



25  
SOICT

ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

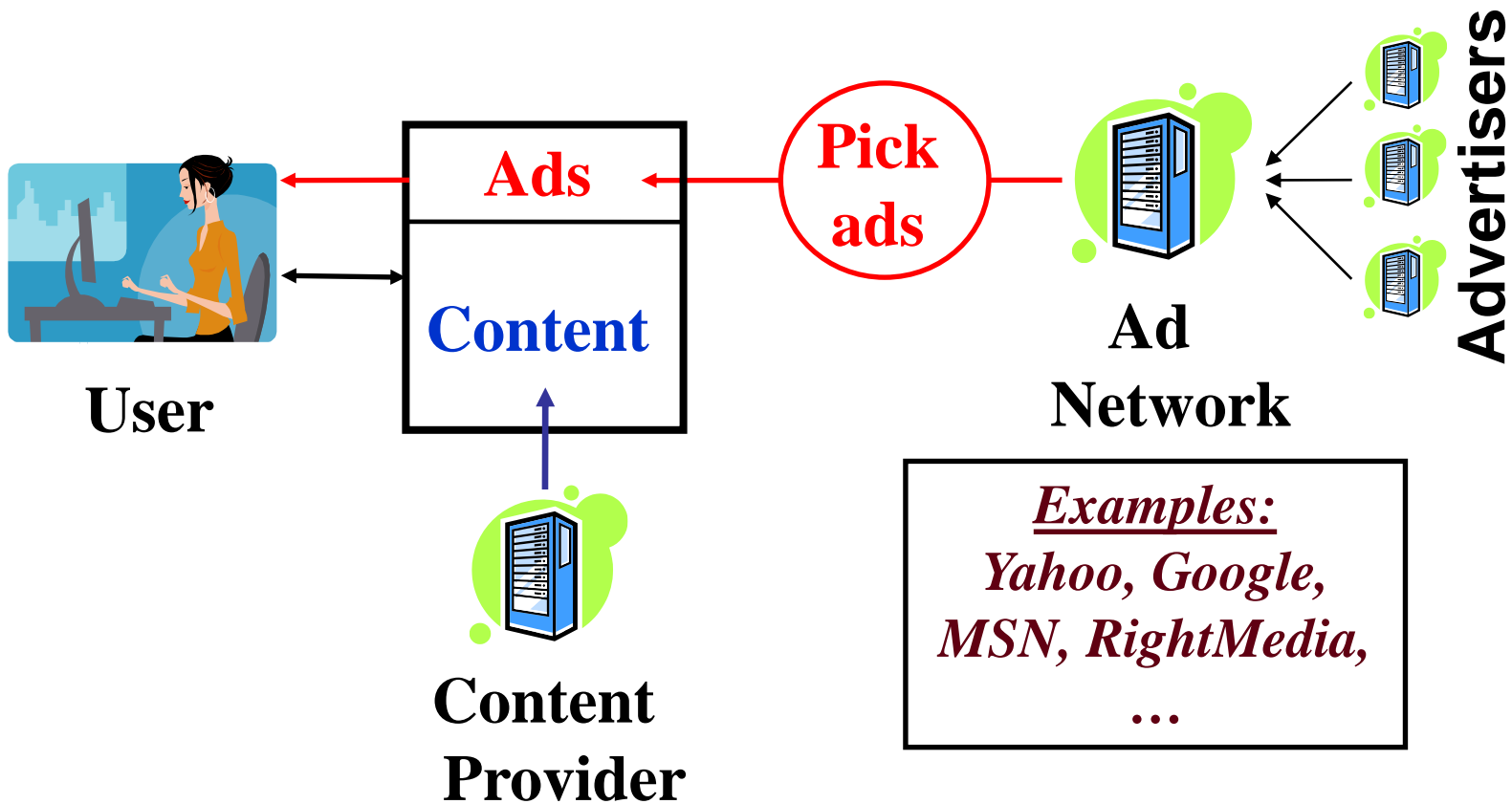
# BÀI 9: CHỦ ĐỀ NÂNG CAO

