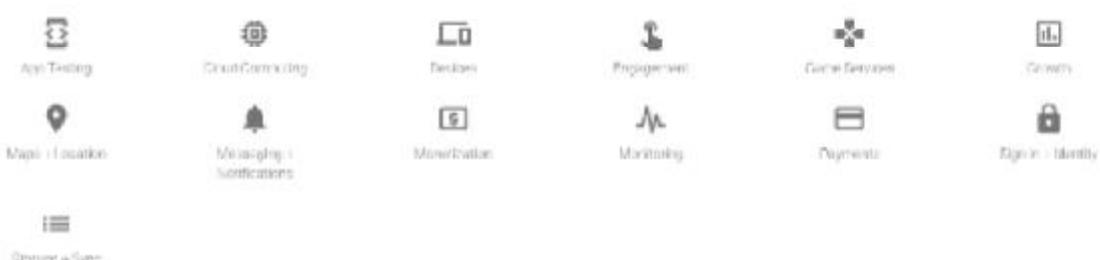


Рисунок 14 - Скріншот сторінки повного переліку служб [google.com, – https://www.google.com.ua/intl/uk/about/products/](https://www.google.com.ua/intl/uk/about/products/)

Результати подати у табличній формі за зразком табл.10.

Таблиця 10– Перелік наявних сервісів пошукових систем

№ з/п	Назва сервісу, служби	Web–адреса	Опис служби
1	 Gmail Електронна пошта	https://mail.google.com	Поштовий сервіс, електронна пошта

6. Провести аналіз можливостей веб-аналітики (або означити їх відсутність) для заданих пошукових систем

Інструменти веб-аналітики пошукових машин чи їх відсутність є важливим фактором у визначенні та оптимізації, у подальшому, позицій сайтів користувача.

Веб-аналітика (Web analytics) — це вимір, збір, аналіз, подання та інтерпретація інформації про відвідувачів веб-сайтів з метою її поліпшення та оптимізації. Основним завданням веб-аналітики є моніторинг відвідуваності веб-сайтів, на підставі даних якого визначається веб-аудиторія і вивчається поведінка веб-відвідувачів для прийняття рішень щодо розвитку і розширення функціональних можливостей веб-ресурсу.

Інструменти веб-аналітики

Збирати статистику можна з допомогою:

- **лічильників** — зовнішніх програм. Для отримання статистики на веб-сторінки сайту встановлюється невеликий фрагмент коду (зазвичай 1-2 кілобайт). Сенс в тому, що при вході на сайт браузер вантажить картинку, яка розміщена на сайті збору інформації. Дані про завантаження лічильника заносяться в базу даних, яка може розміщуватися на сервері постачальника послуги збору та обробки статистики, і потім проглядаються, наприклад, на його сайті;
- **лог-аналізаторів.**

Порівняльна таблиця обох інструментів:

Інструмент	Переваги	Недоліки
Лічильник	Простий і зручний у використанні. Дозволяє отримувати оперативну наочну інформацію.	Вимагає установки на сайт програмного коду. Лічильники можуть не зібрати або втратити дані, якщо веб-сторінка недовантажилась або виникли технічні проблеми при передачі даних. Неможливість збору статистики за контентом, що завантажується, трафіком сайту, закладками відвідувачів у браузері.
Лог-аналізатор	Дозволяє аналізувати помилки роботи сервера, відстежувати хакерські атаки, складати спеціальні звіти. Дає точніші дані про кількість відвідувачів сайту. Дозволяє вирішувати більш вузькі і складні завдання, створювати власні специфічні звіти. Дозволяє збирати статистику за контентом, який завантажується, трафіком, закладами відвідувачів в браузери.	Вимагає високої кваліфікації адміністратора веб-ресурсу для установки і налаштування аналізатора.

Далі наведено перелік деяких видів інструментів веб-аналітики.

– **Лог-аналізатори:**

[WebTrends^{\[en\]}](#)

[WebaLizer](#)

AWStats

– **Лічильники-рейтинги** (дають кількість відвідувачів за день, тиждень, місяць, за всю історію):

Rambler's Top100

[Liveinternet](#)

Рейтинг@Mail.ru

[OpenStat](#)

[HotLog](#)

– **Лічильники-трекери** (тепер – системи інтернет-статистики, які дають сумарну інформацію щодо відвідувань, обрану для певного зрізу, заданого користувачем):

[Google Analytics](#)

Piwik

[Яндекс.Метрика](#)

[Liveinternet](#)

Рейтинг@Mail.ru

[OpenStat](#) (кол. Spylog)

