

УДК 004.4'2; 004.652.4

Матвійчук Т. – ст. гр. СНМ-52

Тернопільський національний технічний університет імені Івана Пулюя

МЕТОДИ ІНФОРМАЦІЙНОГО ПОШУКУ

Науковий керівник: к.т.н., доцент Мацюк О.В.

Matviichuk T.

Ternopil Ivan Pul'uj National Technical University

METHODS OF INFORMATION RETRIEVAL

Supervisor: Matsiuk O.

Ключові слова: пошук, запит, метод.

Keywords: search, inquiry, method.

Інформаційний пошук [1] – це сукупність операцій, необхідних для отримання інформації, що відповідає запиту користувача.

Він виконується за певними правилами, які визначають стратегію пошуку, тобто способи досягнення оптимального результату. Стратегія інформаційного пошуку залежить від типу пошукової задачі, критеріїв видачі і характеру діалогу між споживачами інформації і інформаційно пошукової системи (ІПС).

В загальному вигляді процедура інформаційного пошуку полягає в уточненні інформаційної потреби і формулюванні запиту, визначенні сукупності інформаційних масивів, вилученні інформації з інформаційних масивів, ознайомленні користувача з отриманою інформацією і оцінювання результатів пошуку.

При постановці пошукової проблеми користувач формулює точне визначення і фіксує те, що буде шукати і в якій області знань. Таким чином здійснюється звуження області пошуку.

На етапі створення тезаурусу проблеми користувач складає перелік слів, які найбільш повно відображають предметну область або проблему, що була визначена. Як рекомендують спеціалісти з бібліографічного пошуку, цей перелік повинен мати приблизно 10-15 слів.

В залежності від поставленого завдання тезаурус може бути складений на декількох мовах для пошуку серед вітчизняних та закордонних джерел інформації. Робота над тезаурусом ведеться весь час, і в процесі виявлення нових термінів вони тут же додаються до тезаурусу. Найбільш прийнятною є структура тезаурусу у вигляді семантичних зрізів. В цьому випадку для кожного основного терміну окремо будується таблиця для супутних та шумових слів.

Джерела інформації для пошуку вибираються виходячи з характеру проблеми та можливостей користувача.

На етапі виконання пошуку користувач складає пошукові запити з використанням тезаурусу і реалізує їх методами пошуку, які специфічні для даного ресурсу.

Запити необхідно складати таким чином, щоб область пошуку була максимально конкретизована та звужена. Доцільно віддавати перевагу декільком вузьким запитам ніж одному, але розширеному. В загальному випадку для кожного основного поняття з тезаурусу готується окремий пакет запитів. Після чого проводиться пробне виконання

запитів - для уточнення та доповнення тезаурусу, в тому числі для відсікання шумової інформації.

В результаті пошуку користувач отримує множину документів, які надалі необхідно проаналізувати і вирішити наскільки повно вони покривають поставлену пошукову проблему. Перелік ресурсів, отриманих в результаті запиту, рекомендується обробляти в два етапи. На першому етапі відсікаються вочевидь нерелевантні джерела і уточнення тезаурусу та модифікації подальших запитів. На другому етапі обробки користувач послідовно вивчає кожен з знайдених ресурсів для безпосереднього аналізу інформації, що знаходиться в них.

Якщо, оцінюючи результати пошуку, користувач прийшов до висновку, що знайдена вся необхідна інформація, тоді пошук можна припинити. У зворотній ситуації (неповні відомості) користувачеві доведеться приймати рішення про те, на якому з етапів була допущена помилка, і спробувати виправити її, після чого повторити процес пошуку з цього місця заново. В цьому випадку можливі три варіанти: невірно складений тезаурус проблеми, невірно обране інформаційне джерело користувач скористався недоцільними методами пошуку

Стандартну модель інформаційного пошуку представлено на рисунку 1.



Рисунок 1 – Класична модель інформаційного пошуку

Процес пошуку базується на використанні визначеної моделі пошуку. Модель пошуку характеризується наступними параметрами:

- форма подання документів и запитів;
- критерій змістовної відповідності;
- методи ранжування результатів запитів;
- механізм зворотного зв'язку для оцінювання релевантності документів.

На даний момент широко використовуються наступні методи інформаційного пошуку: метод, базований на булевій моделі; метод, базований на векторно-просторовій моделі; метод, базований на ймовірнісній моделі; метод, базований на дескрипторній моделі; метод, базований на класифікаторах.

Література:

1. Ландэ Д.В. Поиск знаний в INTERNET. Профессиональная работа / Д.В. Ландэ; пер. с англ.; [глав. ред. А.В. Слепцов] – М.: «Вильямс», 2005. – 352 с. – ISBN: 5-8459-0764-0.