

УДК 044.42

АНАЛІЗ ОСНОВНИХ ХАРАКТЕРИСТИК СИСТЕМ INFORMATION RETRIEVAL У КОНТЕКСТІ КОМП'ЮТЕРНИХ НАУК

Астістова Т.І., кандидат технічних наук, доцент

Київський національний університет технологій та дизайну

Лінчак Н.К., студент

Київський національний університет технологій та дизайну

Сомок І.О., студент

Київський національний університет технологій та дизайну

Ключові слова: Information Retrieval, NLP, ML, алгоритм, інтерфейс, пошуковий рушій.

Головною метою автоматизованої системи інформаційного пошуку - IR (Information Retrieval) знаходження неструктурованої інформації (зазвичай тексту) у великих колекціях даних, що відповідає інформаційній потребі користувача.

Найвідомішими прикладами IR-систем є вебпошуковики (Google, Bing), системи пошуку всередині великих корпоративних архівів або бібліотек, а також модулі пошуку в коді (наприклад, GitHub Search).

Оцінка релевантності в системах інформаційного пошуку (Information Retrieval, IR) трансформувалася від простих методів порівняння термів до складних когнітивних архітектур. Для галузі комп'ютерних наук цей процес є ключовим прикладом еволюції обробки природної мови (NLP) та машинного навчання (ML).

Розглянемо основні характеристики IR-систем у контексті комп'ютерних наук:

1. Об'єкт пошуку. На відміну від баз даних (пошук ведеться за чіткими полями та структурами), IR працює з текстами, зображеннями, аудіо та відео, де зміст не є формалізованим.
2. Запит (Query). Користувач вводить запит природною мовою або набором ключових слів.
3. Нечітка відповідність. Система обчислює ступінь відповідності і видає ранжований список результатів.
4. Компоненти:
 - Індексатор. Перетворює документи у зручну для пошуку форму (наприклад, інвертований індекс).
 - Пошуковий рушій. Алгоритм, який зіставляє запит з індексом.
 - Інтерфейс. Засіб взаємодії з користувачем.

Пошуковий рушій та інтерфейс є фундаментальними компонентами будь-якої сучасної системи пошуку інформації.

На основі проведеного аналізу інтерфейсів, було виділено три види інтерфейсів, які виконують різні завдання.

1. Класичний текстовий інтерфейс (Пошуковий рядок). Приклад: Google, Bing, Wikipedia. Головна риса — мінімалізм. Складові інтерфейсу:

- *Елементи*: Поле введення, кнопка пошуку, список результатів (SERP).
- *Приховані механізми*. * Автодоповнення (Suggest). Алгоритм передбачає запит у реальному часі.

2. Фасетний інтерфейс (Навігаційний пошук). Приклад інтернет-магазинів Amazon, Rozetka. Дозволяє звузити вибірку за допомогою конкретних параметрів:

- Бічна панель (Фільтри). Приклад: Ціна, бренд, технічні характеристики, рейтинг).
- Динамічне оновлення. Кожен вибраний фільтр перераховує кількість знайдених документів у реальному часі.
- Хлібні крихти (Breadcrumbs). Шлях користувача (наприклад: *Електроніка > Телефони > Apple*).

3. Діалоговий інтерфейс (Conversational IR). Чат. Приклад: Gemini, ChatGPT, Siri. Параметри:

- Контекстність. Система пам'ятає попередні запити.
- Мультиmodalність. Можливість завантажити картинку або голос як запит.

В таблиці 1 представлено таблицю елементів інтерфейсу.

Таблиця 1

Порівняльна таблиця елементів інтерфейсу

Елемент	Функція	Чому це важливо для IR
Search Box	Введення даних	Початкова точка взаємодії.
Facets/Filters	Уточнення запиту	Допомагає користувачу впоратися з надлишком інформації.
Ranking/Sorting	Упорядкування	Дозволяє вибрати критерій (за ціною, за релевантністю).
Pagination	Розподіл результатів	Не перевантажує користувача всіма даними одразу.

В роботі також було проведено аналіз еволюції алгоритмічних моделей оцінки релевантності, які представлені п'ятьма періодами в часі.

