## Лекція 1. Пошукові системи – основні поняття та принципи роботи

### План

1. Основні поняття
2. Принципи роботи

### Основні поняття

***Пошукова система*** (пошуковик, ПС) – програмно-апаратний комплекс з веб інтерфейсом. Здійснює функції пошуку по мережі Інтернет.

#### Частини пошукової системи

***WEB Spider*** (павук) – програма; працює як інтернет оглядач. Служить для отримання інформації зі сторінок інтернет-ресурсів.

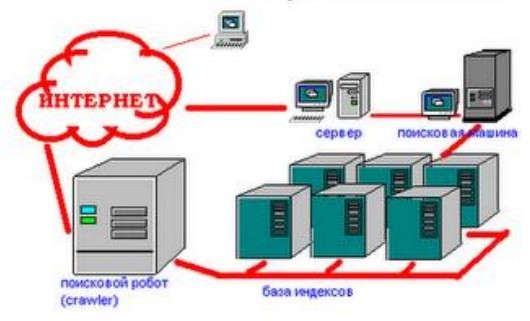
***Crawler*** (краулер) – програма; обробляє посилання, які пов'язують сторінки і вказує шлях павуку.

***Indexer*** (індексер) – програма; займається аналізом інформації на сторінках, текстового і графічного наповнення.

***Database*** (база даних) – база даних, в якій зберігатися індекс пошукової системи і збережена інформація з проіндексованих веб сторінок.

***Search results engine*** (пошукова машина) – програмна частина, яка здійснює оцінку і сортування збереженої інформації.

***Web server*** (веб сервер) – сервер, який взаємодіє між користувачем і пошукової системою.



Сервер Пошукова машина

**ІНТЕРНЕТ**

Пошуковий робот

(crawler)

База індексів

**Рис 3. Принцип роботи пошукової системи.**

***Пошуковий індекс*** – це рубрикатор, який за своєю суттю схожий зі змістом книги. Технологія індексу використовується в пошукових машинах для більш швидкого доступу і оцінки збереженої інформації. Пошуковий індекс включає в себе елементи математики, лінгвістики, когнетивної психології і інформатики. В індексі зберігається інформація про релевантність та тематику сторінки, результати семантичного аналізу, інформація про посилання сторінки і багато іншого.

### Принципи роботи

Наведемо основні поняття.

***Ранжування сайтів*** – сортування результатів пошукової системи на певний запит. Компанія Google враховує більш 200 характеристик для оцінки відповідності сторінки сайту пошуковому запиту. З урахування всіх характеристик ПС формує релевантність сторінки.

***Релевантність*** – це показник відповідності сторінки як інформаційної відповіді щодо запиту пошукової системи. У релевантності враховується не тільки лінгвістична відповідність але і можливість застосування,

«адекватність», даного результату.

***Траст сайту*** (довіра ПС) – одна з характеристик сайту – це значення

«довіри» пошукової системи щодо цього сайту. Непрямими показниками трасту сайту є числові показники якості сайту або сторінки. Також на траст сайту впливаютьтакі показники як:

* Домен сайту
* Вік домена
* Відвідуваність сайту
* Показники ТІЦ та PR.

***PageRank*** (Пейдж ранк) – названий на честь його винахідника Ларі Пейджа. Змінюється від 0 до 10, також може мати значення «не визначений». По суті, є алгоритмом ранжування посилань, тобто показує, на скільки певна сторінка важливіше щодо інших сторінок враховуючи контрольну вагу цих сторінок. Також пейдж ранк можна розглядати як показник ймовірності перебування користувача на конкретній сторінці, враховуючи тільки зв'язки через посилання. Пейдж ранк має експоненційну природу, це означає, що значення PR 10 в десятки тисяч разів більше ніж значення PR 1.

***Тематичний індекс цитування*** (ТІЦ) – технологія пошукової машини Яндекс, яка полягає в визначенні авторитетності інтернет-ресурсів з урахуванням якісної характеристики – посилань на них з інших сайтів. ТІЦ

розраховується за спеціально розробленим алгоритмом, в якому особливе значення приділяється тематичній близькості ресурсу і посилання на нього. ТІЦ визначається сумарною вагою сайтів, які не можуть впливати на ТІЦ сайту, за виключенням тих сайтів, де будь-яка людина може поставити своє посилання без відома адміністратора ресурсу.

#### Просування сайтів з точки зору пошукових систем це:

* Аналіз змісту або структури сайту.

