

Giới thiệu

Bạn tò mò về "phép thuật" đằng sau những kết quả tìm kiếm? Phần này sẽ bật mí tất cả. Chúng ta sẽ học nguyên lý cơ bản của hệ thống truy vấn và hiểu cách các công cụ tìm kiếm hoạt động.

Bạn sẽ nắm vững những kỹ thuật then chốt—những kỹ năng tưởng chừng chỉ dành cho chuyên gia.

1. Tokenization & Tiền Xử Lý Dữ Liệu

- Tokenization:** Bước đầu tiên để xử lý dữ liệu văn bản.
- Thực hành các kỹ thuật tiền xử lý, đảm bảo dữ liệu sẵn sàng cho phân tích.

2. Xây Dựng Các Loại Hệ Thống Truy Vấn

- Hệ thống Boolean:** Sử dụng AND, OR, NOT.
- Mô hình Không Gian Vector (VSM):** Ứng dụng TF-IDF.
- Mô hình Xác Suất Truy Vấn.**

3. Truy Vấn & Xếp Hạng Kết Quả

- Yếu tố then chốt để trả về kết quả tìm kiếm chất lượng cao.

4. Kỹ Năng Lập Trình Thực Tế

- Tokenize và tiền xử lý văn bản.
- Xây dựng & truy vấn **inverted index** (chỉ mục ngược).
- Áp dụng các mô hình truy vấn.

Bạn sẽ đạt được

- Thành thạo tokenization & tiền xử lý.
- Hiểu sâu các mô hình: Boolean, Vector Space, Xác suất.
- Tự tin xây dựng và truy vấn inverted index.
- Kinh nghiệm lập trình ứng dụng ngay.