[HotLog](#)

- **Інтернет-статистика з деталізацією за переглядами** (крім сумарної інформації дають інформацію за переглядами сторінок всередині кожного відвідування):

[Woopra^{\[en\]}](#)

- **Інтернет-аналітика з деталізацією поведінки відвідувача** (дає максимально детальну інформацію з можливістю перегляду всіх дій відвідувачів: рухи миші, кліки, клавіші тощо. За зібраною поведінковою інформацією будуються звіти у вигляді карт активності відвідувачів на сторінці):

[ClickTale](#)

[Яндекс.Метрика](#)

[SpyBOX](#)

Сертифікації в галузі веб-аналітики

На даний момент у світі існує дві пріоритетних сертифікації фахівців в області веб-аналізу:

1. Google Analytics Certified Partner — сертифікація для компаній, які показали свою здатність в аналізі сайтів з допомогою Google Analytics. Сертифікація проводиться після співбесіди в офісах Google фахівцями компанії з веб-аналізу. Для сертифікації необхідно мати три успішні проекти в області веб-аналізу, спрямованих на підвищення конверсії та оптимізацію витрат на інтернет-рекламу.
2. Web Analytics Association Member — членство в міжнародній асоціації веб-аналітики, що дає право на формування стандартів у галузі відстеження веб-даних для інтернет-реклами.

Прикладом визначення інструментів вебаналітики на пошуковій машині може слугувати наявність лічильників та рейтингів пошукової машини (рис. 15).