* Технічні рекомендації по розробці веб-сайту, наприклад: хостинг, переадресація, сторінки з повідомленнями про помилки, використання JavaScript.

* Розробка змісту.

* Керування кампаніями з розвитку бізнесу в Інтернеті. * Аналіз ключових слів.

* Навчання в галузі пошукової оптимізації.

* Досвід роботи на певних ринках і географічних територіях.

Якщо для сайту обрано SEO-просування, бажано спочатку зрозуміти принцип роботи пошукових систем (ПС) для отримання позитивних результатів. Це допоможе побудувати власну стратегію оптимізації Вашого ресурсу під потрібні пошукові запити так, щоб напевно опинитися на першій сторінці видачі пошуковика.

#### Основні завдання ПС

1. Дати користувачу правильну відповідь на введений в рядок пошуку запит і тим самим задовольнити його, щоб не дати приводу скористатися конкуруючої ПС. Для цього в ТОП 10 пошуку повинні бути результати, які максимально підходять під запит користувача.
2. Забезпечити монетизацію трафіку. Тобто, в першу чергу, видати результат, який сприятиме отриманню прибутку власником пошукової системи, наприклад, за допомогою контекстної реклами.

#### Як працює пошукова система?

Пошукових систем на сьогоднішній день досить багато. Природно, що в кожній з них пошук побудований на власних алгоритмах, але в загальному принцип роботи пошукових машин досить схожий.

У той момент, коли Ви створили сайт і бот однієї з пошукових систем дізнався про це (в результаті закупівлі посилань або будь-яким іншим способом), він «приходить і починає бродити» по сторінках.

***Принцип роботи пошукових*** роботів базується на виконанні декількох етапів:

***Збір даних***. За допомогою ботів ПС збирає необхідну інформацію з усіх сторінок ресурсу (тексти, зображення, відео та ін.).

Ці процеси починається після запуску сайту і надалі відбуваються за складеним роботом розкладом. Завдяки цьому час від часу відбувається оновлення даних про сторінки на предмет змін.

***Індексація*** – привласнення кожній сторінці певного індексу, який дозволить проводити швидкий пошук по їх вмісту.

З цього моменту Ваш ресурс або деякі його сторінки можуть бути відображені на сторінках видачі пошуковика по окремим запитам. А те, на якій сторінці він виявиться, залежить від рівня його оптимізації.

***Основні принципи ранжирування сайтів в пошукових системах.*** Для того, щоб сайт потрапив у ТОП 3, або хоча б у ТОП 10, необхідно знати всі основні критерії пошукових систем і намагатися оптимізувати його таким чином, щоб він їм відповідав. Ці критерії можна розділити на дві основні групи

– текстові і нетекстові.

***Текстові критерії ПС***. В цьому випадку ранжування сайту грунтується на оцінці якості текстового вмісту сторінок. Робота з ними проводиться на етапі створення семантичного ядра сайту і наповнення його текстовим контентом.

Мається на увазі, що ПС видає на першій сторінці пошуку результати, найбільш відповідні запиту користувача. Тобто їх вміст найбільш точно відповідає необхідним критеріям і має задовольнити його. Також при цьому звертається увага на правильність заповнення титульної фрази (title), опису сторінки (description) і наявність ключового запиту в заголовках (H1, H2 і т. Д.). ***Нетекстові критерії ПС.*** Робота з ними відбувається після публікації сайту і його індексації. Принципи визначення рейтингу Інтернет-ресурсів на основі нетекстових критеріїв базуються не на оцінці якості їх змісту. Вони

містять дещо інші чинники ранжування сайту.

Це каталожні та посилально-розрахункові критерії. Сюди відносяться індекс цитування PageRank (PR) від Google і CY від Яндекс. Система працює за принципом обліку кількості посилань на Ваш ресурс, оцінка авторитетності ресурсів, які посилаються на сайт, перевірка реєстрацій в каталогах сайтів.

Можна провести аналогію з висновками про якість певного товару на основі відгуків споживачів, постачальників, продавців тощо.

*Додаткові матеріали:*

Визначення SEO в довідковому центрі Google <https://goo.gl/h1TGt>

Заборонені методи просування у вказівках компанії Google для веб-майстрів <https://goo.gl/9tuq>

Детально про PageRank <https://goo.gl/Ywe1i>

Відео «Принцип роботи пошукових систем» <https://goo.gl/4RF1EB> Корисний сайт <https://seo-akademiya.com/>