## Quảng cáo trực tuyến – Khai phá truy vấn

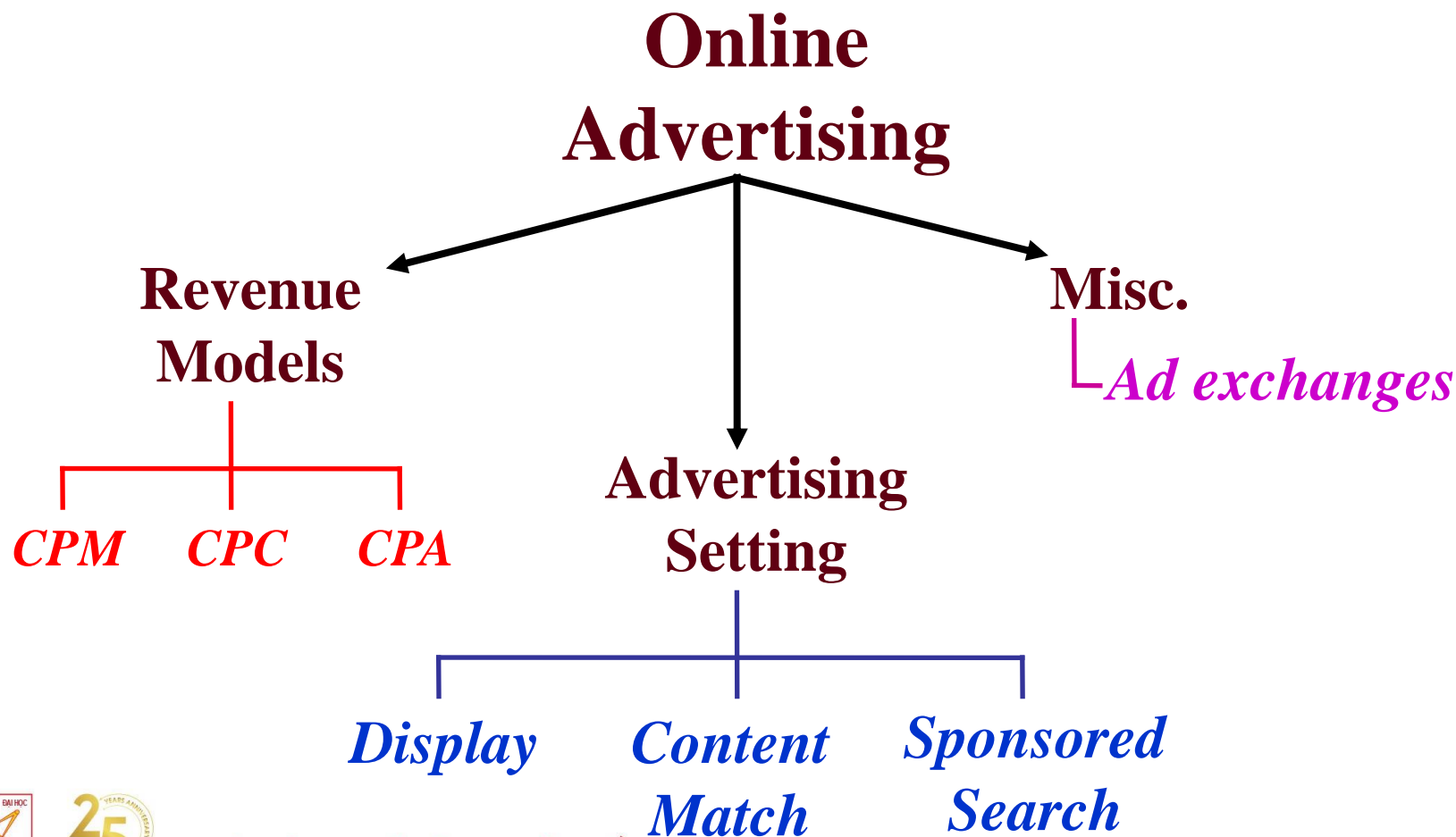
# Nội dung

1. Quảng cáo trực tuyến
2. Quảng cáo trên máy tìm kiếm
