

Лабораторна робота № 6

на тему: Застосування мови запитів пошукової системи Google для розв'язання складних інформаційно-пошукових задач

Мета роботи: ознайомитися з мовою запитів пошукової системи Google, вивчити оператори, які використовуються у мові запитів, та їхній синтаксис, навчитися застосовувати мову запитів для формування складних запитів до пошукової системи з метою вирішення складних інформаційно-пошукових задач.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Для максимального зниження рівня інформаційного шуму у результатах пошуку необхідно використовувати розширені можливості пошукових систем. Доступ до розширених можливостей та їх застосування реалізується або через панель розширеного пошуку (лабораторна робота №3), або через мову запитів до пошукової системи.

Мова запитів передбачає наявність спеціальних операцій пошуку, які дають змогу параметризувати запит, щоб отримати якомога точніший результат. Використовуючи відповідний синтаксис мови запитів, створюється запит до конкретної пошукової системи. Синтаксис операцій пошукової системи Google подано в таблиці 1.

Таблиця 1.

Синтаксис операцій пошукової системи Google

Синтаксис операцій Google	Застосування
“запит”	Цитата, слова підряд в точній формі або конкретна морфологічна форма слова
(складний запит)	Групування в складних запитах. Використовується для того, щоб вказати, що під дію певного оператора потрапляє не одне слово, а група слів
запит1 запит2 або запит1 AND запит2	Логічне “І”: вживання слів в межах одного документа

запит1 запит2	Логічне “АБО”
запит1 OR запит2	
–запит	Логічне “НЕ”
запит1 * запит2	Встановлення відстані між словами
~запит	Пошук синонімічних одиниць
число1..число2	Пошук числових значень у заданому проміжку
allintitle:запит	Пошук усіх слів в заголовку
intitle:запит	Пошук у заголовку сторінки
allintext:запит	Пошук усіх слів запиту у тексті веб-сторінки
intext:запит	Пошук у тексті сторінки
inurl:запит	Пошук в URL-адресі
site:URL	Пошук проіндексованих сторінок вказаного веб-сайту
link:URL	Пошук сторінок, які мають посилання на головну сторінку зазначеного ресурсу (зворотні посилання)
filetype:розширення	Пошук результатів у заданому форматі файла
related:URL	Пошук пов’язаних однотипних сторінок

Розглянемо деякі оператори та їхній синтаксис детальніше.

1) запит1[пробіл]запит2 або запит1 AND запит2

Google інтерпретує слова, які записані у стрічці пошуку через пробіл, як окремі слова. Тому, якщо запит1 і запит2 складаються з одного слова, то пошукова система інтерпретує їх так: “і запит1 і запит2”. Тобто, можна сказати, що пошукова система завжди використовує цей оператор за змовчанням, якщо у стрічці пошуку є більше ніж одне слово.

2) запит1|запит2 або запит1 OR запит2

Можливість вказувати синоніми або альтернативні форми дає оператор OR (записаний літерами верхнього регістру!) чи | (вертикальна риска).

Оператор OR, для якого також можна використовувати | (вертикальну риску), відноситься до пошукових термінів, які безпосередньо прилягають до нього:

Львів OR Тернопіль

Львів | Тернопіль

У першому і другому прикладах пошукова система знайде сторінки, які містять або “Львів” або “Тернопіль” або обидві умови, але не сторінки, які не містять ні “Львів”, ні “Тернопіль”.

Львів OR Тернопіль OR Хмельницький

У третьому і четвертому прикладах будуть знайдені сторінки, що містять одне, два або всі три слова “Львів”, “Тернопіль” і “Хмельницький”.

Примітка: Якщо оператор написати з малої літери, Google інтерпретує його як слово для пошуку, а не як оператор.

Примітка: На відміну від OR, | (вертикальна риска) не повинна бути оточена пробілами.

Для групування складних слів і фраз, використовуються лапки (“”) або дужки (), залежно від того, чи здійснюється пошук точної фрази чи лише групи слів.

(Львівська кав'ярня) OR (Тернопільська площа)

"Львівська кав'ярня" OR "Тернопільське озеро".

