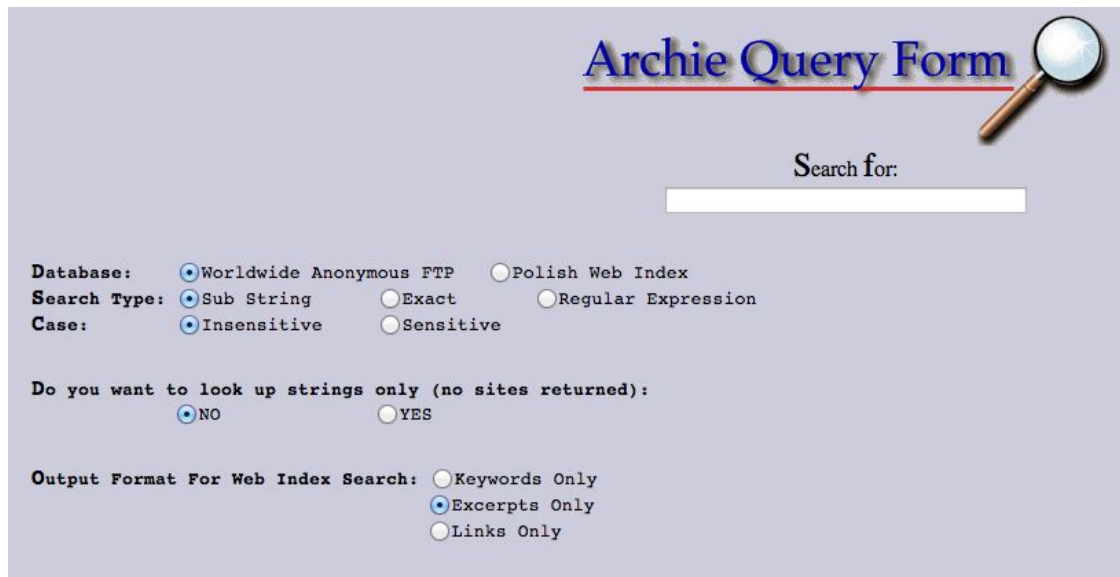


Пошукові системи

Олександра Ярошенко

Звідки беруть початок пошукові системи?

Однією з перших таких систем стала програма під назвою **Archie**, яка з'явилась у 1990 р. Її використовували для пошуку назв файлів, тобто, по суті, назв веб-сторінок. Однак програма Archie не інформувала про вміст веб-сторінок.



The image shows a screenshot of the Archie Query Form interface. The title "Archie Query Form" is displayed in blue text with a red underline, accompanied by a magnifying glass icon. Below the title is a "Search for:" label and a text input field. The form includes several sections with radio button options:

- Database:** ☒ Worldwide Anonymous FTP, ☐ Polish Web Index
- Search Type:** ☒ Sub String, ☐ Exact, ☐ Regular Expression
- Case:** ☒ Insensitive, ☐ Sensitive
- Do you want to look up strings only (no sites returned):** ☒ NO, ☐ YES
- Output Format For Web Index Search:** ☐ Keywords Only, ☒ Excerpts Only, ☐ Links Only

Звідки беруть початок пошукові системи?

Минуло кілька десятиліть, і такі пошукові системи, як Google, Bing, Yahoo, Ask.com, AOL, Baidu та Yandex досягли значних успіхів. Вони використовують неймовірно складні комп'ютерні програми для класифікації величезної кількості веб-сторінок.



Принцип роботи пошукових систем

Коли користувач хоче щось знайти, він вводить слово або фразу, тобто пошуковий запит. Далі пошукова система порівнює цей запит зі своїм каталогом веб-сторінок і показує користувачеві найвідповідніші варіанти на сторінці результатів пошуку.

