ПРАКТИЧНЕ ЗАНЯТТЯ № 7

**Тема:** Інформаційний пошук. Пошукові системи

**Мета:** ознайомитися з інформаційного пошуку у різних пошукових системах.

**Завдання:** Ознайомитися із теоретичним матеріалом та, користуючись прикладами, зробити розширений пошук у GoogleScolar по одній із наведених фраз:

1. Кількість людей, що вмерли у 2019 році від туберкульозу
2. Відкриття у фізіології людини і тварин 1950-1970 рр.
3. Відкриття у цитології 1970-1990 рр.
4. Характеристика ссавці (виключити хижаків).
5. Нобелевські лауреати у 1990-1995 рр.
6. Будова вірусу (тільки відео).
7. Епітеліальна тканина (тільки аудіо-файл).
8. Цитокіни (тільки назви документів).
9. Кількість людей, що захворіли на кір в Україні у 2018-2020 рр.
10. Кількість людей, що відмовилися від вакцинації у 2019 р. в Україні.
11. Характеристика черепах (виключити суходольних черепах).
12. Кількість відкритих видів тварин у 2010-2019 рр.
13. Кількість відкритих видів рослин у 2010-2019 рр.
14. Характеристика тканин (виключити текстильні тканини). (мається на увазі тільки гістологія)
15. Нобелевські лауреати у 1970-1980 рр.
16. Вчені-біологи 1930-1940 рр.
17. Кількість людей, хворих на гіпертонію у 2019 рр. у віці 30-40 рр.
18. Серцево-судинна система (тільки назви документів).
19. Серцево-судинна система (тільки відео-файл).
20. Домашні тварини (щоб у результаті пошуку були наявні і собаки, і коти, і рибки, і черепахи одночасно у одному запиті, а сайти, що містять один вид домашніх тварин або інші не видавав у пошууку)

Наукова інформація - відомості, сукупність даних, знань, детальна, систематизована подача певного відібраного матеріалу.

- є наслідком науково-дослідницької діяльності, яка формується в процесі пізнання навколишнього світу і відображає його закономірності.

- документовані або публічно оголошені відомості про досягнення науки, техніки, виробництва, отримані в процесі науково-дослідної, дослідно-конструкторської, виробничої діяльності

Для інформаційного пошуку розробляють:

●алгоритми

●підходи

●стратегії

Для здійснення створюють:

●методи інформаційного пошуку

●засоби інформаційного пошуку

●комп'ютерні пошукові програми

а також:

представлення даних, інформації, знань

представлення інформації в сучасних інформаційних сховищах

багатомовний інформаційний пошук

одночасний інформаційний пошук

розподілений інформаційний пошук

суспільний інформаційний пошук

Основні джерела наукової інформації

Первинні

●Монографія ● Періодичні видання – журнал

●Спеціальні видання (аналітичні, статистичні збірники, доповіді тощо)

●Патентно-ліцензійні видання

●Дисертації

●Матеріали конференцій

Вторинні:

●Енциклопедії, індекси, покажчики, бази даних

Рецензування — процедура розгляду та оцінки наукових статей та монографій дослідниками – фахівцями в тій же галузі (Peer review -«розгляд рівними», колегами)

Інформаційний Пошук

●пошук неструктурованої документальної інформації

●пошук інформації в документах, пошук самих документів, добуття метаданих з документів, пошуку тексту, зображень, відео та звуку у локальних реляційних базах даних, у гіпертекстових базах даних (в т.ч. Інтернет)

●міждисциплінарна область науки, знаходиться на перетині когнітивної психології, інформатики, інформаційного дизайну, лінгвістики, семіотики, бібліотечної справи та статистики. (Вперше виділив як міждисциплінарну галузь відомий угорський дослідник Золтон Жулен у 1989 році.)

●об'єктом інформаційного пошуку є текстова інформація, зображення, аудіо, відео інформація

Найвідомішим прикладом можна назвати пошукові системи в Інтернеті

Ефективність пошукових систем:

●Пошукова відповідність (precision) визначає частку відповідних документів серед відібраних на запит. Пошукова відповідність визначає якість отриманого результату інформаційного пошуку.

●Пошукова якість (recall) визначає частку отриманих системою відповідних до запиту документів серед загального числа відповідних до запиту документів у збірці. Загальне число відповідних до запиту документів завжди є невідомим і може бути встановлене лише при повному перегляді збірки людиною.

●Крім того роботу пошукових систем оцінюють швидкодією — часом, за який отримують список відповідних до запиту документів.