3. Khai phá truy vấn

# 1. Quảng cáo trực tuyến



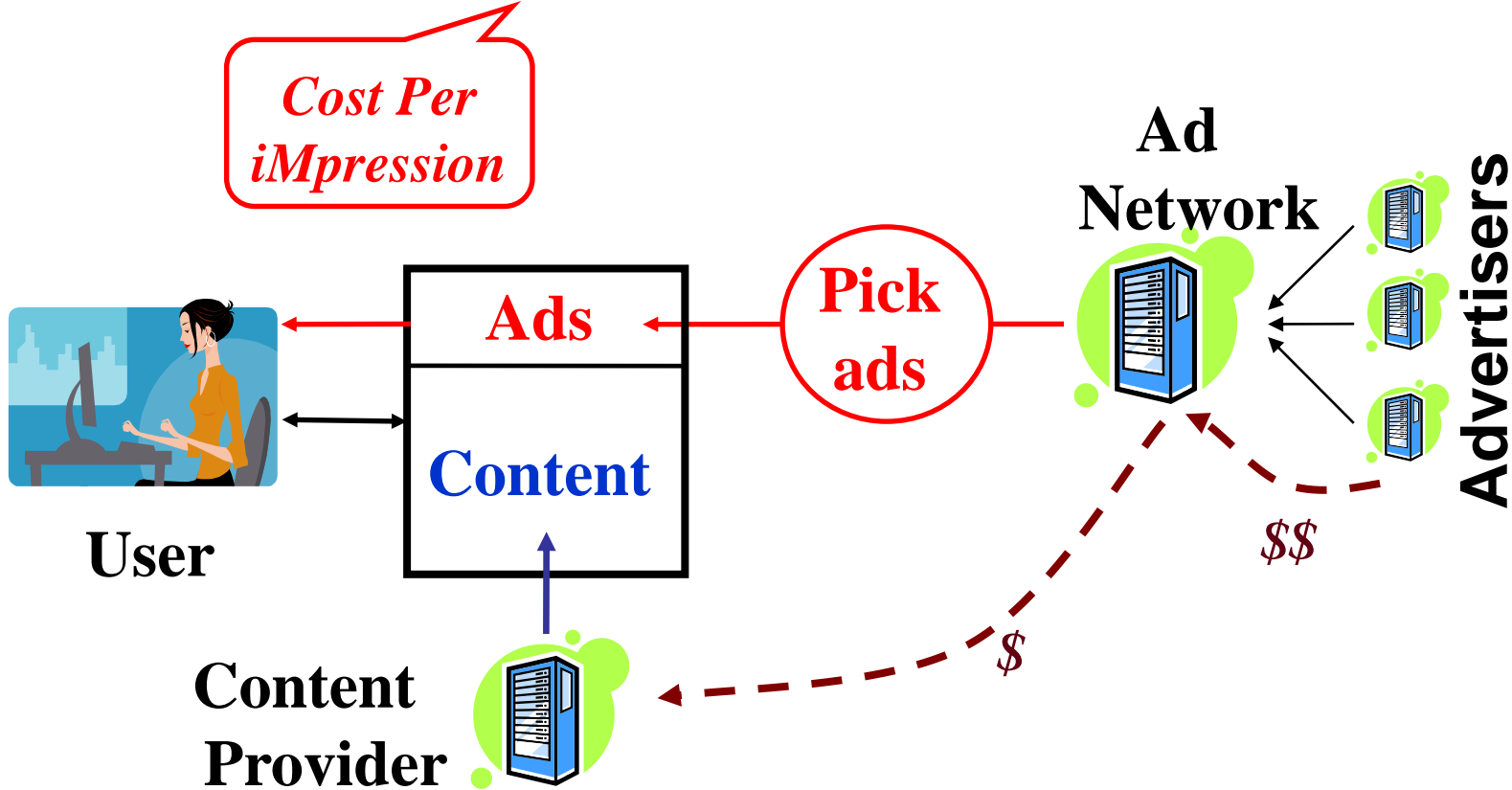
# Các mô hình quảng cáo trực tuyến