Принципи роботи більшості пошукових систем однакові.



how to

how to **tie a tie**

how to **screenshot on mac**

how to **get away with a murderer**

how to **write a check**

how to **hard boil eggs**

how to **make money**

how to **boil eggs**

how to **screenshot on pc**

how to **draw**

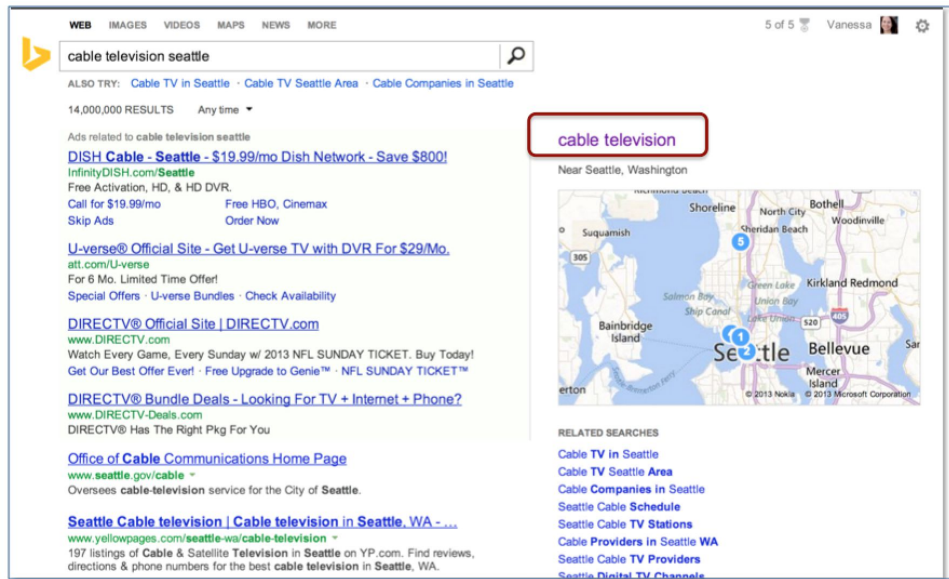
how to **write a cover letter**

Google Search

I'm Feeling Lucky

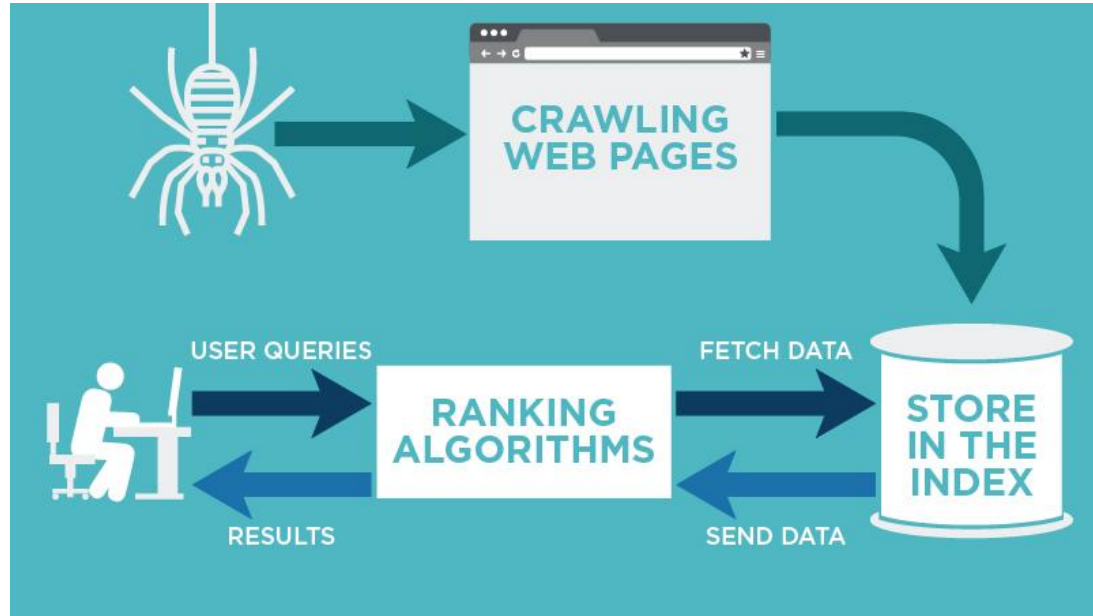
Мета пошукової системи

Мета пошукової системи – створити список найбільш релевантних результатів, щоб допомогти користувачам знайти потрібний вміст. На сторінці результатів пошуку відображаються переважно посилання на сайти, але також і зображення, рекламні оголошення, карти, відео тощо.



