

# **TRUY XUẤT THÔNG TIN**

## **CHƯƠNG III – ĐÁNH GIÁ MÔ HÌNH TRUY XUẤT THÔNG TIN**

# NỘI DUNG TRÌNH BÀY

- ❖ ĐÁNH GIÁ HỆ THỐNG TRUY VẤN THÔNG TIN
- ❖ CÁC TẬP DỮ LIỆU KIỂM TRA CHUẨN
- ❖ ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG
- ❖ ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

# ĐÁNH GIÁ HỆ THỐNG TRUY VẤN THÔNG TIN

## ❖ MỤC TIÊU ĐÁNH GIÁ

- Đảm bảo hệ thống có khả năng đáp ứng yêu cầu.
- So sánh với các hệ thống khác.
- Thử nghiệm và cải tiến thuật toán truy xuất thông tin.

# ĐÁNH GIÁ HỆ THỐNG TRUY VẤN THÔNG TIN

## ❖ TIÊU CHÍ ĐÁNH GIÁ

- Khả năng đáp ứng yêu cầu thông tin.
- Khả năng hiểu của hệ thống.
  - Hiểu nội dung tập tài liệu
  - Hiểu nội dung câu truy vấn
- Sự thân thiện với người sử dụng của hệ thống

# ĐÁNH GIÁ HỆ THỐNG TRUY VẤN THÔNG TIN

## ❖ SỰ LIÊN QUAN GIỮA TÀI LIỆU VÀ TRUY VẤN

- Tài liệu trả về chính xác đối với truy vấn.
- Tài liệu trả về đáp ứng một phần so với yêu cầu của truy vấn.
- Cung cấp nguồn thông tin liên quan để đáp ứng thông tin của truy vấn.
- Cung cấp thông tin cơ bản về vấn đề nêu trong truy vấn.
- Gợi ý một vấn đề khác có liên quan.

# ĐÁNH GIÁ HỆ THỐNG TRUY VẤN THÔNG TIN

## ❖ CHỈ TIÊU ĐÁNH GIÁ

- Độ bao quát thông tin so với yêu cầu.
- Hình thức trình bày kết quả.
- Sử dụng phức tạp hay đơn giản.
- Thời gian truy xuất thông tin
- Kích thước dữ liệu mà hệ thống sử dụng.
- Độ phủ (Recall)
- Độ chính xác (Precision)

→ Độ phủ và độ chính xác là 2 chỉ tiêu đánh giá tính hiệu quả của hệ thống

# CÁC TẬP DỮ LIỆU KIỂM TRA CHUẨN

## ❖ CÁC TẬP DỮ LIỆU THỜI KỲ ĐẦU

- Số lượng tài liệu ít.
  - Số lượng truy vấn ít.
- Chỉ thích hợp cho thử nghiệm ban đầu.

Các bộ dữ liệu gồm:

- CACM (3204 tài liệu).
- CISI (1460 tài liệu)
- CRAN (1397 tài liệu)
- INSPEC (12684 tài liệu)
- MED (1033 tài liệu)

# CÁC TẬP DỮ LIỆU KIỂM TRA CHUẨN

## ❖ CÁC TẬP DỮ LIỆU THỜI KỲ ĐẦU

- Truy vấn và đánh giá độ liên quan:
  - Thực hiện thủ công bằng cách đưa ra truy vấn theo ý định của các tác giả
  - Các tài liệu liên quan tương ứng được xác định trên toàn bộ tập tài liệu và được lập thành danh sách.



# CÁC TẬP DỮ LIỆU KIỂM TRA CHUẨN

## ❖ TẬP DỮ LIỆU CỦA TREC

- Số lượng hơn 1,5 triệu tài liệu, chiếm dung lượng khoảng 5GB với các loại (2013): tin tức, tài liệu của chính phủ, bằng phát minh sang chế ở Hoa Kỳ, ...
- Truy vấn và đánh giá độ liên quan:
  - Do chuyên viên thông tin đưa ra và đánh giá.
  - Đánh giá độ liên quan dựa trên tài liệu trả về, không đánh giá trên toàn bộ tập tài liệu
- Chỉ tiêu đánh giá: Độ phủ (Recall) và độ chính xác (Precision)

