

УДК 004.89; 004.71

Плетюк І. – ст. гр. СНм-51

Тернопільський національний технічний університет імені Івана Пулюя

МОДЕЛІ ІНФОРМАЦІЙНОГО ПОШУКУ

Науковий керівник: асистент Маєвський О.В.

Інформаційний пошук є одним з найважливіших напрямків дослідження в соціальній мережі та й загалом в мережі Інтернет. Тому, коли величезні масиви інформації стали доступними широкому колу користувачів, виникла потреба в локальному пошуці в певних Інтернет-ресурсах. Згідно зі статистикою, зібраною з найбільших пошукових систем, середня довжина користувацького запиту складає лише два слова в 75% випадків. Отже, точність таких запитів невисока і часто не задовольняє потреби користувачів в інформації.

Модель інформаційного пошуку включає три ключових аспекти.

- Модель представлення документу. Під документом будемо розуміти деякий об'єкт, що містить інформацію у фіксованому вигляді. Документи можуть містити тексти на потрібній або формалізованій мові, зображення, звук і т.д.
- Формат представлення запиту. Під запитом будемо розуміти формалізований спосіб вираження інформаційних потреб користувача системи. Для цього використовується мова пошукових запитів, синтаксис яких залежить від системи.
- Функція відповідності документа запиту. Ступінь відповідності запиту знайденого документа (релевантність) — суб'єктивне поняття, адже результати пошуку, притаманні для одного користувача можуть бути не притаманними для іншого.

Існують такі моделі інформаційного пошуку.

- векторно-просторового представлення (vector space model);
- ймовірності появи пошукового терміну в документі (probabilistic retrieval);
- з побудовою мовної моделі для кожного документу (language models);
- з побудовою мережі припущень, яка використовується для встановлення відповідності документу пошуковому запиту (inference network);
- з булевим індексуванням (Boolean indexing);
- з використанням латентного семантичного індексування (latent semantic indexing);
- з побудовою нейромереж (neural networks);
- з використанням генетичних алгоритмів (genetic algorithms);
- з використанням нечітких множин (fuzzy set retrieval).