Як працює пошукова система

Базові завдання, які вони виконують: вивчення виявленого вмісту, до якого є доступ (це називається "**сканування**"), класифікація кожного фрагмента вмісту (**індексування**), а також вибір вмісту, найбільш корисного для автора пошукового запиту (визначення **рейтингу**).



У кожную пошукову систему закладена своя програма, проте всі вони діють подібно.



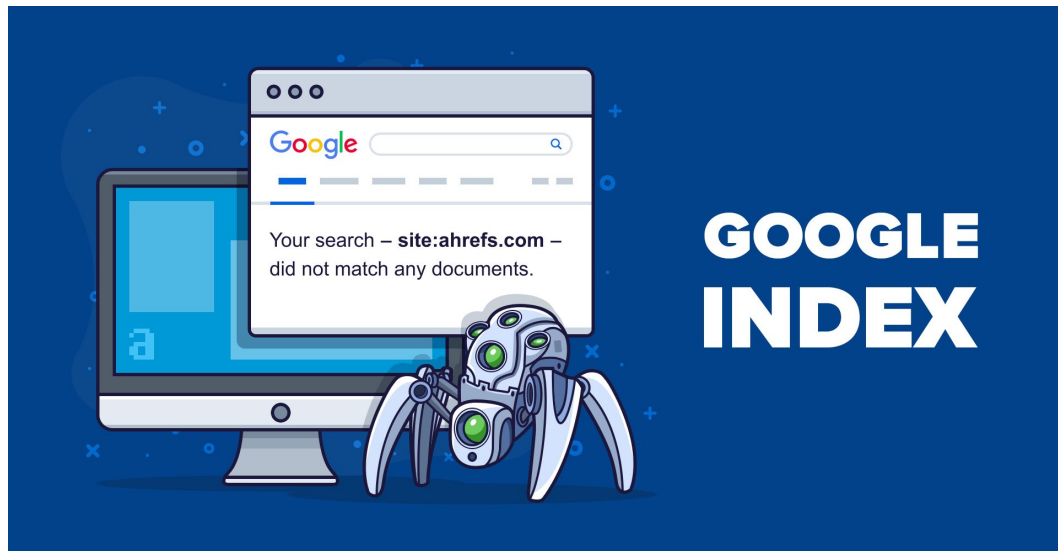
Розгляньмо докладніше, як пошукові системи сканують Інтернет, виявляючи такий вміст, як веб-сторінки, зображення та відео.

Веб-сканери

Кожна система використовує комп'ютерні програми для **аналізу веб-сторінок**. Такі програми називають **веб-сканерами, павуками або роботами**. Роботи переходять від сторінки до сторінки за посиланнями. Вони ніколи не зупиняються. Їх призначення – перегляд веб-сторінок для виявлення нових посилань або вмісту для включення в індекс.

Індексування

Індексування – друга складова процесу. Це величезний список усіх сторінок і вмісту, виявленого роботами. Пошукова система використовує індекс як джерело інформації, відображуваної у результатах.



Але не вся інформація, виявлена роботом, додається до індексу. Пошукові системи можуть знаходити багато копій одного вмісту, розміщеного на різних сайтах.

Рейтинг

Що відбувається, коли ви вводите пошуковий термін? 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**. Метод, за яким це робиться, приховано. Без перебільшення, існують сотні способів визначення рейтингу, зокрема аналіз лексики, використаної на сторінці, кількості пов'язаних веб-сайтів і новизни вмісту. Але якою б не була використовувана формула, мета залишається тією ж самою – спробувати знайти для користувача саме те, що він шукає.

CiteSeer^x

Перша академічна пошукова система (в першу чергу в галузі інформатики та інформаційних технологій) на основі власного процесу сканування та незалежного індексу цитування. **CiteSeer** розпочала роботу у 1998 році і мала багато нових можливостей: Автономна індексація цитування дозволила автоматично створювати **індекс цитування**, який можна використовувати для пошуку та оцінки літератури.

<https://citeseerx.ist.psu.edu/index>

Сьогодні цей сервіс потребує розвитку як з точки зору функціонування, так і покриття.