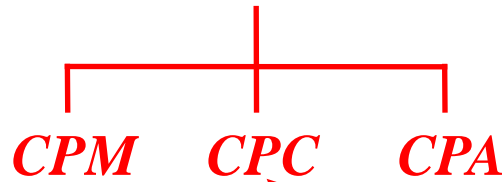
# CPM

CPM    CPC    CPA

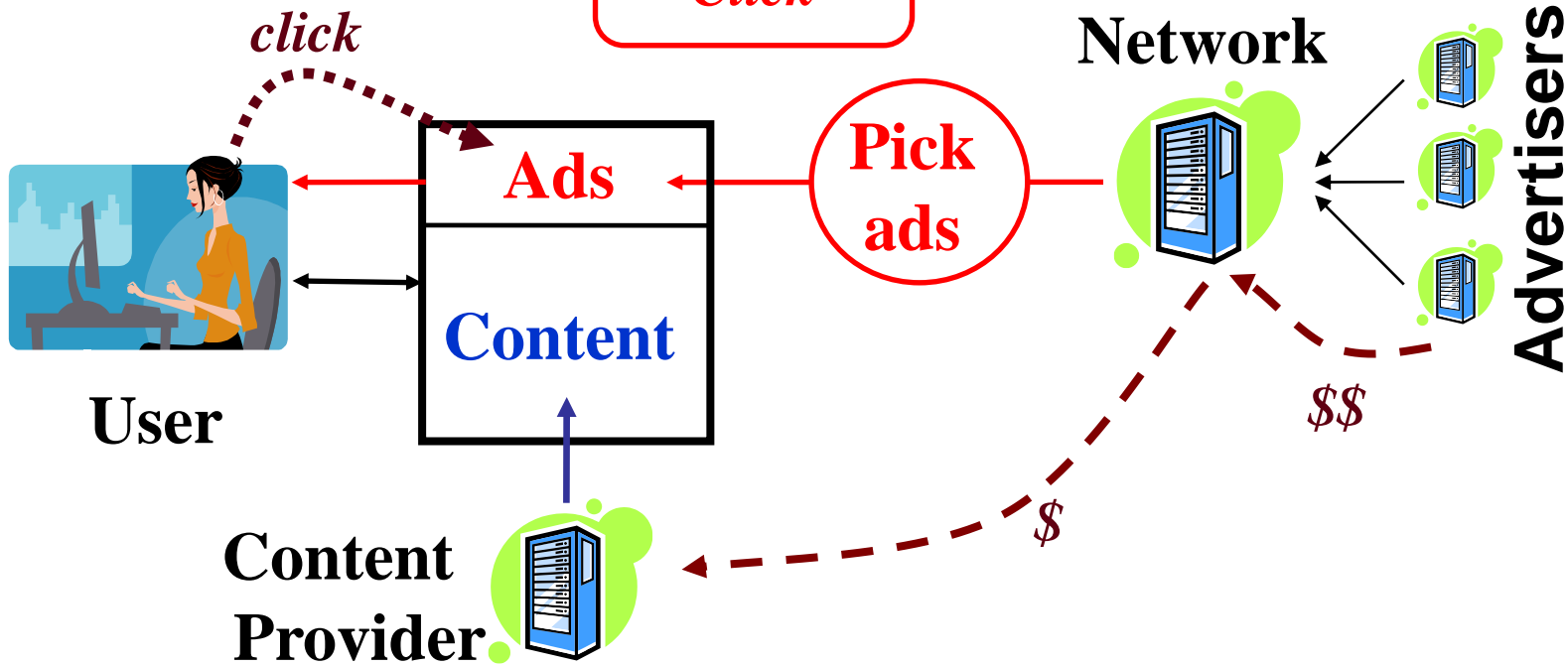
*Cost Per  
iMpression*