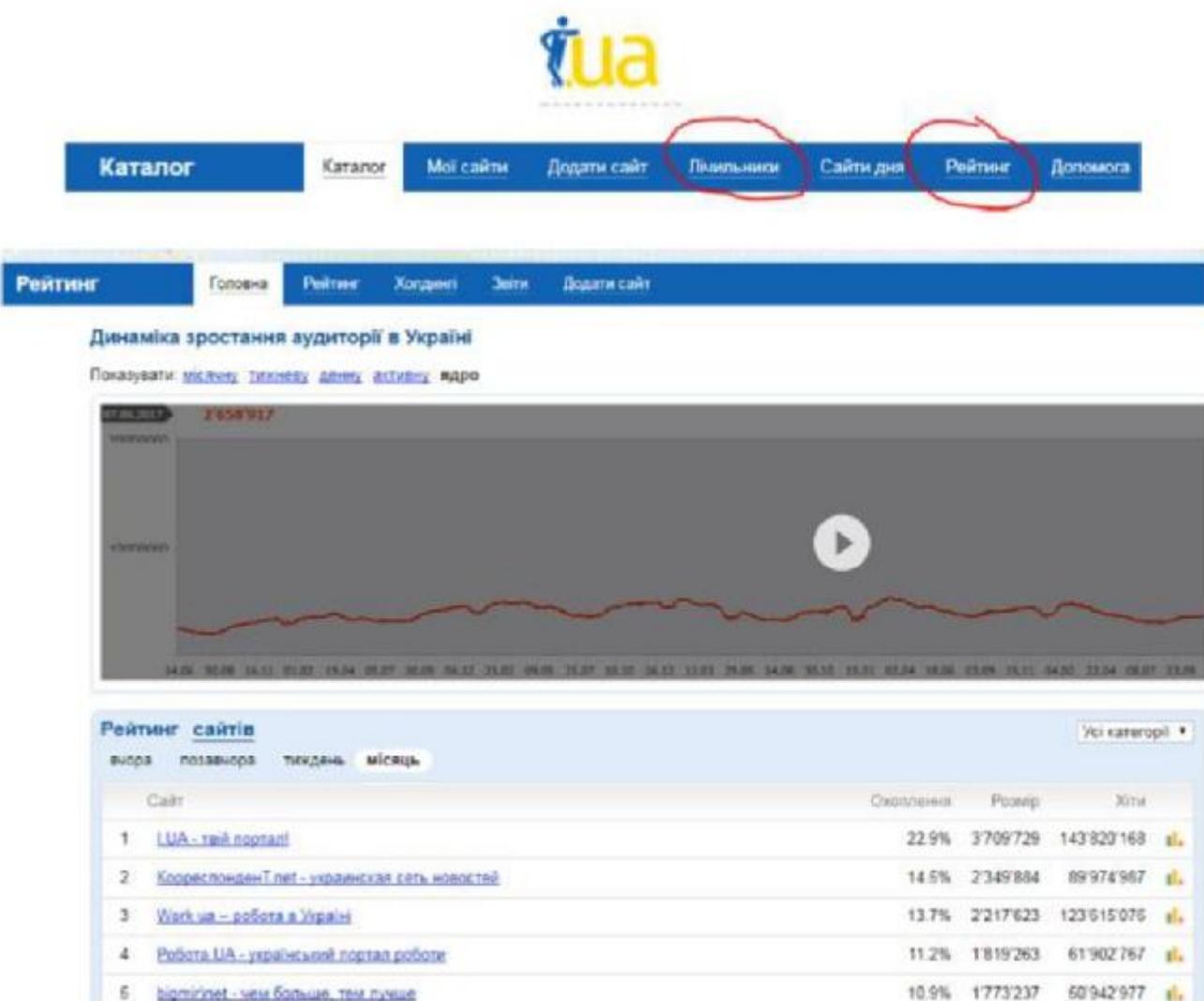


Рисунок 15 - Скріншот сторінки пошуковика <http://www.i.ua/>, <http://top.i.ua/>

Література

1. Ресурси WIKI.
2. Айверсон У. Популярные Web-сервисы. Практика использования Real World Web Services / Уилл Айверсон. – КУДИЦ-Образ, 2005. 240 стр., ISBN 5-9579-0077-X, 0-596-00642-X

ДОДАТОК 1. ОСНОВНІ ПОШУКОВІ СИСТЕМИ

Хронологія			Хронологія		
Рік	Система	Подія	2006	2007	2008
1993	W3Catalog^[en]	Запуск		Quaero	Заснування
	Aliweb	Запуск		Live Search	Запуск
	JumpStation^[en]	Запуск		ChaCha^[en]	Запуск (бета)
1994	WebCrawler^[en]	Запуск		Guruji.com^[en]	Запуск (бета)
	Infoseek^[en]	Запуск	2007	wikiseek	Запуск
	Lycos	Запуск		Sproose^[en]	Запуск
1995	AltaVista	Запуск		Wikia Search	Запуск
	Daum	Заснування		Blackle.com^[en]	Запуск
	Open Text^[en] Web Index	Запуск	2008	DuckDuckGo	Запуск
	Magellan	Запуск		Tooby	Запуск
	Excite	Запуск		Picollator^[en]	Запуск
	SAPO	Запуск		Viewzi^[en]	Запуск
	Yahoo!	Запуск		Cuil	Запуск
1996	Dogpile^[en]	Запуск		Boogami^[en]	Запуск
	Inktomi^[en]	Заснування		LeapFish^[en]	Запуск (бета)
	Рамблер	Заснування		Forestle^[en]	Запуск
	HotBot^[en]	Заснування		VADLO	Запуск
	Ask Jeeves	Заснування		Powerset	Запуск
1997	Northern Light^[en]	Запуск	2009	Bing	Запуск
	Яндекс	Запуск		KAZ.KZ	Запуск
1998	Google	Запуск		Yebol^[en]	Запуск (бета)
1999	AlltheWeb^[en]	Запуск		Mugurdy^[en]	Закриття
	GenieKnows^[en]	Заснування		Scout^[en]	Запуск
	Naver	Запуск	2010	Cuil	Закриття
	Teoma	Заснування		Blekko	Запуск (бета)
	Vivisimo^[en]	Заснування		Viewzi	Закриття
2000	Baidu	Заснування	2012	WAZZUB	Запуск
	Exalead^[en]	Заснування	2014	Спутник	Запуск (бета)
2003	Info.com^[en]	Запуск			
2004	Yahoo! Search	Фінальний запуск			
	A9.com^[en]	Запуск			
	Sogou^[en]	Запуск			
2005	MSN Search	Фінальний запуск			
	Ask.com	Запуск			
	Нигма	Запуск			
	GoodSearch^[en]	Запуск			
	SearchMe^[en]	Заснування			

ДОДАТОК 2. МОВА ПОШУКОВИХ ЗАПИТІВ

2. Оператори пошукових запитів

Оператори пошукових запитів допомагають одержати корисну інформацію про індексацію сайту, виявити проблеми й навіть розібратися в нюансах роботи пошукових алгоритмів.

2.1. Загальні для Яндекс і Google оператори

Зазначимо, що якщо Яндекс точно відповідає операторам, то **Google може їх проігнорувати**, якщо вважатиме, що є кращі результати.

