

# Mô Hình Truy Xuất Thông Tin Boolean (Boolean Retrieval Model)

## 1. Mục Tiêu Bài Học

- Hiểu được khái niệm và cách hoạt động của mô hình Boolean
- Sử dụng được các phép toán logic (AND, OR, NOT) để truy vấn văn bản
- Biết ưu và nhược điểm của mô hình này
- Thực hành với ví dụ minh họa cụ thể

## 2. Khái Niệm Cơ Bản

**Boolean Retrieval Model** là một mô hình truy xuất thông tin trong đó:

- Tài liệu và truy vấn đều được biểu diễn bằng các tập hợp từ (terms).
- Người dùng sử dụng các phép **toán logic** để tìm tài liệu phù hợp.
- Kết quả truy vấn là **danh sách các tài liệu** thỏa mãn điều kiện logic đó.

## 3. Các Phép Toán Logic Cơ Bản

Phép Toán	Ý Nghĩa	Ví dụ Truy Vấn
AND	Cả hai từ đều phải xuất hiện	máy AND học
OR	Một trong hai từ xuất hiện	máy OR học
NOT	Loại bỏ tài liệu chứa từ đó	máy AND NOT học
Kết hợp	Dùng ngoặc để nhóm biểu thức phức tạp	(máy AND học) OR AI

## 4. Ví Dụ Cụ Thể

### Tập Tài Liệu

Tài liệu	Nội dung
D1	"Tôi yêu học máy và AI"
D2	"Học sâu là một nhánh của AI"
D3	"Máy học khác với lập trình truyền thống"
D4	"Tôi học lập trình Python"

Truy Vấn 1: học AND máy

- Phân tích:
  - Tìm các tài liệu chứa cả 2 từ: học và máy
- Kết quả:
  - D1 (chứa cả "học" và "máy")
  - D3 (cũng chứa cả hai)

Kết quả: D1, D3

Truy Vấn 2: AI OR Python

- Phân tích:
  - Chỉ cần một trong hai từ xuất hiện
- Kết quả:
  - D1, D2 (chứa "AI")
  - D4 (chứa "Python")

Kết quả: D1, D2, D4

Truy Vấn 3: học AND NOT AI

- Phân tích:
  - Tài liệu có "học" nhưng không có "AI"
- Kết quả:
  - D4 (có "học", không có "AI")

Kết quả: D4

5. Biểu Diễn Dưới Dạng Ma Trận Boolean

Tài liệu	học	máy	AI	Python
D1	1	1	1	0
D2	1	0	1	0
D3	1	1	0	0
D4	1	0	0	1

Truy vấn "học AND máy"  $\Rightarrow$  Chỉ những hàng có cả hai giá trị là 1 tại cột "học" và "máy".

## 6. Ưu và Nhược Điểm

### Ưu điểm:

- Đơn giản, dễ hiểu
- Truy vấn chính xác và rõ ràng
- Hiệu quả với các tập tài liệu nhỏ

### ❑ Nhược điểm:

- Không hỗ trợ tìm kiếm mờ (fuzzy search)
- Không xếp hạng mức độ liên quan giữa các tài liệu
- Không linh hoạt nếu người dùng không biết chính xác từ khóa

## 7. Ứng Dụng Thực Tế

- Truy vấn luật trong cơ sở dữ liệu pháp lý
- Hệ thống quản lý tài liệu nội bộ
- Công cụ tìm kiếm cơ bản trong ứng dụng nhỏ

## 8. Kết Luận

- Boolean Retrieval là nền tảng của hệ thống tìm kiếm hiện đại
- Dù đơn giản, nhưng nó tạo nền móng cho các mô hình nâng cao hơn như: TF-IDF, BM25, hay Vector Space Model.
- Khi kết hợp với xử lý ngôn ngữ tự nhiên (NLP), nó trở nên mạnh mẽ hơn

**Tác giả: Đỗ Ngọc Tú**  
**Công Ty Phần Mềm VHTSoft**

---

Phiên bản #1

Được tạo 4 tháng 5 2025 14:15:14 bởi Đỗ Ngọc Tú

Được cập nhật 6 tháng 5 2025 09:52:34 bởi Đỗ Ngọc Tú