

1. ВСТУП ДО ІНФОРМАЦІЙНОГО ПОШУКУ

1.1. Визначення поняття “інформаційний пошук”, його завдання та цілі

Пошук (search) – це сукупність операцій, які пов'язані з визначенням місцезнаходження об'єктів з заданими характеристиками або ознаками.

Інформаційний пошук або пошук інформації (information retrieval, data search) – це пошук неструктурованої інформації, одиницею представлення якої є інформація у довільних форматах. Завдання інформаційного пошуку стосується пошуку інформації в документах, пошук самих документів, вилучення метаданих з документів, пошуку тексту, зображень, відео і звуку у локальних реляційних і гіпертекстових базах даних.

Термін «інформаційний пошук» був вперше введений Кельвіном Муерзом (Calvin Mooers) у виступі на конференції у 1950 році.

Залежно від ступеня залучення до інформаційного пошуку технічних засобів і участі в ньому людини розрізняють: "ручний", "машинний" і "автоматизований" інформаційний пошук.

В автоматизованих інформаційних системах інформаційний пошук забезпечується і здійснюється із залученням лінгвістичних, інформаційних, програмно-технічних, технологічних, організаційних засобів і складених з них комплексів .

Безпосередньо інформаційний пошук здійснюється засобами інформаційно-пошукової системи, що є підсистемою АІС. Основними критеріями якості результатів інформаційного пошуку є повнота, точність і оперативність пошуку.

Як критерій інформаційного пошуку, так і його результати є недетерміновані, тобто невизначеними. Цими признаками інформаційний пошук відрізняється від ‘пошуку даних’.

Пошук або видобування даних (data mining) – процес аналітичного дослідження великих масивів інформації з метою виявлення певних

закономірностей і систематичних взаємозв'язків між змінними, які потім можна застосувати до нових сукупностей даних.

Пошук даних відбувається у масивах з структурованою інформацією і його результат завжди детермінований.

Інформаційний пошук на практиці це послідовність операцій, спрямованих на надання інформації зацікавленим особам, процес відшукування в деякій множині текстів (документів) всіх тих, які присвячені темі, зазначеній в запиті, або містять потрібні споживачеві факти чи відомості.

Предметом пошуку виступає інформаційна потреба користувача, виражена у формі інформаційного запиту.

Потреба людей в інформації постійно змінюється і трансформується, тому її неможливо виразити і описати однозначно. Проте інформаційну потребу можна представити у вигляді деякої послідовності її окремих значень, виражених природною мовою у фіксовані моменти часу. Це окреме значення є інформаційним запитом, з яким користувач звертається до системи.

Інформаційний пошук повинен забезпечувати вирішення таких основних завдань:

1. пошук релевантної інформації, тобто такої, яка відповідає інформаційним потребам;
2. пошук аналогічної інформації в суміжних областях, тобто забезпечувати розширення вихідної інформації);
3. узагальнення та уточнення отриманої інформації;
4. аналіз та оцінка інформації, виходячи з власних завдань.

Інформаційний пошук має три основні цілі:

1. Пошук необхідних **відомостей про джерело** і встановлення його наявності в системі інших джерел. Здійснюється шляхом розшукування бібліографічної інформації та бібліографічних посібників (інформаційних видань), спеціально створених для більш ефективного пошуку і використання інформації (літератури, книги).
2. Пошук **самих інформаційних джерел** (документів і видань), в яких є або може міститися потрібна інформація.

3. Пошук фактичних відомостей, що містяться в інформаційних джерелах, наприклад літературі, книзі тощо. Наприклад, відомості про історичні факти і події, про технічні характеристики машин і процесів, про властивості речовин і матеріалів, про біографічних даних з життя і діяльності письменника, вченого і т.п.

1.2. Види інформаційного пошуку

Розрізняють такі види інформаційного пошуку:

- Залежно від мети – адресний (формально-механічний) та семантичний (тематичний);

Адресний пошук – процес пошуку за ознаками, зазначеним у запиті. Для здійснення такого пошуку потрібні наступні умови:

1. Наявність точної адреси документа
2. Забезпечення певного порядку розташування документів чи веб-сторінок в сховище системи.

Адресами документів можуть виступати як адреси веб-серверів і веб-сторінок, так і елементи бібліографічного запису, адреси зберігання документів у сховищі.

Семантичний пошук – процес пошуку документів за їхнім змістом.

Умови:

- Переклад змісту документів і запитів з природної мови на інформаційно-пошукову мову та складання пошукових образів документа та запиту.
- Складання пошукового опису, в якому вказується додаткова умова пошуку.

Принципова різниця між адресним і семантичним пошуком полягає в тому, що при адресному пошуку документ розглядається як об'єкт з точки зору форми, а при семантичному пошуку – з точки зору змісту.

При семантичному пошуку знаходиться безліч документів без вказівки адрес. У цьому принципова відмінність каталогів і картотек. Бібліотека - збори бібліографічних записів без вказівки адрес.

•Від об'єкта пошуку – **документальний і фактографічний**;

Документальний пошук – процес пошуку в сховищі інформаційно-пошукової системи первинних документів або в базі даних вторинних документів, які відповідають запиту користувача.

Два види документального пошуку:

1. Бібліотечний, спрямований на знаходження первинних документів.
2. Бібліографічний, спрямований на знаходження відомостей про документи, представлених у вигляді бібліографічних записів.

Фактографічний пошук – процес пошуку фактів, що відповідають інформаційному запиту.

До фактографічних даними відносяться відомості, отримані з документів, як первинних, так і вторинних, які одержуються безпосередньо з джерел їх виникнення.

Розрізняють два види:

1. Документально-фактографічний, полягає в пошуку в документах фрагментів тексту, що містять факти.
2. Фактологічний (опис фактів), що передбачає створення нових фактографічних описів в процесі пошуку шляхом логічної переробки знайденої фактографічної інформації.

•Від ступеня використання технічних засобів – ручний або автоматизований.

Однак вони тісно взаємопов'язані між собою. Наприклад, щоб знайти інформаційне джерело (документ або видання), треба знати певну сукупність бібліографічної інформації (фактів), що його характеризує, і що відрізняє його від багатьох інших. Наприклад, двох різних книжок написаних тим самим автором, на одну і ту ж тему, випущених тим же видавництвом, в одному і тому ж році і т.д. Отже, потрібно спочатку виконати бібліографічний пошук. І, навпаки, щоб провести фактографічний пошук у будь-якій галузі знань або практики потрібно спочатку знайти ті літературні джерела (документи, видання), в яких можуть бути цікаві факти. Тому спочатку треба провести бібліографічний та документальний пошук.

Всі види інформаційного пошуку перетинаються, так як їх цілі та об'єкти часто взаємопов'язані. Наприклад, документний та фактографічний види пошуку можуть бути як адресними, так і семантичними.