Розглянемо детальніше Пошукові системи

Однією з перших пошукових систем була система Archie, яку розробили 1990 р. в Канаді. Її використовували для пошуку назв веб-сторінок. Однак програма не вміла шукати по змісту веб-сторінок.

Минуло кілька десятиліть, і з’явились пошукові системи Google, Bing, Yahoo, Ask.com, AOL, Baidu, Yandex та інші. Вони використовують складні комп’ютерні алгоритми для класифікації величезної кількості веб-сторінок.

●У рейтинзі Alexa (https://www.alexa.com/topsites) серед 50 найбільш відвідуваних сайтів, ці пошукові системи займають місця:

●1 - Google.com (США)

●4 - Baidu.com (Китай)

●10 - Yahoo.com (США)

●16 - Live.com (США)

●26 - Yandex.ru (Росія)

●27 - Bing.com (США)

Принципи роботи більшості пошукових систем однакові. Коли користувач хоче щось знайти, він вводить якесь слово чи фразу, - це називається пошуковий запит. Далі система порівнює цей запит зі своїм каталогом веб-сторінок і показує користувачеві найвідповідніші варіанти на сторінці результатів пошуку.

Мета пошукової системи – відібрати найбільш відповідні результати, щоб допомогти користувачам знайти потрібний вміст. На сторінці результатів відображаються переважно посилання на сайти, зображення, рекламні оголошення, карти, відео тощо. Базові завдання, які вони виконують: Сканування – тобто вивчення виявленого вмісту, класифікація кожного фрагмента вмісту (індексування), і визначення рейтингу - вибір найбільш корисного для автора змісту.

Розгляньмо докладніше, як пошукові системи сканують Інтернет, виявляючи такий вміст, як веб-сторінки, зображення та відео. Кожна система використовує комп’ютерні програми для аналізу веб-сторінок. Такі програми називають веб-сканерами, павуками або роботами. Вони постійно переходять від сторінки до сторінки, шукаючи нові посилання або вміст для включення в індекс.

Індексування – друга складова процесу. Це величезний список з усіх сторінок, що знайшли роботи. Пошукова система використовує індекс як джерело інформації, що відображається у результатах.

Але не вся інформація, виявлена роботом, додається до індексу. Пошукові системи можуть знаходити багато копій одного і того самого вмісту.

Коли ви вводите пошуковий запит, система зіставляє слова й фрази з індексом і шукає збіги. Але що робити, коли кількість збігів перевищує сотні мільйонів? На цьому етапі й виконується наступне важливе завдання – визначення рейтингу. Алгоритм, за яким це робиться, приховано. Це може відбуватись через аналіз розмітки сайту, кількість пов’язаних веб-сайтів і новизни вмісту. Мета - спробувати знайти для користувача саме те, що він шукає.

Зараз ми поговоримо про спеціалізовані пошукові системи наукової інформації.

Перша академічна пошукова система (в першу чергу в галузі інформатики та інформаційних технологій) - CiteSeer розпочала роботу у 1998 році. Сьогодні цей сервіс потребує вдосконалення.

Google Scholar - це спеціалізована система пошуку наукових документів. Завдяки масовому використанню веб-сканерів, вона отримує метадані з найбільш релевантних джерел у мережі. Окрім основної сторінки пошуку, вона включає: Google Scholar Citations (авторські профілі) та Google Scholar Metrics (журнали).

Microsoft Academic Search - це розробка Microsoft Research Asia, що фокусується не лише на профілях авторів, а й організацій та журналів. Цей двигун є одним з найбільш інноваційних рішень у сфері наукової інформації, наявної в Інтернет. Він повністю черпає свої дані з бази даних CrossRef і використовує автоматичний індекс цитування.

Університет Цінхуа (Tsinghua) в Китаї вперше представив ArnetMiner як концепцію особистого профілю науковця, інструмент візуалізації та можливість оцінки авторів та установ. Однак сервіс також містить кілька проблем та невідповідностей.

Серед інших пошукових систем можна згадати RefSeek - пошукова система, яка надає інструмент пошуку, спрямований на студентів та дослідників, що сканує документи, знайдені в Інтернеті.

BASE - розпочала роботу у 2004 році у Німеччині. Спеціалізується на відкритих джерелах і побудована на відкритих протоколах, що дозволяють збирати документи в з інституційних сховищ та цифрових бібліотек. Однак BASE наразі не пропонує повнотекстовий пошук.

WorldWideScience – є прикладом об'єднаної пошукової системи, що працює на основі бібліографічних баз даних наукових установ у всьому світі. Її метою є глобальний доступ до національних наукових джерел інформації.