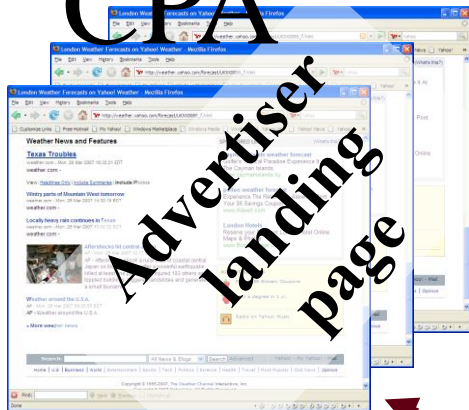
# CPC



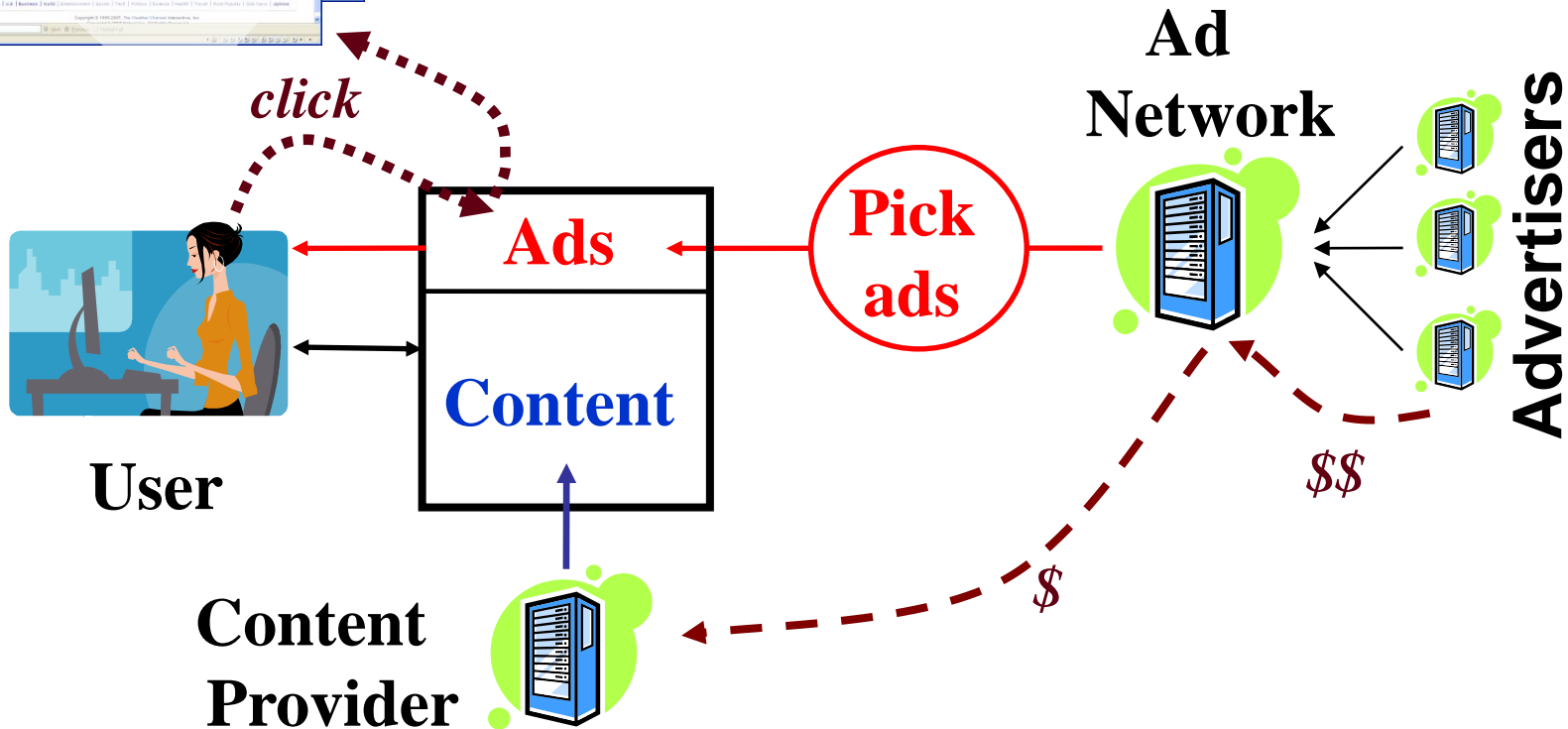
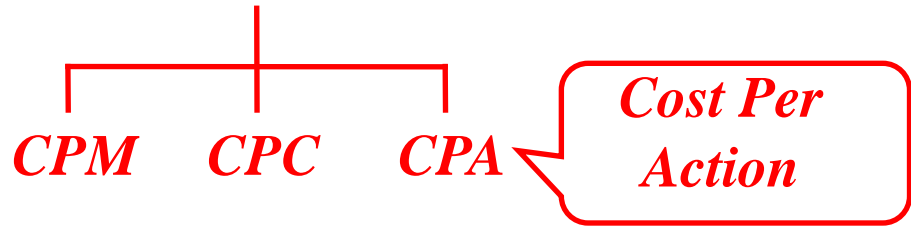
*Cost Per Click*