3) запит1 * запит1

Символ зірочки (*), відомої також як “шаблон”, використовується, щоб замінити одне або декілька слів у фразі (укладеній в лапки). Кожен символ * представляє лише одне або декілька слів. Google розглядає * в якості заповнювача для одного або декількох слів. Наприклад, запит у формі *["Природа - це * яку * серцем"]* говорить Google, щоб знайти сторінки, що містять фразу, яка починається з "Природа - це", а потім одне або декілька слів, тоді слово “яку”, а тоді "серцем". Фразами, які відповідають усім цим вимогам є:

- *Природа — це казка, яку треба читати серцем*
- *Природа - це казка, яку читаєш серцем*
- *Природа – це казка, яку потрібно берегти всім серцем.*

4) -запит

Символ “-” усуває з результатів пошуку сторінки, які містять термін або фразу, перед яким стоїть цей знак. Для того, щоб усунути з результатів пошуку певний термін, необхідно перед цим терміном поставити знак “-“. Не слід ставити пробіл між символом та терміном. Якщо термін складний або складається з кількох слів, то його треба взяти в лапки.

- *львівська політехніка -інститут –державний*
- *львівська політехніка -"віртуальне середовище"*

5) ~запит

Тильда (~) – це оператор, який дозволяє знайти синоніми до слова, якому передує символ (~).

Оператор (~) бере слово, яке стоїть безпосередньо після нього, і шукає для цього конкретного слова його синоніми. Він також шукає терміни з альтернативними закінченнями. Оператор тильда працює найкраще, коли застосовується до загальних термінів та термінів з багатьма синонімами. Як і оператор (-), ~ (тильда) не відокремлюється пробілом від слова, для якого здійснюється пошук.

- *~лінгвістика*
- *~дешеві авіаквитки*

6) число1..число2

Для того, щоб знайти сторінки, на яких містяться числа з певного діапазону, використовується оператор (..). Для цього необхідно вказати два числа – межі діапазону, між ними застосувати оператор (..).

Наприклад:

- *nokia 500..700 – пошук всіх моделей у заданому проміжку*
- *гетьман Хмельницький 1600..1700*

7) allintitle:запит

intitle:запит

Ці два оператори подібні за своєю функцією – вони виконують пошук лише у заголовку об'єкта, пошук якого здійснюється. Однак, ці оператори по-різному використовують складові запиту, які стоять безпосередньо після нього.

Якщо запит формується з оператором **allintitle**, наприклад *[allintitle: університет ECTS]*, то пошукова система здійснюватиме пошук у заголовку сторінки абсолютно всіх слів (у цьому випадку – обох), які є у стрічці пошуку (або всіх до наступного оператора).

Якщо запит формується з оператором **intitle**, наприклад *[intitle: університет ECTS]*, то пошукова система здійснюватиме пошук у заголовку сторінки лише слова “університет”, а слово “ECTS” шукатиме будь-де на сторінці.

Таким чином, *[allintitle: університет ECTS] = [intitle: університет intitle: ECTS]*

8) **allintext:запит**

intext:запит

Так само, як і у попередньому випадку, ці два оператори подібні за своєю функцією – вони виконують пошук у тексті об'єкта, пошук якого здійснюється. Однак, ці оператори по-різному використовують складові запиту, які стоять безпосередньо після нього.

Наприклад:

- *[allintext: університет ECTS]* – пошук обох слів лише у тексті веб-сторінки
- *[intext: університет ECTS]* – пошук слова “університет” у тексті сторінки, а слова “ECTS” – будь-де на сторінці, зокрема в заголовку.

9) **inurl:запит**

Оператор **inurl** виконує пошук у межах URL-адреси веб-сайту. Наприклад, такий оператор використовується тоді, коли необхідно знайти офіційні портали і веб-сайти українських міст. Можна припустити, що сайт міста може містити в адресі складову “misto”. Тоді, для відбору всіх сторінок, в адресі яких є така складова, використовується оператор **inurl**:

Наприклад: `[inurl:misto]`

10) **site:URL**

Оператор **site** використовується для пошуку всіх існуючих сторінок заданого веб-сайту, або для пошуку конкретної інформації в межах заданого сайту. Наприклад, щоб дізнатися, скільки сторінок належить до офіційного веб-сайту львівської політехніки, достатньо ввести запит *[site:lp.edu.ua]*.