- **Оператор «+» і «-»** — пошук документів, які обов'язково містять (або обов'язково не містять) зазначене слово. Можна використати декілька операторів в одному запиті, причому як «мінус», так й «плюс».
- **Пошук за цитатою, оператор лапок «»** — пошукова система буде шукати точний збіг фрази. Можна використати кілька разів в одному запиті. Можна додати «мінус» перед одним із запитів.
- **Оператор «*»** — в Яндексі використовується для вказівки пропущеного слова в цитаті. Застосовується тільки з оператором «». В Гуглі використовується для вказівки пропущених слів у запиті.
- **Оператор «|»** — Пошуки сторінок, які містять кожне зі слів, пов'язаних цим оператором.
- **Оператор «~»** — в Яндексі шукає документи, у яких слово, зазначене після оператора, не міститься в одній пропозиції зі словом до оператора. Гугл: шукає документи із зазначенним словом і його синонімами.

2.2. Оператори пошуку для Яндекса

- **Оператор !** — пошук документів, де слово втримується тільки в заданій формі. Можна шукати навіть слова із заголовними буквами.
- **Оператор !!** — пошук документів, де слово втримується в будь-якій формі, у будь-якому відмінку.
- **Оператор & (амперсант)** — пошук документів, де слова, зв'язані оператором, перебувають в одній пропозиції.
- **Оператор && (подвійний амперсант)** — те ж саме, тільки слова в межах одного документа.

- **Оператор <<** — пошук слів у межах документа, але релевантність (вона впливає на положення в результатах пошуку) розраховується тільки по першому слову (яке міститься до оператора).
- **Оператор /n**, де n — максимальна відстань між заданими словами — пошук документів, у яких задані слова розташовуються в межах n слів одне щодо одного. Додатково можна задати прямий (+) або зворотний (-) порядок проходження слів у знайдених документах.
- **Оператор /(m n)**, де m — мінімальна відстань між заданими словами, n — максимальна відстань між заданими словами: пошук документів, у яких задані слова розташовуються на відстані не менше m і не більше n слів одне від одного. Додатково можна задати прямий (+) або зворотний (-) порядок проходження слів у знайдених документах.
- **Оператор && /n**, де n — максимальна відстань між пропозиціями, що містять слова запиту: пошук документів, у яких слова запиту (розділені оператором) розташовуються в межах n пропозицій одне від одного. Порядок, у якому йдуть слова запиту, не враховується.
- **Оператор ()** — угрупування слів при складних запитах. Всередині укладеної в дужки групи також можуть бути використані будь-які оператори.

2.2.1 Документні оператори Яндекса

- **title:** — пошук по заголовках сторінок;
- **url:** — пошук по сторінках на заданому URL, наприклад url:aiwastudio.ru/blog/*;
- **site:** — пошук по всіх піддоменах і сторінках зазначеного сайту;
- **inurl:** — пошук по сторінках, розміщених на даному хості. Ідентичний операторові url: із заданим ім'ям хоста;
- **domain:** — пошук по сторінках, розташованих на даному домені;
- **mime:** — пошук по документах за заданим типом файлу;
- **lang:** — пошук по сторінках заданою мовою;
- **date:** — пошук по сторінках з обмеженням по даті їхньої останньої зміни; рік зміни вказується обов'язково. Місяць і день можна замінити символом *;
- **cat:** — Пошук по сторінках сайтів, зареєстрованих в Яндекс.Кatalозі, тематична рубрика або регіон яких збігаються із заданим.

Перераховані вище оператори не обов'язково запам'ятовувати, тому що розширеній пошук Яндекса (використання фільтрів) по суті є інтерфейсом до

цих операторів. Тобто ви можете вибирати настроювання пошукової форми й одержувати результат, ніби вводили оператори вручну.

2.2.2 Недокументні оператори:

- **Intext** – шукає тільки ті документи, текст яких містить слова запиту, тобто не в метатегах або ще десять, а саме в тексті;
- **image** – шукає всі документи, у яких міститься зображення із заданим ім'ям;
- **Anchormus** – шукає посилання на музичні файли, що містять зазначений запит в анкорі;
- **Linkmus** – шукає всі сторінки, з яких є посилання на зазначений музичний файл;
- **Inlink** – для пошуку в тексті посилань;
- **Linkint** – пошук внутрішніх посилань на певний документ;
- **Anchorint** – пошук документів, що містять зазначений запит у текстах своїх посилань на свої внутрішні документи;
- **idate** – шукає документи із заданою датою останньої індексації;
- **style** – пошук за значенням атрибута stylesheet тегу link;
- **applet** – пошук за значенням атрибута code тегу applet;
- **script** – пошук за значенням атрибута src тегу script;
- **object** – пошук по вмісту атрибутів тегу object;
- **action** – пошук за значенням атрибута action тегу form;
- **profile** – пошук за значенням атрибута profile тегу head;
- **inpos** – пошук тексту в межах заданих позицій елементів на сторінці (inpos:0..100).