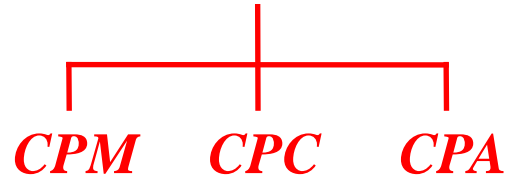
# CPA



Advertiser landing page



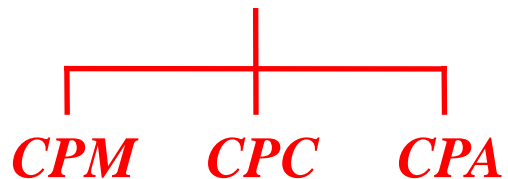
# Doanh thu - CPM



- Giả sử một quảng cáo được hiển thị N lần tại một vị trí
- CPM:  $\text{Doanh thu} = N * \text{CPM}$



# Doanh thu - CPC



- Giả sử một quảng cáo được hiển thị N lần tại một vị trí
- CPM: Doanh thu =  $N * CPM$
- CPC: Doanh thu =  $N * \text{CTR} * CPC$

*Phụ thuộc  
vào cơ chế  
đấu giá*

*Click-through Rate  
(xác suất click vào một  
quảng cáo)*

# Doanh thu - CPA

*CPM*   *CPC*   *CPA*

- Giả sử một quảng cáo được hiển thị N lần tại một vị trí
- CPM: Doanh thu =  $N * CPM$
- CPC: Doanh thu =  $N * CTR * CPC$
- CPA: Doanh thu =  $N * CTR * \text{Cony. Rate} * CPA$

*Conversion Rate*  
(*xác suất người dùng thực hiện  
một hành động khi xem trang  
quảng cáo*)

## 2. Quảng cáo trên máy tìm kiếm

The image shows a screenshot of a Yahoo! search results page for the query "recipe indian food". The browser window title is "recipe indian food - Yahoo! Search Results - Mozilla Firefox". The search bar contains "recipe indian food". The search results are displayed in a list format. Annotations are present:

- A blue box labeled "Câu truy vấn" (Query) points to the search bar.
- A red box labeled "Quảng cáo trả tiền" (Paid advertisement) points to the sponsored results on the right side of the page.
- A blue oval highlights the search bar and the "Web" tab.
- A red oval highlights the first two organic search results: "Recipe Indian Food" and "Indian Food".
- A red oval highlights the sponsored results on the right, including "Indian Food", "Recipe India Food", "Chinese Food Recipe Books on Cataloglink", and "\$19.97 Over 500 Chinese Recipes Cookbook".

