

УДК 004.4'6, 004.451.83

Торяник А., Неділенько В. – ст. гр. СНсп-43

Тернопільський національний технічний університет імені Івана Пулюя

РОБОТА ПОШУКОВИХ СИСТЕМ

Науковий керівник: асистент Маєвський О.В.

Пошукові системи або машини (Search Engines) безперервно розвиваються і вдосконалюються. Проте існують деякі незмінні принципи їх роботи.

Всі пошукові системи об'єднують те, що вони розташовані на спеціально виділених потужних серверах і володіють високопродуктивними каналами зв'язку. Кількість одночасно обслуговуваних відвідувачів досягає на найбільш популярних пошукових машинах багатьох тисяч, а найбільш відомі системи обслуговують в добу мільйони клієнтів. Відбір і підготовка інформації для видачі її клієнтам-користувачам у різних пошукових системах здійснюється по-різному. Відповідно, і назви таких систем різні.

Першим типом пошукових систем-машин є наочні або тематичні каталоги (Subject Catalogs, Subject Guides). Компанія, яка володіє каталогом, проводить безперервну роботу по вивченню і впорядкуванню вмісту матеріалів різних www-серверів, і в результаті створюється деякий ієрархічний каталог. Жодного наукового принципу ділення знань в такому каталозі немає. Каталог створюється на підставі уявлень про звичайні вимоги основної маси користувачів. Тому немає жодних гарантій того, що каталог охоплює всі розділи знань. В той же час саме свідомість відбору, в який втручаються модератори, роблять такі каталоги дуже зручними і корисними.

Другий тип пошукових систем — це чисто пошукові машини. Вони зв'язані з використанням самостійного автоматичного збору і машинного аналізу інформації, що знаходиться в різних частинах Інтернету, для чого створюються спеціальні допоміжні програми з різними назвами. Такі програми періодично досліджують вміст всіх ресурсів Інтернету, переміщаючись по різних ресурсах. Відповідно вони називаються роботами або ботами. Крім того, такі програми називаються ще спайдерами та краулерами. Такі програми починають досліджувати і «викачувати» з різних URL-адрес інформацію, що міститься в них, але не можуть оцінити реальний вміст ресурсів і документів. Вони будують спеціальні словопоказники, які називаються індексами. Ці індекси і лежать в основі другого типу пошукових машин, часто названих просто автоматичними індексами. Всі ці програми відвідують кожний ресурс, через певний час, інколи тільки раз в місяць. Частота відвідин залежить від багатьох обставин - зокрема, від загальної відвідуваності ресурсу. Програми-спайдери в різних пошукових системах різні, тому і індексація в різних пошукових машинах теж різна.

Слід зазначити, що пошуковим машинам велика частина ресурсів або недоступна, або ж вони її в своїй роботі просто ігнорують. В даний час пошукові системи і каталоги в чистій формі майже не зустрічаються. Велика частина пошукових систем в тій або іншій мірі поєднують обидві форми роботи.

Серед пошукових систем особливе місце займають системи, в яких активна роль належить експертам-модераторам. Ці експерти відбирають матеріал для каталогів, індексують його і часто пишуть професійні резюме вмісту тих електронних ресурсів, які є в каталозі системи. Робота таких пошукових систем дуже дорога, і з цієї причини кількість відбитих в них ресурсів набагато менша, ніж в основних пошукових системах, але цей недолік окупується високою якістю обробки матеріалів.