2.3 Оператори пошуку для GOOGLE

- **Оператор «..» дві крапки** – використовується для пошуку діапазонів між числами;
- **оператор «@»** – для пошуку по тегах у соц. мережах;
- **оператор «#»** – пошук за хештегами.

2.3.1. Документні оператори Google

- **site:** – аналогічно Яндексу шукає по зазначеному сайту або домену;
- **link:** – пошук сторінок, що посилаються на зазначений сайт;
- **related:** – пошук сторінок зі схожим вмістом;
- **info:** – за допомогою цього оператора можна одержати відомості про веб-адресу, у тому числі посилання на кеш-версію сторінки, схожі сайти, а також сторінки, що посилаються на зазначену вами;
- **cache:** – перегляд кеш-версії сторінки;

- **filetype:** – пошук у зазначених типах файлів, можна вказати розширення;
- **movie:** – пошук інформації про фільми;
- **daterange:** – пошук сторінок, проіндексованих за зазначений проміжок часу;
- **allintitle:** – пошук сторінок, у яких слова із запиту перебувають в title;
- **intitle:** – теж саме, але частина запиту може втримуватися й в іншій частині сторінок;
- **allinurl:** – пошук сторінок, що містять всі слова запиту в url;
- **inurl:** – теж саме, але для одного слова;
- **allintext:** – тільки в тексті;
- **intext:** – для одного слова;
- **allinanchor:** – пошук за словами в анкорах;
- **define:** – пошук сторінок з визначенням зазначеного слова.

2.4. Оператори пошукової системи Bing

- **contains:** – залишає результати із сайтів, які містять посилання на типи файлів, які ви вказуєте;
- **ext:** – повертає тільки веб-сторінки з розширенням, що ви вказуєте;
- **filetype:** – повертає тільки веб-сторінки, створені з типом файлу, що ви вказуєте;
- **inanchor:** або **inbody:** або **intitle:** – ці ключові слова повертають веб-сторінки із заданим терміном у метаданих, наприклад якорі, тексті й назві сайту;
- **ip:** – знаходить сайти, які розміщені по певній IP-адресі;
- **language:** – повертає веб-сторінки певною мовою;
- **loc:** або **location:** – повертає веб-сторінки з певної країни чи регіону;
- **prefer:** – дає пріоритет умові пошуку або іншому операторові, щоб зосередити результати пошуку;
- **site:** – повертає веб-сторінки, які належать зазначеному сайту;
- **feed:** – знаходить канали RSS або Atom на веб-сайті по термінах, які ви шукаєте;
- **hasfeed:** – знаходить веб-сторінки з каналами RSS або Atom на веб-сайті по термінах, які ви шукаєте;
- **url:** – перевіряє, є чи зазначений домен або веб-адреса в індексі Bing.

ДОДАТОК 3. GOOGLE ОПЕРАТОРИ В ПОШУКОВИХ ЗАПИТАХ

Для уточнення результатів пошуку можна використовувати спеціальні оператори й пунктуацію. За винятком випадків, описаних нижче, пунктуація в пошукових запитах не враховується.

Пунктуація й символи

Хоча перераховані нижче символи підтримуються, їхне включення в запити не завжди покращує результати пошуку. Якщо пунктуація не зробить пошук точнішим, Google може запропонувати результати, отримані без неї.

Символ	Інструкції
+	Пошук +сторінок або груп крові+ Приклади: +Chrome й AB+
@	Пошук по тегах у соціальних мережах Приклад: @agoogler
\$	Пошук цін Приклад: nikon \$400
#	Пошук популярних хештегів для актуальних тем Приклад: #деньнародження
-	Якщо ввести дефіс перед словом або адресою сайту, сторінки із цим словом або із цього сайту будуть виключені з результатів пошуку. Це корисно при пошуку слів, які мають кілька значень, наприклад "Ягуар" (марка машини) і "ягуар" (тварина) Приклади: ягуар швидкість -автомобіль і панди -site:wikipedia.ru
"	Якщо помістити слово або фразу в лапки, у результатах пошуку будуть показані тільки сторінки із цими словами в тому ж порядку, у якому вони зазначені в лапках. Використовуйте лапки, тільки якщо ви шукаєте точне слово або фразу. У протилежному випадку з пошуку будуть виключені багато корисних результатів Приклад: "мільйон червоних троянд"
*	Зірочка замінює будь-яке невідоме слово в запиті Приклад: "сім * відміряй, один * відріж"
..	Щоб одержати результати, що містять числа в заданому діапазоні, розділіть значення двома крапками без пробілу Приклад: Фотоапарат 5000..9000 гривень

Оператори в пошукових запитах

Оператори пошуку – це слова, що додають до пошукових запитів для уточнення результатів. Їх необов'язково запам'ятовувати: всі ці можливості доступні на сторінці «Розширений пошук» Google.