Google Scholar

Google Scholar - це спеціалізована система пошуку наукових документів. Завдяки масовому використанню веб-сканерів, вона отримує метадані з найбільш релевантних джерел у мережі. Вона використовує технологію **PageRank** для сортування документів та власний алгоритм індексування для вимірювання впливу кожного матеріалу. Окрім основної сторінки пошуку, вона включає: **Google Scholar Citations** (авторські профілі) та **Google Scholar Metrics** (журнали).



Слабкі сторони - це не надто прозорі результати, не надто великий вибір опцій для звуження результатів пошуку

<https://scholar.google.com.ua/>

Microsoft Academic Search

Microsoft Academic Search - це розробка Microsoft Research Asia, що фокусується не лише на профілях авторів, а й організацій та журналів. Окрім оригінальних візуалізацій та **інструментів бенчмаркінгу**, 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, класифікований за алгоритмом **PopRank**.

<https://academic.microsoft.com/home>



Основними його
недоліками є
недосконалий механізм
пошуку, орієнтований лише
на документи та авторів та
повільна швидкість
оновлення.

AMiner

Університет Цінхуа (Tsinghua) в Китаї вперше представив **AMiner** (або **ArnetMiner**) як концепцію особистого профілю науковця, інструмент візуалізації та можливість оцінки авторів та установ. Однак сервіс також містить кілька аномалій та невідповідностей, таких як дефіцитний механізмів пошуку, відсутність прозорості джерел та хаотичний дизайн.



Ці досить серйозні проблеми викликають питання довіри користувача до цього інструменту як надійного джерела не лише для оцінки досліджень, але й для пошуку академічної інформації.

BASE

BASE (Bielefeld Academic Search Engine) - пошукова система, що розпочала роботу у 2004 році (Німеччина). Спеціалізується на відкритих джерелах і побудована на відкритих протоколах (OAI-PMH), що дозволяють збирати документи в основному з інституційних сховищ та цифрових бібліотек. Користувачі можуть шукати бібліографічні метадані, включаючи анотації, якщо вони доступні. Однак BASE наразі не пропонує повнотекстовий пошук.



<https://www.base-search.net>

WorldWideScience

WorldWideScience - це приклад об'єднаної пошукової системи, що працює на основі бібліографічних баз даних наукових установ у всьому світі.

Мета: глобальний доступ до національних наукових джерел інформації.

<https://worldwidescience.org>

Однак ця модель дає велику кількість дубльованих результатів і вимагає великих витрат часу.



Оператори пошуку (на прикладі Google Scholar)

Полегшити роботу та звузити результати пошуку допоможуть **оператори пошуку**:

Оператор “...” - пошук по точній фразі

Наприклад, щоб
знайти рядок з
вірша

1 result (0.18 seconds)

► [Poetry Corner « POWO.co.ke](http://www.powo.co.ke)
www.powo.co.ke/media-2/
Like black tape he is stuck to me. Cant dare resist me. Halts in front of me.
Mesmerized by me. Awaits approval from me. Time to beseech me. Relish his desires
...

Оператори пошуку (на прикладі Google Scholar)

Оператор ~ - щоб знайти
СИНОНІМИ

Наприклад, *vocabulary, glossary,
definitions*

chemistry ~vocabulary

About 17,000,000 results (0.09 seconds)

► [Chemistry Vocabulary - Common chemistry terms explained](#)
[misterguch.brinkster.net/vocabulary.html](#)
80+ items – **Chemistry vocabulary** terms defined for the enlightenment of ...
absolute temperature This is a temperature reading made relative to ...
absolute zero This is the lowest temperature possible. If you ...

[Chemistry Glossary - Chemistry Glossary Terms Starting with the ...](#)
[chemistry.about.com/od/chemistryglossary/a/glossarya.htm](#)
Use this **chemistry glossary** to look up the definitions of **chemistry terms** beginning with the letter A.

[Online Chemistry Dictionary / Glossary! everyone's online source fo...](#)
[www.chemistry-dictionary.com/](#)
Free online **chemistry dictionary** containing over 1800 **chemistry terms** and definitions.
Browsing - Search - Chemistry Terms starting with A - Random

[Chemistry Glossary](#)
[www.shodor.org/unchem/glossary.html](#)
A basic **Chemistry glossary** suitable for incoming freshman chemistry students.