Складний запит, наприклад *[site:lp.edu.ua intitle:інститут]* дозволить виявити, які сторінки веб-сайту Львівської політехніки містять у заголовку слово “інститут”.

11) **link:URL**

Оператор **link** дозволяє здійснити пошук тих сторінок, які містять посилання на головну сторінку веб-сайту, URL-адреса якого введена у запиті.

Наприклад, для того, щоб дізнатися, хто посилається на сторінку Вашого веб-сайту, тобто перевірити зворотні посилання, достатньо сформулювати запит такого вигляду: [link:misto.ridne.net].

12) filetype:розширення

Оператор **filetype** дозволяє обмежити формат файлів у результатах пошуку. Сьогодні доступні такі формати файлів:

- Adobe Acrobat PDF (pdf)
- Adobe Postscript (ps)
- Autodesk DWF (dwf)
- Google Планета Земля KML (kml)
- Google Планета Земля KMZ (kmz)
- Microsoft Excel (xls)
- Microsoft Powerpoint (ppt)
- Microsoft Word (doc)
- Rich Text Format (rtf)
- Shockwave Flash (swf)

Наприклад, пошук реферату про картографічні сервіси у форматі PDF або DOC: *[allintext: реферат картографічний сервіс filetype:doc OR filetype:pdf]*.

ХІД РОБОТИ

ЗАВДАННЯ 1:

1. Надайте п'ять (5) власних прикладів ВДАЛОГО! застосування мови запитів для розв'язання завдання, яке Ви сформулюєте особисто. Розробіть завдання і запити для різних операторів пошуку. Реалізуйте кожен з розроблених запитів і надайте відповідні копії з екрану. Заповніть таблицю 1, структура якої подана нижче.

Таблиця 1.

Запропоноване вирішення пошукового завдання

№	Формулювання мети пошуку	Структура запиту	Отриманий результат

2. У таблиці нижче наведений перелік міністерств України та їхніх офіційних веб-сайтів. Залежно від Вашого порядкового номера у списку

підгрупи, визначте відповідний веб-сайт для виконання подальших завдань.

Порядковий номер студента у списку підгрупи	Назва міністерства	Офіційний веб-сайт
1.	Міністерство соціальної політики України	www.mlsp.gov.ua
2.	Міністерство юстиції України	www.minjust.gov.ua
3.	Міністерство оборони України	www.mil.gov.ua
4.	Міністерство освіти і науки України	www.mon.gov.ua
5.	Міністерство фінансів України	www.minfin.gov.ua
6.	Міністерство економічного розвитку і торгівлі України	www.me.gov.ua
7.	Міністерство культури України	mincult.kmu.gov.ua
8.	Міністерство екології та природних ресурсів України	www.menr.gov.ua
9.	Міністерство інфраструктури України	www.mtu.gov.ua
10.	Міністерство охорони здоров'я України	www.moz.gov.ua
11.	Міністерство закордонних справ України	www.mfa.gov.ua
12.	Міністерство з питань надзвичайних ситуацій	www.mns.gov.ua
13.	Міністерство енергетики та вугільної промисловості України	mpe.kmu.gov.ua
14.	Міністерство внутрішніх справ України	mvs.gov.ua
15.	Міністерство аграрної політики України	minagro.gov.ua

- Для відповідного веб-сайту визначте кількість сторінок, які посилаються на його головну сторінку. Надайте відповідну копію з екрану з відповідною назвою рисунка.
- Для відповідного веб-сайту визначте кількість сторінок, які містить заданий веб-сайт. Надайте відповідну копію з екрану з відповідною назвою рисунка.
- Визначте, хто зараз займає посаду міністра відповідного міністерства. Надайте відповідну копію з екрану з відповідною назвою рисунка.
- Визначте, скільки є на відповідному веб-сайті таких сторінок, де не згадується особа міністра відповідного міністерства, але на сторінці фігурує слово “закон” (у будь-якій формі). Надайте відповідну копію з екрану з відповідною назвою рисунка.