Search Results

- **Recipe Indian Food**  
www.MonsterMarketplace.com - Browse and compare great deals on recipe indian food.
- **Indian Food**  
sanfrancisco.citysearch.com - Find great Indian restaurants in your area today. Search here.

1. **indian food recipe**  
indian food recipe ... Title: Indian Food Recipe. Yield: 4 Servings. Ingredients: 1 bunch ... to the echo by: Jonathan Kandell Indian Food Recipes Put ...  
recipes.chef2chef.net/recipe-archive/43/231458.shtml - 13k - Cached - More from this site

2. **Recipe Gal: Indian Foods**  
Indian Recipes from Recipe Gal's Archives ... All Food Posters: Travel Posters: Indian Recipes: Indian Breads Indian Chicken Recipes ...  
www.recipegal.com/indian - 10k - Cached - More from this site

3. **Indian Recipes, Indian Food Recipe, South Indian Recipes, Indian ...**  
indian recipes, indian food recipe, south indian Recipes, indian cooking Recipes, ... Indian Recipes, Indian Food Recipe, South Indian Recipes, Indian Cooking Recipe, ...  
www.india4world.com/indian-recipe - 17k - Cached - More from this site

4. **Paav Bhaaji - Recipe for Paav Bhaaji - Pao Bhaaji**

SPONSOR RESULTS

**Indian Food**  
Buy indian food at SHOP.COM  
Search our free shipping offers.  
www.SHOP.com

**Recipe India Food**  
Find and Compare prices on recipe india food at Smarter.com.  
www.smarter.com

**Chinese Food Recipe Books on Cataloglink**  
Find chinese food recipe books on CatalogLink.  
www.CatalogLink.com

**\$19.97 Over 500 Chinese Recipes Cookbook**  
100% Satisfaction Guaranteed  
243-Page Chinese Cookbook. Only \$19.97.

# Mô hình quảng cáo trên máy tìm kiếm

*Display*      *Content Match*      *Sponsored Search*



*Search Query*

**Text ads**

**Canon EOS Digital Camera**  
Circuit City - Official Site. Free Shipping on Orders \$24 and up.  
[www.CircuitCity.com](http://www.CircuitCity.com)

**Canon Eos Digital Camera**   
Most Digital Cameras Ship Free! 2008 Circle of  
[www.BuyDig.com](http://www.BuyDig.com)

**Canon EOS Digital Cameras**  
800+ Compact & SLR Digital Cameras. Canon EOS Digital Cameras on Sale.  
[www.NexTag.com](http://www.NexTag.com)

**Camera - EOS (SLR) Camera Systems - Digital Cameras - Canon USA ...**  
Canon offers a range of digital, compact film, and SLR cameras.  
[usa.canon.com/consumer/controller?act=ProductCatIndex1Act&... - 51k - Cached](http://usa.canon.com/consumer/controller?act=ProductCatIndex1Act&...)

**EOS DIGITAL Top Page - English**  
CANON IINC. official home page, information on digital camera  
[www.canon.com/eos-d - Cached](http://www.canon.com/eos-d)

**Canon Global (NYSE: CAJ)**  
The reformation and the expansion that healthily becomes it hangs, and Canon Inc. ... EOS-1D Series Digital SLR Cameras and the Cover for Motor Drive Coupler for EOS ...  
[www.canon.com - Cached](http://www.canon.com)

**Canon EOS SLR Camera Systems**  
Official Canon site for the EOS line of digital and film SLR cameras.  
[www.canon.com - Cached](http://www.canon.com)