●JURN – пошуковик, що працює на ядрі Google Custom Search і шукає серед репозитаріїв відкритого доступу та інституційних репозитаріїв. Заснований в 2009 році для пошуку наукових публікацій в галузі гуманітарних наук, в 2014 році відбулось оновлення та розширення сервісу, він починає індексувати матеріали з біології, медицини, економіки.

Пошукові стратегії на прикладі Google Scholar

Google Академія проводить пошук серед різних наукових дисциплін та джерел, включаючи рецензовані статті, дисертації, книги, анотації, статті академічних видань, професійних асоціацій, сховища препринтів, сайтів вищих навчальних закладів та освітніх організацій

Для проведення простого пошуку введіть пошуковий запит і натисніть кнопку пошуку..Можна скористатись Розширеним пошуком

Google Академія підтримує і більшість операторів розширеного пошуку: + додати в пошуковий запит загальні назви, букви або цифри, які, як правило, ігноруються пошуковою системою;

- відкинути всі результати, які містять заданий термін;

«...» лише результати, в яких міститься конкретній фразі;

OR результати, які міститимуть один з пошукових термінів;

intitle:.. пошук лише у заголовку документів.

This exact wording or phrase (“ ”)

One or more of these words (OR)

Any of these unwanted words (-)

File type (filetype:)

Search within a site or domain (site:)

Numeric range (..)

●Google Академія дозволяє досліднику укладати свою бібліотеку бібліографічних посилань. Проводячи пошук у Google Академії Ви завжди можете зберегти той, чи інший результат пошуку у свою бібліотеку, натиснувши на посилання «Зберегти».

●Для того, щоб перейти у свою бібліотеку бібліографічних посилань натисніть «Моя бібліотека»

●Переглядаючи бібліографічні посилання у своїй бібліотеці, можна процитувати потрібну публікацію – натисніть «Процитувати», скопіюйте та вставте відформатоване посилання, або скористайтеся одним із наведених посилань, щоб імпортувати її в бібліографічний менеджер.

Ось приклад цитування

Розглянемо декілька прикладів пошуку.

**Приклад 1.**

Припустимо, Вам потрібно дізнатись показник

щорічної смертності від СНІДу.

Для цього варто використовувати фразу ‘Each year, N people die of AIDS’

Як краще сформулювати цей запит, якщо нам не відомо значення N? Ми беремо фразу у лапки “” тому що шукаємо по точній фразі. Якщо не взяти у лапки, пошук видасть всі результати де зустрічаються слова EACH, YEAR, PEOPLE, DIE, OF, AIDS окремо.

Оператор зірочка \* (англ asterisk (\*)) – дає змогу гнучко пошукати заповнюючи собою порожні місця (“fill-in-the-blank” space in the search): “each year \* people die of AIDS”

**Приклад 2.**

Припустимо, Вам потрібно знайти матеріали про гробниці з раннього династичного періоду Єгипту за період 2686-3050 рр. до н.е. Проведіть пошук у Google Scholar так, аби до запиту egypt dynastic tombs додати пошук по всім рокам між 2686-3050 рр. до н.е.

Ps. egypt dynastic tombs 2686-3050 - неправильна відповідь, бо

такий пошук знайде лише схожі цифри. Оператор дві крапки .. Дозволяє знайти кожне значення числа між (включно з) числами, вказаними з права і ліва від двох крапок. (англ. the two dots make Google look for every number in between (and including) the numbers on either side of the dots) Тому правильна відповідь: Egypt Dynastic tombs 2686..3050 знайде всі числа, починаючи з 2686, 2687, 2688 і так далі аж до 3050.

**Приклад 3.**

Припустимо, Вам потрібно знайти матеріал у форматі Flash (формат файлів мультимедіа та векторної графіки), що стосується частин мікроскопу. Додайте до запиту

parts of a microscope пошук по типу файлу swf

Правильна відповідь: parts of a microscope filetype:swf Оператор filetype має бути написаний з маленької літери.

Висновки

➔пошукові системи постійно сканують Інтернет у пошуках вмісту, класифікують його й відображають найрелевантніші результати, щоб допомогти користувачам знайти саме те, що вони шукають.

➔важко дати загальне визначення того, що таке академічна пошукова система, оскільки, хоча всі вони містять наукову інформацію та є доступними в Інтернеті, їх функціональні можливості та алгоритми значно відрізняються один від одного.

➔деякі діють як бібліографічні бази даних (BASE) інші виступають спеціалізованими пошуковими системами (Google Scholar); тоді як деякі виробляють наукометричні показники (Microsoft Academic Search, AMiner). ➔цілком можливо, що нова покращена модель академічної пошукової системи з'явиться в майбутньому, конденсуючи та спрощуючи існуючі веб-інструменти.