Оператор	Інструкції
site:	Використовуйте цей оператор, щоб одержати результати з певних сайтів або із зазначених доменів Приклади: олімпійські ігри site:ntv.ru й олімпійські ігри site:.net
related:	Використовуйте цей оператор, щоб знайти сайти зі схожим контентом Приклад: related:kommersant.ru
OR	Використовуйте цей оператор, щоб знайти сторінки, що містять хоча б одне з декількох слів Приклад: олімпіада OR олімпійські ігри
info:	Використовуйте цей оператор, щоб одержати відомості про адресу в Інтернеті, у тому числі кешовану версію сторінки, схожі сайти, а також сторінки з посиланнями на зазначену вами Приклад: info:google.ru
cache:	Використовуйте цей оператор, щоб довідатися, як виглядала певна сторінка, коли роботи Google сканували її востаннє. Наприклад, якщо вас цікавить кешована версія сайту МГУ, використовуйте такий запит: cache:msu.ru

Примітка. При пошуку з використанням операторів не додавайте пробіл між оператором і пошуковим запитом. Наприклад, пошуковий запит **site:korrespondent.ru** спрацює, а запит **site: korrespondent.ru** – ні.

ДОДАТОК 4. YANDEX ОПЕРАТОРИ В ПОШУКОВИХ ЗАПИТАХ

Мова запитів дозволяє точніше сформулювати ваш пошуковий запит.

Найбільш популярні пошукові оператори представлені у фільтрах розширеного пошуку, з іншими можна ознайомитися в таких розділах:

- морфологія;
- пошуковий контекст;
- документні оператори.

4.1. Морфологія

При пошуку з урахуванням морфології приймаються до уваги:

- форма заданого слова (відмінок, рід, число, відмінювання й т.д.);
- частина мови (іменник, прикметник, дієслово й т.д.).

За замовчуванням Яндекс шукає всі форми слова, зазначеного в запиті.

Наприклад, при запиті [розповів] пошук буде здійснюватись по дієслівних формах: «розповісти», «розповім», «розповідати» і т.д., але не по однокорінних словах типу «розповідь», «оповідач».

Для обмеження області пошуку можна використовувати спеціальні оператори, які дозволяють отримати у видачі тільки документи, що містять запитуване слово в заданій формі.

Оператор	Опис	Синтаксис	Приклад запиту
!	Пошук слова в заданій формі. Можна використовувати кілька операторів ! в одному запиті	!слово	[!розповів] Будуть знайдені документи зі словом «розповів» у заданій формі.
!!	Пошук слова, початкова форма якого зазначена в запиті	!!слово	[!!день] Будуть знайдені документи, що містять кожну з форм слова «день» (наприклад, «дні», «судень»).

4.2. Пошуковий контекст

Ви можете конкретизувати пошуковий запит за допомогою операторів, які уточнюють наявність і взаємне розташування запитуваних слів у документі.

Яндекс шукає всі форми слова, зазначеного в запиті. Виключення становлять випадки, коли використовуються оператори ! і ".

Оператор	Опис	Синтаксис	Приклад запиту
+	Пошук документів, у яких обов'язково є присутнім виділене слово. Можливе використання декількох операторів + в одному запиті.	слово ₁ + слово ₂	[Шевченко +Канів +Київ] Будуть знайдені документи, у яких обов'язково є слова «Канів» й «Київ» і може бути присутнім слово «Шевченко».
"	Пошук за цитатою. Пошук документів, що містять слова запиту в заданій послідовності й формі.	"слово ₁ слово ₂ ... слово"	["Щедрий вечір добрим людям"] Будуть знайдені документи, що містять дану цитату.