1. Класичні статистичні моделі (1960-ті — 1980-ті).

- Булева модель;
- Векторна модель (Vector Space Model, VSM);
- TF-IDF (Term Frequency-Inverse Document Frequency).

2. Ймовірнісні моделі та BM25 (1990-ті).

- Принцип імовірнісного пошуку;
- Окарі BM25.

3. Моделі машинного навчання та ранжування (2000-ні).

- Pointwise, Pairwise, Listwise підходи;
- Градієнтний бустинг (XGBoost, CatBoost, LightGBM).

4. Семантичний пошук та нейромережі (2010-ті — 2020).

- Word Embeddings (Word2Vec, GloVe);
- BERT та архітектура Transformer.

5. Сучасний етап: LLM та генеративний пошук (2023 — сьогодні)

Для реалізації цього підходу використовуються такі методи:

- Dense Retrieval (Векторні бази даних);
- RAG (Retrieval-Augmented Generation): Моделі (як GPT-4);
- Cross-Encoders: Нейромережі.

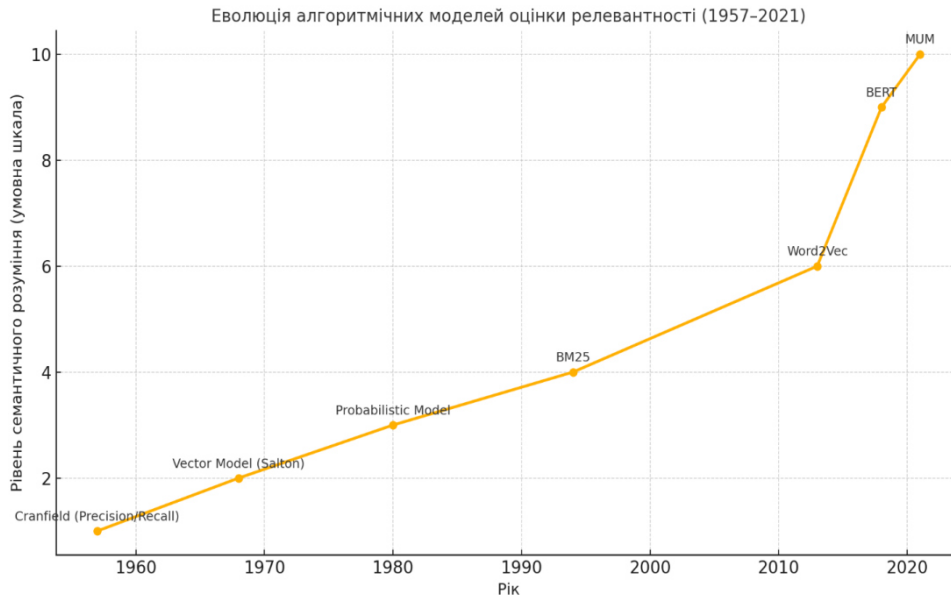


Рисунок 1 – Еволюція алгоритмічних моделей оцінки релевантності

На рисунку 1 показано узагальнену динаміку розвитку основних моделей обробки текстових даних. Вісь X відображає роки появи моделей, а вісь Y — умовну шкалу рівня здатності алгоритму аналізувати контекст.

З графіка видно, що починаючи з 2010 року спостерігається різкий ріст цієї здатності, що співпадає з поширенням нейронних мереж і моделей Висновки.

1. Еволюція моделей релевантності демонструє рух від лексичного до семантичного розуміння і, нарешті, до когнітивного синтезу.

2. Еволюція систем Information Retrieval оцінки відображає загальний тренд Computer Science: від жорстких алгоритмічних правил до гнучких нейромережових структур, що здатні до узагальнення. Майбутнє систем IR полягає в інтеграції мультимодальності (пошук за зображеннями, відео та текстом одночасно) та забезпеченні повної прозорості.

Список використаних джерел

1. Дорохин С., Вавилов П. Інформаційні системи в управлінні проектами. Видавництво Національного технічного університету України, 2018. - 386 с.
2. Petrov A. Database Internals: A Deep Dive into How Distributed Data Systems Work. O'Reilly Media, 2019. – 632 p.