

УДК 004.891

Василик К., Притолук Г. – ст. гр. СНм-51

Тернопільський національний технічний університет імені Івана Пулюя

ПОШУКОВІ СИСТЕМИ

Науковий керівник: ас. Маєвський О.В.

Vasylyk K., Prytol'uk G.

Ternopil Ivan Pul'uj National Technical University

SEARCH ENGINES

Supervisor: Majevskiy A.

Ключові слова: пошукові системи, програма

Keywords: search engines, program

Одним з перших способів організації доступу до інформаційних ресурсів мережі стало створення каталогів сайтів, в яких посилання на ресурси групувалися згідно з тематикою. Першим таким проектом став сайт Yahoo. Після того, як число сайтів в каталозі Yahoo значно збільшилося, була додана можливість пошуку інформації за каталогом. Першою повноцінною пошуковою системою став проект WebCrawler.

В даний час існує 3 основних міжнародних пошукових системи – Google, Yahoo і Bing Search, що мають власні бази і алгоритми пошуку. Більшість інших пошукових систем використовує в тому чи іншому вигляді результати вищеперерахованих. Наприклад, пошук AOL і Mail.ru використовують базу Google, а AltaVista, Lycos і AllTheWeb – базу Yahoo.

Типова пошукова система складається з наступних основних компонентів:

Spider (павук) – це програма, яка завантажує веб-сторінки тим же способом, що і браузер користувача. Відмінність полягає в тому, що браузер відображає інформацію, що міститься на сторінці, павук же не має ніяких візуальних компонент і працює безпосередньо з HTML-текстом сторінки.

Crawler (краулер, «мандрівний» павук) – програма, яка автоматично проходить по всіх посиланнях, знайденим на сторінці. Виділяє всі посилання, присутні на сторінці.

Indexer (індексатор) – програма, яка аналізує веб-сторінки, завантажені павуками. Індексатор розбирає сторінку на складові частини і аналізує їх. Виділяються і аналізуються різні елементи сторінки.

Database (база даних) – сховище завантажених і оброблених сторінок, які пошукова система завантажує і аналізує. Інша назва – індекс пошукової системи.

Search engine results engine (система видачі результатів) – видобуває результати пошуку з бази даних. Система видачі результатів займається ранжуванням сторінок. Вона вирішує, які сторінки задовольняють запиту користувача, і в якому порядку вони повинні бути відсортовані. Це відбувається згідно алгоритмам ранжування пошукової системи. Ця інформація є найбільш цінною і цікавою.

Детальна реалізація пошукових механізмів може відрізнитися один від одного (наприклад, зв'язка Spider + Crawler + Indexer може бути виконана у вигляді єдиної програми, яка завантажує відомі веб-сторінки, аналізує їх і шукає по посиланнях нові ресурси), проте всім пошуковим системам властиві описані загальні риси.