

## 6. ПРИНЦИПИ ФУНКЦІОНУВАННЯ ПОШУКОВИХ СИСТЕМ

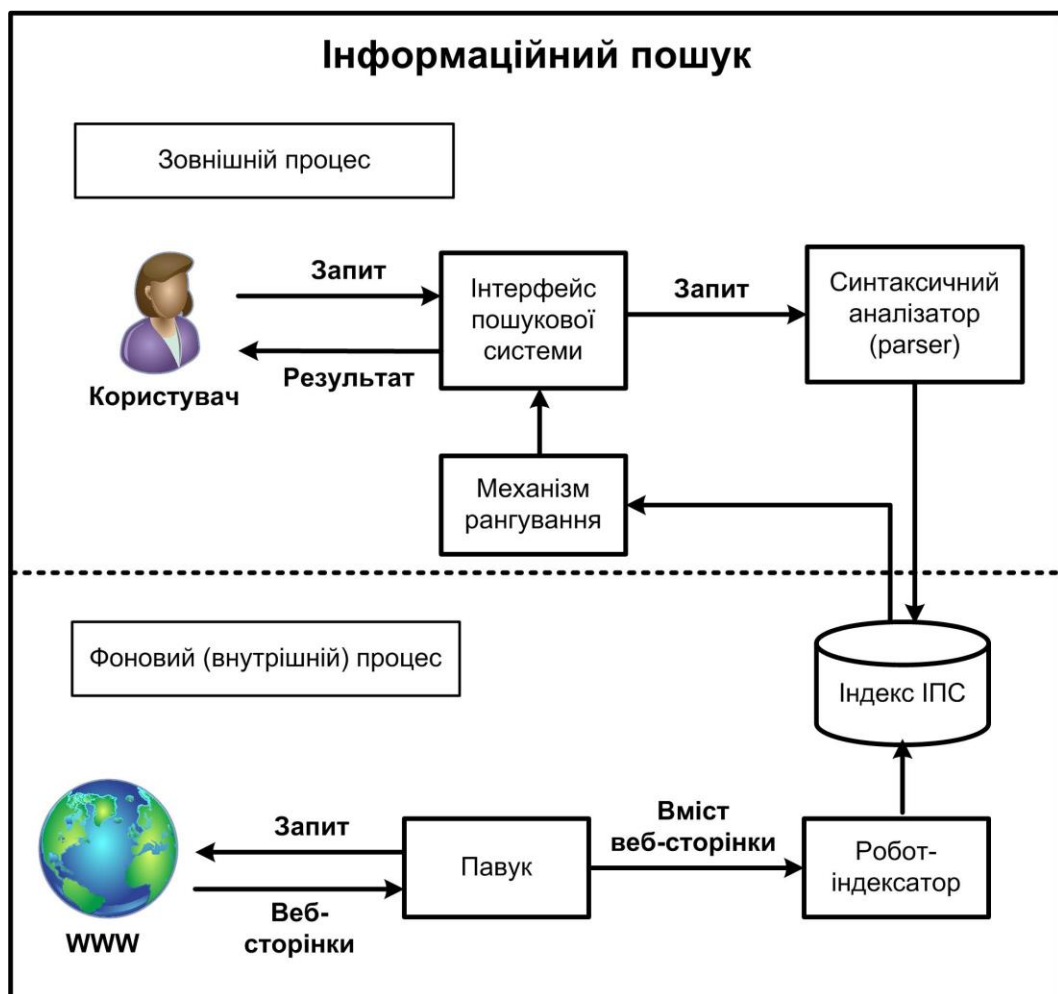


Рис. 2. Складові процесу інформаційного пошуку у веб

Інформаційний пошук у веб умовно можна поділити на 2 процеси:

- 1) фоновий або внутрішній процес
- 2) зовнішній процес.

Фоновий процес – це, фактично, сукупність процесів індексування інформаційних ресурсів та оновлення індексу пошукової системи. Фоновий процес відбувається непомітно для користувачів та без їхньої участі. Він відбувається “на фоні” роботи пошукової системи безпосередньо з користувачем. Однак, без фонових процесів, основний процес, тобто робота з користувачем є неможливою. Це пояснюється тим, що пошукова система в момент надходження запиту від користувача, здійснює пошук лише в межах попередньо сформованого індексу. Якщо індексу не існує, то пошук

відбуватися не може. Окрім цього, індекс пошукової системи повинен періодично оновлюватися для того, щоб результати пошуку були актуальними. Тому фоновий процес відбувається безперервно. Павуки обходять за посиланнями задану множину ресурсів, завантажують вміст сторінок, витягують посилання на нові сторінки з отриманих ресурсів. Вміст сторінок передається до робота-індексатора, який аналізує сторінки і вносить отриману інформацію до індексу ПС.

Зовнішній процес – це сукупність процесу опрацювання запиту користувача і видавання результатів.

Через інтерфейс пошукова система отримує від користувача запит. Запит передається до синтаксичного аналізатора і піддається морфологічному аналізу. Генерується інформаційне оточення кожного документа, що міститься в базі, яке буде згодом відображене у вигляді сніпета. Сніпет (англ. snippet – уривок, фрагмент) – невеликий фрагмент тексту сайту, який виводиться в результатах пошуку, що дозволяє оцінити зміст веб-сторінки без необхідності переходу за посиланням, і в якому виділяються ключові слова, що вводяться користувачем в пошукову форму.

Дані, отримані з індексу, передаються в якості вхідних параметрів до спеціального модуля рангування. В ньому відбувається обробка даних зі всіх документів, в результаті чого, для кожного документа розраховується власний рейтинг, що характеризує його релевантність запиту, введеного користувачем, і різних складових цього документа, що зберігаються в індексі пошукової системи .

Залежно від вибору користувача цей рейтинг може бути скоректований додатковими умовами (наприклад , параметрами розширеного пошуку). Далі генерується сніпет, тобто, для кожного знайденого документа з таблиці документів витягуються заголовки, короткий текстовий фрагмент-анотація, в якому виділені ключові слова, а також посилання на сам документ.

Отримані результати пошуку передаються користувачеві у вигляді SERP (Search Engine Result Page ) – сторінки видачі пошукових результатів.