[Chemistry Definitions A](#)
[home.nas.net/~dbc/cic_hamilton/dictionary/a.html](#)
50+ items – **Chemistry Dictionary**. Terminology "A". A · B · C · D · E · F ...
Absolute Entropy (of a substance) The increase in the entropy of a ...

Оператори пошуку (на прикладі Google Scholar)

Оператор * - заповнює собою
порожні місця (*"fill-in-the-blank"*
space in the search)

"each year * people die of aids"

About 249,000 results (0.32 seconds)

[Increased funding fails to curb AIDS / \\$5 billion not enough to halt ...
articles.sfgate.com/.../17437131_1_hiv-last-year-aids-experts-unaid](#)

Jul 7, 2004 – **Each year, 3 million people die of AIDS**, including 500000 children.
advertisement | your ad here. Since it surfaced in California more than two ...

[Words of Death: It's a Girl](#)

[liveaction.org/blog/words-of-death-its-a-girl/](#)

Jan 21, 2012 – In comparison: **each year, some 2.8 million people die of AIDS**, 1.27
million of malaria. Or, put in the most horrible terms: violence against ...

[Subjects for Prayer](#)

[www.stpancraschurch.org/index.php?id=106](#)

Currently **each year 3 million people die of AIDS**, nearly 2 million of tuberculosis, and
1 million of malaria. 7. Ensure environmental sustainability. Currently 1.1 ...

[HIV-outreach Programme in Ghana](#)

[www.ibiswestafrica.com/index.php?option=com...view...](#)

... died from AIDS related diseases. **Each year about 18000 people die of AIDS** in
Ghana, a figure that still decreases partly due to access to new AIDS drugs.

[HIV/AIDS for Florida Healthcare Professionals \(4 CH\) | Continuing ...](#)

[www.nursingceu.com/courses/342/index_mand.html](#)

by N Evans - [Related articles](#)

... with HIV, approximately 1 new infection every 9.5 minutes (Hall et al, 2008). And
each year more than 18000 people die of AIDS in the U.S. (CDC, 2010a).

Оператори пошуку (на прикладі *Google Scholar*)

Оператор **OR** знайде хоча б один з існуючих термінів, або всі, якщо існують всі

Також, оператор **+** додає ще один термін чи фразу, а оператор **-** виключають з результатів пошуку будь-які згадки про термін чи фразу

political cartoons analysis burlesque OR parody OR satire OR irony



About 5,820,000 results (0.19 seconds)

[A Semiotic Analysis of Political Cartoons](#)

[cseweb.ucsd.edu/~clbailey/271/Cartoons.htm](#)

This paper presents a formal **analysis** of **political cartoons** using methods ... is the essence of caricature, or **satire**, which is the basis for **political cartoons'** effect. ...

[How to Analyze Political Cartoons: 6 steps - wikiHow](#)

[www.wikihow.com > ... > Writing > Works > Comics](#)

They can also be **satirical** or even serious in tone, depending on the audience, the artist, and the idea illustrated. Use these steps to **analyze** a **political cartoon** ...

[Political Analysis Through Satire - PBS](#)

[www.pbs.org/now/classroom/satire.html](#)

Jul 11, 2003 – Identify, **analyze**, and explain examples of **political satire**. ... Luther King, Jr.'s "I Have a Dream," campaign advertisements, **political cartoons**) ...

[Political Satire Cartoon Analysis/Discussion - MrTESOL](#)

[www.englishcafe.com/.../political-satire-cartoon-analysisdiscussion-1...](#)

Feb 24, 2009 – Take a good look at this **cartoon** shown in the UK press recently Funny isn't it, but can you see the reasons behind the artists (somewhat ...

[Cartoon Analysis Guide - Library of Congress](#)

[www.loc.gov > ... > Classroom Materials > Presentations and Activities](#)

Cartoon Analysis Guide for the Exploring **Political** Cartoons from the civil rights era ... When you look at a cartoon, see if you can find any **irony** in the situation the ...

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.