# CÁC TẬP DỮ LIỆU KIỂM TRA CHUẨN

## ❖ TẬP DỮ LIỆU CỦA TREC

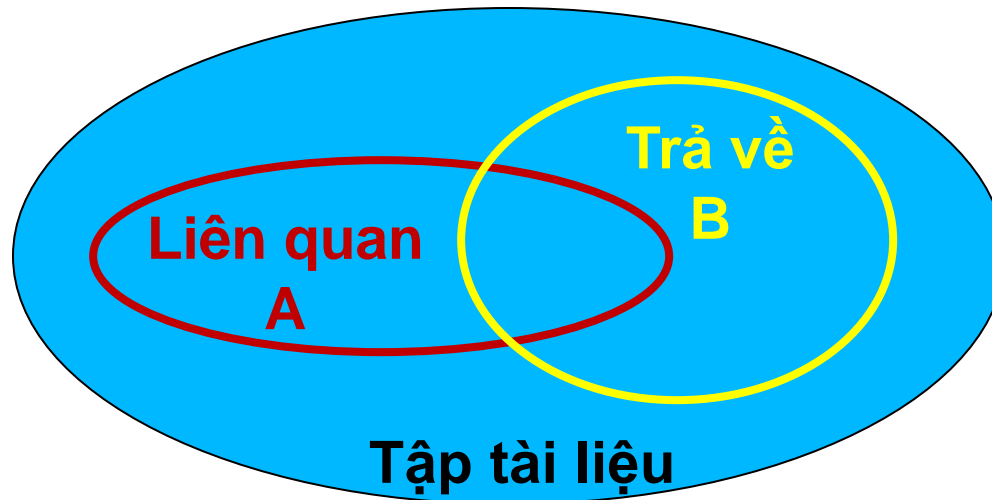
- Ngoài truy xuất thông tin dạng tài liệu, TREC còn tổ chức đánh giá truy xuất thông tin với các dạng:
  - Hệ thống hỏi đáp (QA System)
  - Tìm kiếm thông tin trên Web (Web Searching)
  - Tìm theo phim, âm thanh, hình ảnh.
  - Tìm xuyên ngôn ngữ (Cross-lingual).

# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

Độ phủ cho 1 truy vấn:  $R = |A \cap B| / |A|$

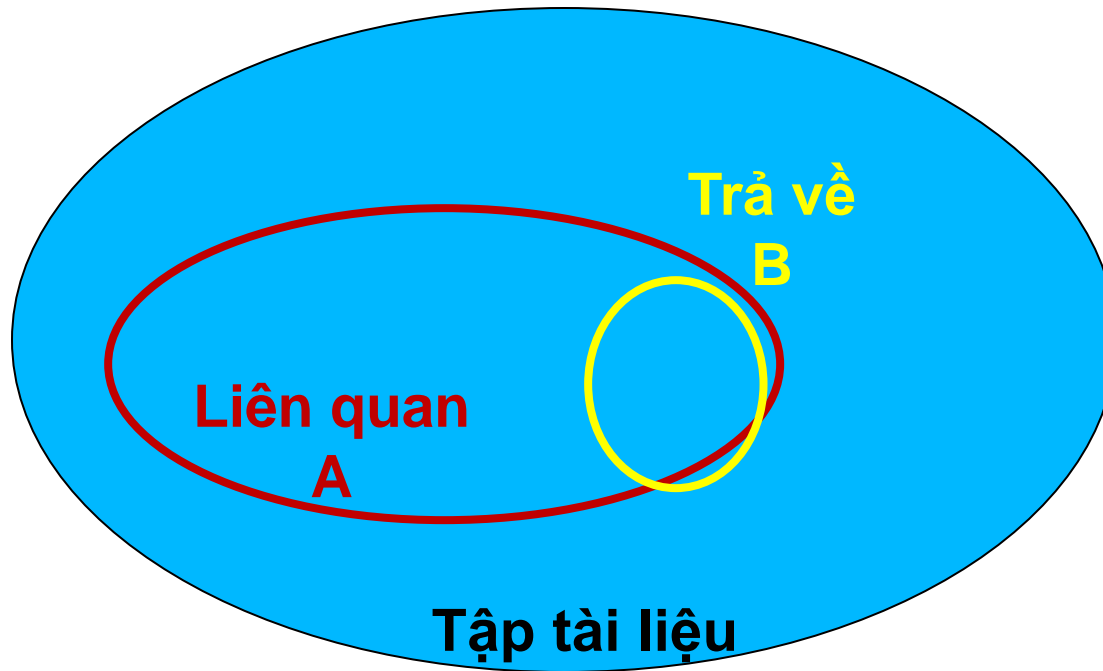
Độ chính xác cho 1 truy vấn:  $P = |A \cap B| / |B|$



# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

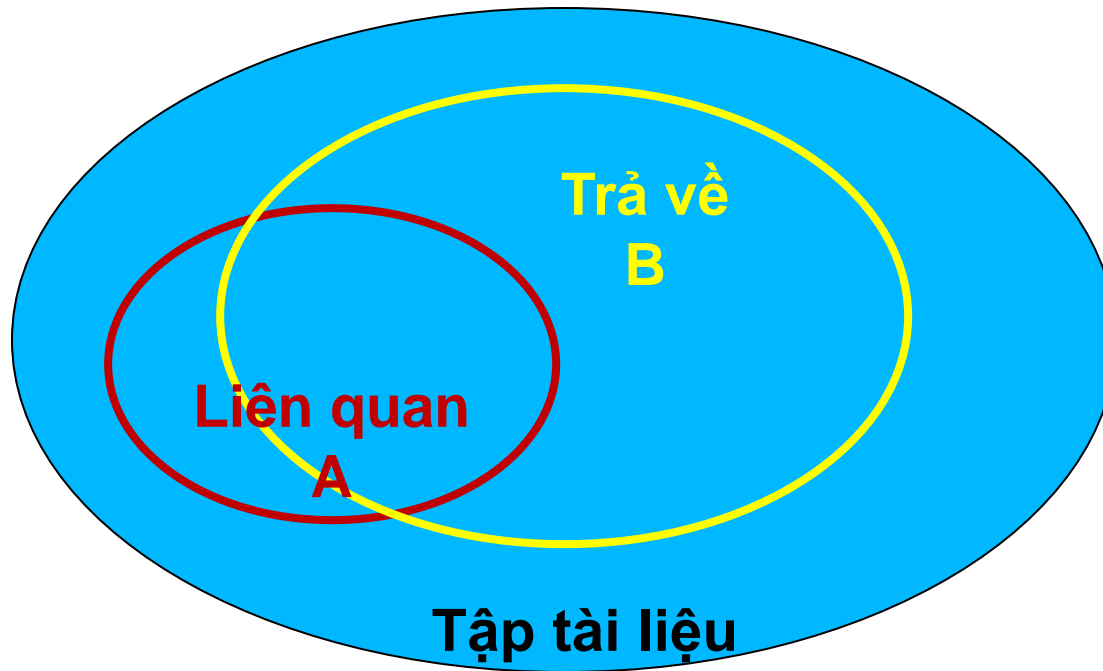
Trường hợp độ chính xác cao, độ phủ thấp



# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

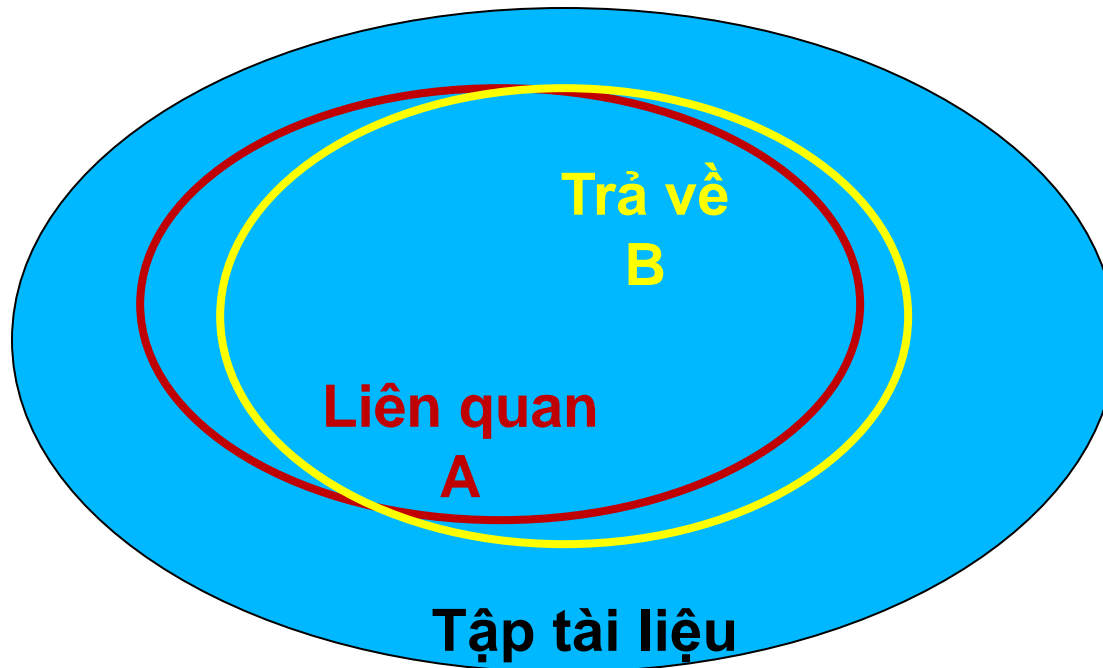
Trường hợp độ chính xác thấp, độ phủ cao



# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

Trường hợp độ chính xác cao, độ phủ cao



# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

Độ phủ và độ chính xác trung bình tương ứng với tập dữ liệu thử nghiệm được tính là trung bình cộng của độ phủ và độ chính xác tương ứng với từng truy vấn

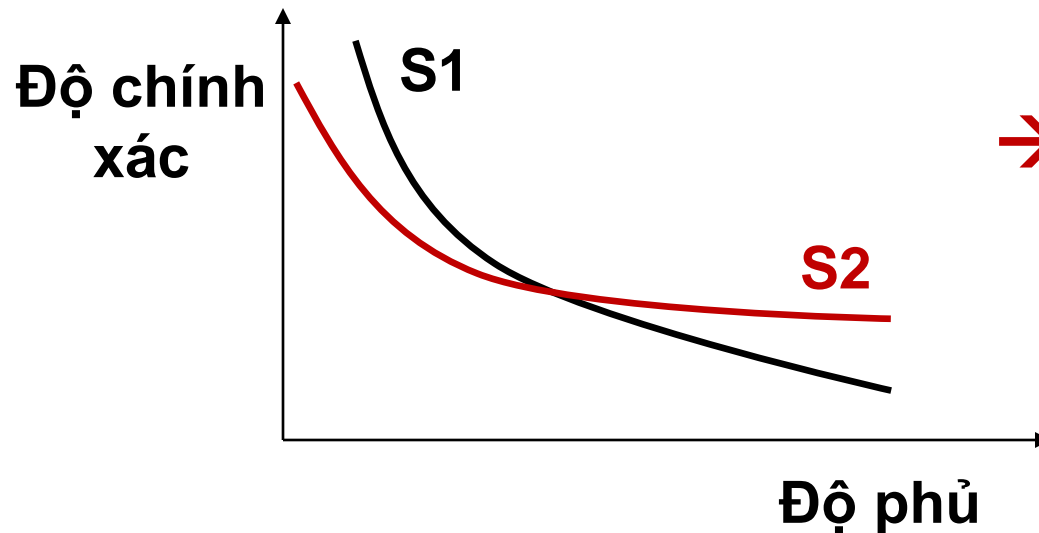
$$\overline{P} = \frac{1}{|Q|} \sum_{q \in Q} P_q$$

$$\overline{R} = \frac{1}{|Q|} \sum_{q \in Q} R_q$$

# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

- Có sự đánh đổi giữa độ phủ và độ chính xác
- Đồ thị thể hiện độ phủ và độ chính xác tính trung bình của 2 hệ thống IR S1 và S2 trên nhiều truy vấn có dạng:



→ Khó so sánh



# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ ĐO F (F-MEASURE)

Kết hợp độ phủ và độ chính xác tương ứng với 1 truy vấn thành một số tương tự như trung bình

$$F = \frac{1}{\alpha \left( \frac{1}{P} \right) + (1 - \alpha) \frac{1}{R}} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

Với  $\beta^2 = \frac{1 - \alpha}{\alpha}$

# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ ĐỘ ĐO F (F-MEASURE)

Trường hợp cân bằng giữa độ phủ và độ chính xác ( $\beta=1$ )

$$F = \frac{2PR}{P + R}$$

Độ đo F trung bình của bộ dữ liệu thử nghiệm bằng trung bình cộng của độ đo F tương ứng với từng truy vấn

$$\overline{F} = \frac{1}{|Q|} \sum_{q \in Q} F_q$$

# ĐÁNH GIÁ CÁC KẾT QUẢ CHƯA XẾP HẠNG

## ❖ VÍ DỤ

Cho tập tài liệu liên quan đến truy vấn là:

$$R = \{3, 5, 9, 25, 39, 44, 56, 71, 89, 123\}$$

Cho kết quả truy vấn gồm tập các tài liệu được xếp hạng như sau:

$$\{123, 84, 56, 6, 8, 9, 511, 129, 187, 25, 38, 48, 250, 113, 3\}$$

$$\text{Độ phủ:} \quad R = 5/10 = 0.5$$

$$\text{Độ chính xác} \quad P = 5/15 = 0.33$$

$$\text{Độ } F_1 \quad F_1 = 2PR / (P + R) = 0.398$$

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ NHƯỢC ĐIỂM CỦA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

- Độ phủ và độ chính xác có quan hệ với nhau
- Đối với tập tài liệu lớn thì không thể xác định độ phủ.