7. Знайдіть сторінки на відповідному веб-сайті, на яких не згадується слово “міністерство” або “міністр” (у точній формі), а у заголовку сторінки фігурує слово “президент” або “кабмін” (у будь-якій формі). Надайте відповідну копію з екрану з відповідною назвою рисунка.
8. Знайдіть звіти міністерства (у формі документа) за період 2008-2011 роки. Надайте відповідну копію з екрану з відповідною назвою рисунка.
9. Знайдіть ті веб-сторінки відповідного веб-сайту, адреса яких будь-яким чином пов’язана з 2011 роком. Надайте відповідну копію з екрану з відповідною назвою рисунка.
10. Знайдіть ті веб-сторінки відповідного веб-сайту, в яких в заголовку згадується Верховна Рада України і слово *закон*. Надайте відповідну копію з екрану з відповідною назвою рисунка.

ЗАВДАННЯ 2:

1. У таблиці нижче наведений перелік міст України. Залежно від Вашого порядкового номера у списку підгрупи, визначте відповідне місто за списком для виконання подальших завдань.

Порядковий номер студента у списку підгрупи	Назва міста
1.	Ужгород
2.	Мукачево
3.	Івано-Франківськ
4.	Чернівці
5.	Луцьк
6.	Тернопіль
7.	Вінниця
8.	Донецьк
9.	Сімферополь
10.	Харків
11.	Запоріжжя
12.	Кіровоград
13.	Луганськ
14.	Ялта
15.	Черкаси

2. Знайдіть рік заснування міста. Надайте відповідну копію з екрану з відповідною назвою рисунка.
3. Знайдіть сторінки на яких згадується теперішній або колишній мер міста. Надайте відповідну копію з екрану з відповідною назвою рисунка.
4. Знайдіть сторінки, адреса яких будь-як пов’язана з назвою міста. Надайте відповідну копію з екрану з відповідною назвою рисунка.
5. Знайдіть сторінки, в яких у назві сторінки фігурує назва міста, а у тексті – фразу, що це моє рідне місто. Надайте відповідну копію з екрану з відповідною назвою рисунка.

6. Знайдіть файли текстових документів або електронних таблиць, в яких у тексті згадується місто, туризм і прибуток. Надайте відповідну копію з екрану з відповідною назвою рисунка.
7. Знайдіть інформаційні сторінки (типу info), в заголовку яких згадується місто. Надайте відповідну копію з екрану з відповідною назвою рисунка.
8. Подумайте, як можна знайти сайти новин. Запропонуйте спосіб (способи) пошуку сторінок новин про місто. Надайте відповідну копію(-ї) з екрану з відповідною назвою рисунка.

ЗАВДАННЯ 3:

1. За допомогою мови запитів до пошукової системи, визначте пропущені слова. Надайте відповідні копії з екрану з відповідними назвами рисунків.
 - Без ... нема науки
 - У суперечці ... та ... перемагає той, хто справжній.
 - Знати багато мов — значить мати багато ... до одного ...
 - Краще різка ... , ніж зрадницька ...
2. Доберіть два власні приклади використання мови запитів для визначення пропущених слів та конструкцій. Надайте відповідні копії з екрану з відповідними назвами рисунків.

СТРУКТУРА ТА ОФОРМЛЕННЯ ЗВІТУ ДО ЛАБОРАТОРНОЇ РОБОТИ

Звіт з лабораторної роботи оформлюється на аркушах формату А4, які заповнюються з однієї сторони. Текст повинен бути рукописним або друкованим на принтері. Використання кольорових чорнил дозволяється лише для ілюстративних матеріалів.

Звіт до лабораторної роботи формується відповідно до змісту і повинен містити такі розділи:

- титульна сторінка;
- мета роботи;
- порядок виконання роботи;
- опис усіх етапів виконання роботи;
- опис отриманих результатів;
- висновки за результатами роботи.

Звіт зшивається з лівої сторони листів формату А4. Титульна сторінка звіту обов'язково друкується на принтері.