*	<p>Пошук за цитатою з пропущеним словом (словами).</p> <p>Один оператор * відповідає одному пропущеному слову. Можливе використання декількох операторів * в одному запиті.</p> <p>Увага! Використовується тільки в складі оператора «».</p>	<p>"слово₁ * слово₂ ... слово"</p> <p>Оператор відокремлюється пробілами.</p>	<p>[«Щедрий вечір * людям»]</p> <p>Будуть знайдені документи, що містять дану цитату, включаючи пропущене слово.</p>
&	<p>Пошук документів, у яких слова запиту, об'єднані оператором, зустрічаються в одній пропозиції.</p> <p>Можливе використання декількох операторів & в одному запиті.</p>	<p>слово₁ & слово₂</p> <p>Оператор відокремлюється пробілами.</p>	<p>[культура & єгипет]</p> <p>Будуть знайдені документи, у яких є одночасно слова «культура» й «єгипет».</p> <p>[культура & єгипет & сучасний]</p> <p>Будуть знайдені документи, у яких є одночасно слова «культура», «єгипет» й «сучасний».</p>
&&	<p>Пошук заданих слів у межах документа. При визначенні релевантності знайдених документів (і їхнього положення в результатах пошуку) ураховуються всі слова запиту.</p> <p>Можливе використання декількох операторів && в одному запиті.</p>	<p>слово₁ && слово₂</p> <p>Оператор відокремлюється пробілами.</p>	<p>[Індія && Великобританія]</p> <p>Будуть знайдені документи, у яких присутні слова «Індія» й «Великобританія», незалежно від відстані між ними в тексті.</p>
<<	<p>Пошук заданих слів у межах документа. При визначенні релевантності знайдених документів запиту (і їхнього положення в результатах пошуку) ураховується тільки слово (слова), розміщене до оператора.</p>	<p>слово₁ << слово₂</p> <p>Оператор відокремлюється пробілами.</p>	<p>[Індія << Великобританія]</p> <p>Будуть знайдені документи, у яких присутні слова «Індія» й «Великобританія», але визначення релевантності знайдених документів буде виконуватися тільки по слову «Індія».</p>
	<p>Пошук документів, у яких є присутнім будь-яке слово із запиту.</p> <p>Можливе використання декількох операторів в одному запиті.</p>	<p>слово₁ слово₂ ... слово</p> <p>Оператор відокремлюється пробілами.</p>	<p>[яхта човен корабель]</p> <p>Будуть знайдені документи, у яких є присутнім хоча б одне зі слів запиту: «яхта», «човен» або «корабель».</p>

	<p>Пошук документів, у яких відсутнє задане слово.</p> <p>Виключається тільки слово, перед яким стоїть оператор.</p> <p>Ідентичний операторові ~.</p> <p>Можливе використання декількох операторів «~» в одному запиті.</p>	слово ₁ слово ₂ ... ~ слово	<p>[зоопарк ~ московський]</p> <p>Будуть знайдені документи, у яких є присутнім слово «зоопарк», але немає слова «московський».</p>
~~	<p>Пошук документів, у яких відсутнє задане слово.</p> <p>Виключається тільки те слово, перед яким стоїть оператор.</p> <p>Ідентичний операторові ~.</p> <p>Можливе використання декількох операторів ~~ в одному запиті.</p>	слово ₁ слово ₂ ... ~~ слово Оператор відокремлюється пробілами.	<p>[зоопарк ~~ київський]</p> <p>Будуть знайдені документи, у яких є присутнім слово «зоопарк», але немає слова «київський».</p>
~	<p>Пошук документів, у яких задане слово може бути присутнім на сторінці, але не в одній пропозиції зі словом, зазначенним до оператора.</p>	слово ₁ ~ слово ₂ Оператор відокремлюється пробілами.	<p>[золоте ~ руно]</p> <p>Будуть знайдені документи, у яких слово «руно» може бути присутнім, але не в одній пропозиції зі словом «золотий».</p>
0	<p>Угруповання слів при складних запитах.</p> <p>Усередині укладеної в дужки групи також можуть бути використані будь-які оператори.</p>	слово ₁ оператор (слово ₂ слово ₃ ... слово) Наявність пробілу перед символом (визначається синтаксисом оператора, що є перед ним.	<p>[машина -(магазин купити)]</p> <p>Будуть знайдені документи, у яких є слово «машина» і відсутні слова «магазин» й «купити».</p> <p>[машина +(магазин !Продаж)]</p> <p>Будуть знайдені документи, у яких може бути присутнім слово «машина» й обов'язково є «магазин» й «Продаж». При цьому слово «Продаж» — у заданій формі й із заголовної букви.</p>

4.3. Документні оператори приклади використання

Уточнити пошуковий запит можна з допомогою даних, що відносяться до службової інформації про сторінки. Наприклад, можна обмежити пошук по типу файла, хосту і т.д.

Як правило, документний оператор вказується після тексту пошукового запиту й відокремлюється від нього пробілом. В одному запиті одночасно можна використати декілька операторів, так само розділяючи їх пробілом. Наприклад, [пошук site:www.yandex.ru mime:pdf].

