О.М. Дишлюк

*Національний авіаційний університет*

ПОШУКОВІ СИСТЕМИ МЕРЕЖІ INTERNET:
ТЕХНОЛОГІЇ ТА ЗАСОБИ

Шукаючи інформацію в мережі *Internet*, ми іноді не замислюємося над тим, як саме здійснюється пошук потрібної нам інформації: чому саме в такому порядку, а не в іншому перелічені посилання на *Web*-сторінки, як проходить вибір корисних повідомлень із існуючих в мережі, і взагалі, що відбувається «за лаштунками» пошукових систем.

На сьогодні найбільш популярними серед російсько-українських пошукових систем в мережі *Internet* є *Google, Yandex, Meta, Bigmir, Rambler* та інші. Для забезпечення їх функціональності використовується комплекс програм, що складається із трьох базових компонентів: веб-паук, індексатор, пошукова машина.

Веб-паук представляє собою програму, яка обходить сторінки із заданою *URL*-адресою, скачує їх в базу даних, потім архівує та перекладає її в сховище. Індексатор виконує обробку сторінок: «читає» всі документи та одержує з них слова, розміщуючи їх в спеціальну базу даних – індекс. Спеціальна програма складає таблиці перенумерації документів бази та об’єднує вміст всіх частин. Серед сторінок з однаковими адресами вибираються найбільш свіжа версія. Коректність та точність обробки даних забезпечує пошукова машина.

Основними критеріями якості роботи пошукової машини є релевантність, повнота бази та врахування морфології слів. Основним методом для оцінки релевантності є *TF-IDF*-метод, а якість посилань контролюється алгоритмом ранжування.

На результат пошуку впливають: загальний об’єм бази даних, механізм індексації, здатність пошукової системи розуміти запити, ранжування результатів пошуку.