

4. СПЕЦИФІКА ПОШУКУ В ІНТЕРНЕТІ

На перший погляд пошук в Інтернеті мало чим відрізняється від звичайного інформаційного пошуку, наприклад, від обробки SQL запиту до бази даних або від завдання пошуку файлу на комп'ютері. Однак, існує низка суттєвих відмінностей.

Перша відмінність пошуку в мережі Інтернет від звичайного полягає в тому, що алгоритм пошуку по базі даних, наприклад, передбачає, що її структура заздалегідь відома пошуковій машині і авторові запиту. В Інтернет сторінки утворюють мережу, а не каталожну структуру, що також впливає на алгоритми пошуку, а формат даних, що розміщуються на інтернет ресурсах, ніким не контролюється.

Друга відмінність є одним із наслідків першої: запит при пошуку в мережі подається не у вигляді набору значень параметрів (критеріїв пошуку), а у вигляді тексту, що написаний людиною на природній мові. Таким чином, перед тим, як почати пошук, пошукова система трансформує запит у зрозумілу для себе форму.

Третя відмінність є менш очевидною, але не менш принциповою: в каталозі або базі даних усі елементи є рівноправні. В інтернеті існує конкуренція, а також недобросовісні ресурси, тому, окрім звичайного пошуку, потрібно ще застосовувати фільтри для інформаційного сміття, а також методи рангування результатів видачі.

Четверта відмінність між роботою пошукової машини в інтернеті і роботою будь-якої іншої інформаційної системи, що здійснює пошук у різних каталогах і базах даних, полягає у тому, що пошукова інтернет машина не шукає інформацію серед тих ресурсів, що є в інтернеті на момент надходження запиту, а формує відповідь на підставі власного інформаційного сховища, що називається індексом, де зберігаються “досьє” на всі відомі їй ресурси, і ці досьє система періодично оновлює. Іншими словами, пошукова машина працює не з оригіналом, а з проекцією області допустимих значень пошуку. Всі останні зміни в інтернеті можуть відобразитися в результатах пошуку тільки після того,

як відповідні сторінки будуть проіндексовані, вперше або повторно – додані в індекс пошукової системи.