→ Cần phương pháp đánh giá trong trường hợp tập tài liệu lớn.

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP CHIA CẮT

- Áp dụng đối với tập tài liệu lớn, không thể xác định tập tài liệu liên quan trên toàn bộ tập tài liệu.
- Phương pháp thực hiện:
  - Kết quả trả về của 1 truy vấn từ hệ thống phải được xếp hạng.
  - Chọn các tập kết quả là  $n$  tài liệu đầu tiên ( $n = 5, 10, 20, 50, 100, 500$ )
  - Tính độ chính xác tại mỗi tập kết quả  $P@n$
  - Tính độ chính xác trung bình (có thể có trọng số) từ độ chính xác của mỗi tập kết quả.

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ VÍ DỤ:

Cho kết quả truy vấn gồm tập các tài liệu được xếp hạng như sau: (tài liệu tô đậm là có liên quan)

{**123**, 84, **56**, 6, 8, **9**, 511, 129, 187, **25**, 38, 48, 250, 113, **3**}

Độ chính xác tại mức cắt 5:  $P@5 = 2/5 = 0.4$

Độ chính xác tại mức cắt 10:  $P@10 = 4/10 = 0.4$

Độ chính xác trung bình:  $P = 0.4*0.5+0.4*0.5=0.4$

(giả sử chọn trọng số của mỗi mức cắt là như nhau)

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP TƯƠNG QUAN GIỮA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

- Áp dụng khi có thể xác định tập tài liệu liên quan trên toàn bộ tập tài liệu.
- Phương pháp thực hiện:
  - Kết quả trả về của 1 truy vấn từ hệ thống phải được xếp hạng.
  - Tại mỗi thứ hạng có tài liệu liên quan trong kết quả, tính độ chính xác và độ phủ.
  - Dựa trên những điểm tính được độ phủ và độ chính xác để nội suy những điểm cần so sánh.

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP TƯƠNG QUAN GIỮA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

- Nguyên tắc nội suy như sau:
  - Giả sử có các điểm tính được độ phủ là  $i = 1, 2, \dots$ , giá trị độ phủ và độ chính xác tương ứng là  $r_i$  và  $p_i$
  - Khi đó, độ chính xác  $p$  tương ứng với độ phủ  $r$  cần nội suy được xác định:

$$p = \max(p_{r'}) \text{ với } r \leq r'$$



# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP TƯƠNG QUAN GIỮA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

- Mục đích của phương pháp nhằm so sánh độ chính xác của các hệ thống tại một độ phủ  $r$  xác định.

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ VÍ DỤ

Cho tập tài liệu liên quan đến truy vấn là:

$$R = \{3, 5, 9, 25, 39, 44, 56, 71, 89, 123\}$$

Cho kết quả truy vấn gồm tập các tài liệu được xếp hạng như sau: {**123**, 84, **56**, 6, 8, **9**, 511, 129, 187, **25**, 38, 48, 250, 113, **3**}

Yêu cầu: Tính p tại điểm có  $r = 0.15$ .

- Tài liệu liên quan đầu tiên ở vị trí 1:

$$p_1 = 1/1 = 1.0; \quad r_1 = 1/10 = 0.1$$

- Tài liệu liên quan thứ hai ở vị trí 3:

$$p_2 = 2/3 = 0.66; \quad r_2 = 2/10 = 0.2$$

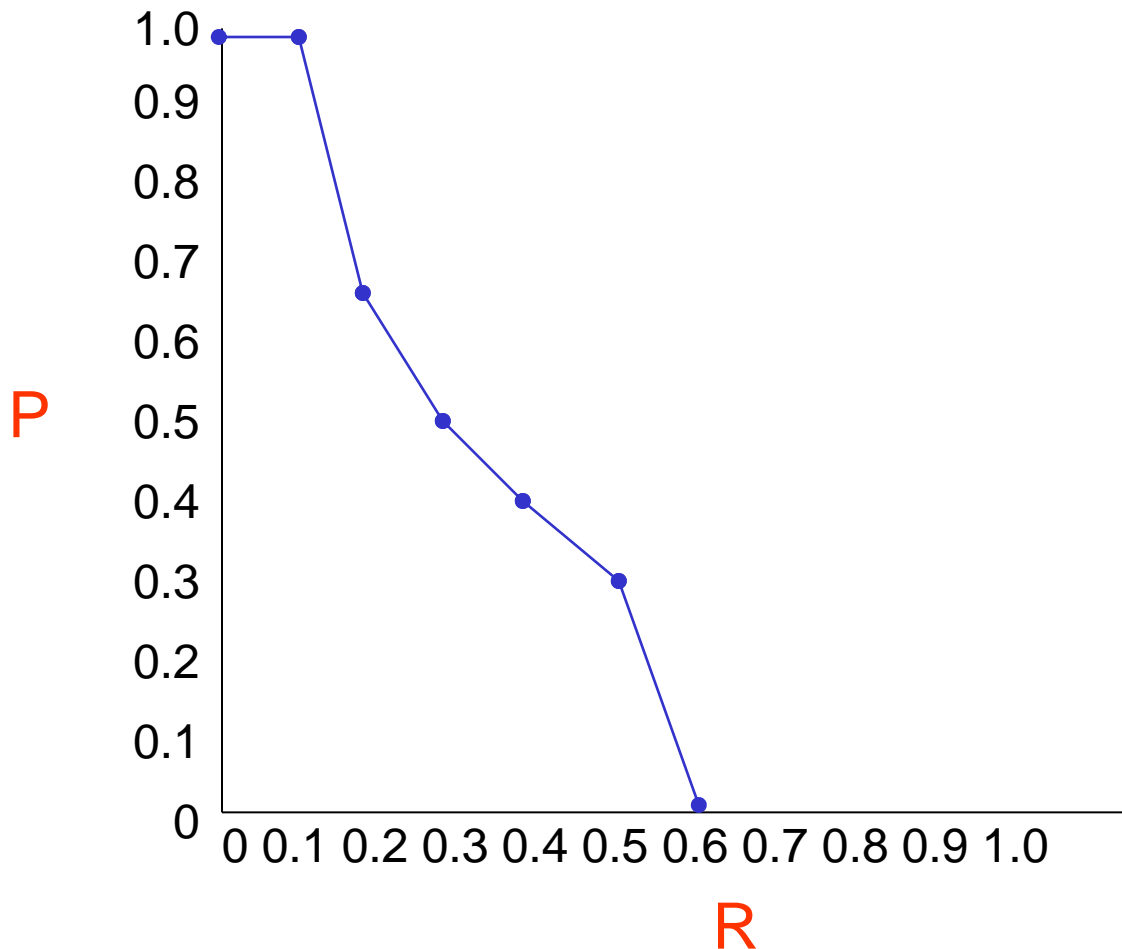
# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ VÍ DỤ

- Tài liệu liên quan thứ 3 ở vị trí 6:  
 $p_3 = 3/6 = 0.5; \quad r_3 = 3/10 = 0.3$
- Tài liệu liên quan thứ 4 ở vị trí 10:  
 $p_4 = 4/10 = 0.4; \quad r_4 = 4/10 = 0.4$
- Tài liệu liên quan thứ 5 ở vị trí 15:  
 $p_5 = 5/15 = 0.33; \quad r_5 = 5/10 = 0.5$
- Với  $r = 0.15, p = p_2 = 0.66$

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

❖ VÍ DỤ



# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP TƯƠNG QUAN GIỮA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

### Độ chính xác trung bình không nội suy

- Average Precision
- Tính theo từng truy vấn
- Bằng độ chính xác trung bình tại các độ phủ khác nhau của truy vấn:

$$AP = 1/|r| * \sum_r p_r$$

# ĐÁNH GIÁ CÁC KẾT QUẢ ĐÃ XẾP HẠNG

## ❖ PHƯƠNG PHÁP TƯƠNG QUAN GIỮA ĐỘ PHỦ VÀ ĐỘ CHÍNH XÁC

### Độ chính xác trung bình kỳ vọng

- Mean Average Precision
- Tính trên tất cả truy vấn
- Bằng giá trị trung bình của độ chính xác trung bình của từng truy vấn:

$$\text{MAP} = 1/|Q| * \sum_Q p_q$$