При використанні операторів, де як параметр задається ім'я хоста (url, host й rhost), варто вказувати головне дзеркало сайта. Наприклад, [host:lib.ru], а не [host:www.lib.ru]. Довідатися, чи є сайт головним дзеркалом, можна, додавши хост до бази Яндексу.

Оператор	Опис	Синтаксис	Приклад запиту
url:	Пошук по сторінках, розміщених за заданою адресою (URL). Щоб знайти всі документи, адреси яких починаються із заданого значення, поставте наприкінці URL символ *. URL у запиті варто взяти в лапки, якщо він містить один з наступних символів: ' " () _	url:повний URL url:ім'я хоста/категорія/*	[пошук url:"uk.wikipedia.org/wiki/Ранжування_(пошукові системи)"] Будуть знайдені документи, що містять слово «пошук» і розміщені за заданою адресою. [яндекс url:uk.wikipedia.org/wiki/*] Будуть знайдені документи, що містять слово «яндекс» й адреси яких починаються з uk.wikipedia.org/wiki.
site:	Пошук по всіх піддоменах і сторінках зазначеного сайту.	site:URL сайту	[яндекс site:ukr.net] Будуть знайдені документи, що містять слово «яндекс» і розміщені на піддоменах і сторінках сайту ukr.net.
host:	Пошук по сторінках, розміщених на даному хості. Ідентичний операторові url: з заданим іменем хоста.	host:ім'я хоста у форматі www.домен другого рівня.домен верхнього рівня	[пошук host: www.yandex.ru"]] Будуть знайдені документи, що містять слово «пошук» і розміщені на хості www.yandex.ru.
rhost:	Пошук по сторінках, розміщених на даному хості. Ідентичний операторові host:, але ім'я хоста записується у зворотному порядку: спочатку домен верхнього рівня, потім	rhost:ім'я хоста у форматі домен верхнього рівня.домен другого рівня.www rhost:ім'я хоста у форматі домен	[новини rhost:com.livejournal.www] Будуть знайдені документи, що містять слово «новини» і розміщені на домені livejournal.com. [новини rhost:com.livejournal.*] Будуть знайдені документи, що

	домен другого рівня й т.д. Для пошуку по всім піддоменах заданого домену наприкінці URL поставте символ *.	верхнього рівня домен другого рівня.*	містять слово «новини» і розміщені на всіх піддоменах домену livejournal.com.
domain:	Пошук по сторінках, розташованих на заданому домені.	domain:домен	[яндекс domain:ua] Будуть знайдені документи, що містять слово «яндекс» і розміщені на домені ua.
mime:	Пошук по документах за заданим типом файлу. Підтримувані типи файлів представлені на сторінці розширеного пошуку.	mime:тип файла	[анкета закордонний паспорт mime:doc] Будуть знайдені документи у форматі doc, що містять слова «анкета» й «закордонний паспорт».
lang:	Пошук по сторінках заданою мовою: українська (uk) англійська (en) французька (fr) німецька (de) турецька (tr) російська (ru)	lang:мова	[passsport lang:en] Будуть знайдені документи англійською мовою, що містять слово «passport».
date:	Пошук по сторінках з обмеженням по даті їхньої останньої зміни. <i>Примітка.</i> Рік зміни вказується обов'язково. Місяць і день можна замінити символом *.	Зміна сторінки: відповідає зазначеній даті: date:ПППРММДД перебуває в інтервалі: date:ПППРММДД.. ПППРММДД перебуває до або після зазначеної дати (<, <=, >, >=): date:<ПППРММДД Дата зміни сторінки зазначена частково: date:ПППР**	Будуть знайдені документи, дата зміні яких: відповідає 10.10.2014: [фестиваль date:20141010]; перебуває в інтервалі між 10.10.2014 й 10.11.2014 включно: [фестиваль date:20141010..20141110]; пізніше 10.11.2014: [фестиваль date:>20141110]; відповідає 2014 року: [фестиваль date:2014**].

Кафедра комп'ютерних наук

Звіт
до лабораторної роботи № 3 з курсу
«Інтернет-технології опрацювання консолідованих
інформаційних ресурсів»

**Тема: «Дослідження алгоритмів роботи відомих пошукових
систем, алгоритми ранжування WEB-сайтів»**

Зразок титулки

Виконав:

ст. _____
група _____

Перевірив:

ст. викл. каф. КН Козбур Г.В.

Дата _____ 201 ____ р.

Оцінка	
--------	--