**Canon EOS 5D Digital Camera Body SLR Cameras - BPhotoVideo.com**  
Find Canon EOS 5D Digital Camera (Camera Body) SLR Digital Cameras (\$2199.95) and other digital photography slr digital cameras at B&H Photo Video, world's leading ...  
[bphotovideo.com/cf.../Canon\\_0296B007\\_EOS\\_5D\\_Digital\\_Camera.html](http://bphotovideo.com/cf.../Canon_0296B007_EOS_5D_Digital_Camera.html)

**Pick ads**

*Match ads to the query*



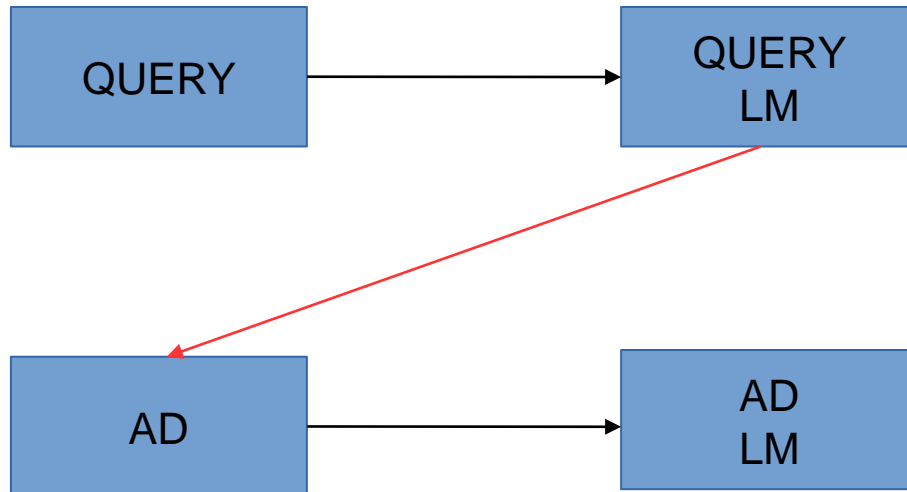
# Tối đa hóa doanh thu

- Bài toán đối với công ty quảng cáo
- Lựa chọn các quảng cáo để có doanh thu tối đa
  - Phù hợp với truy vấn
  - Chi phí quảng cáo
  - Chất lượng trang quảng cáo

# Tính điểm dựa trên nội dung

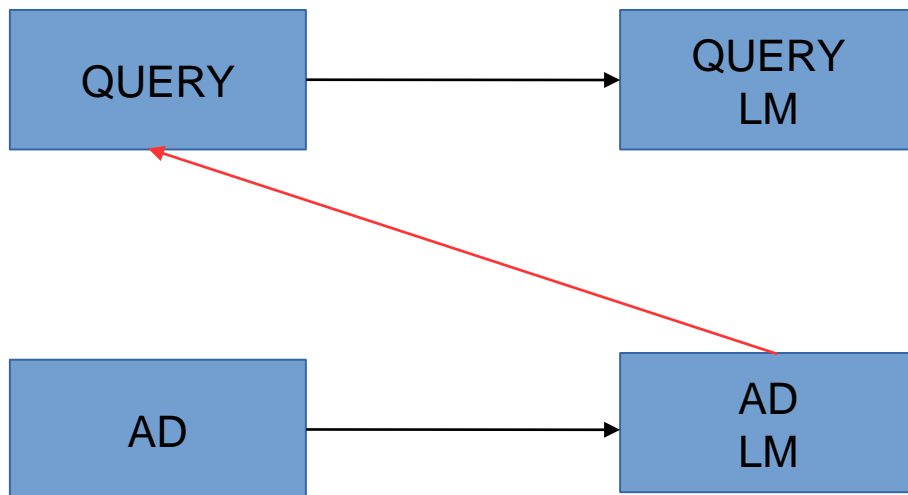
- Coi quảng cáo như một văn bản
- So sánh mức độ liên quan của câu truy vấn với quảng cáo
- Phương pháp
  - Mô hình không gian véc-tơ
  - Mô hình ngôn ngữ

# Mô hình ngôn ngữ



$P(ad|query\ LM)$

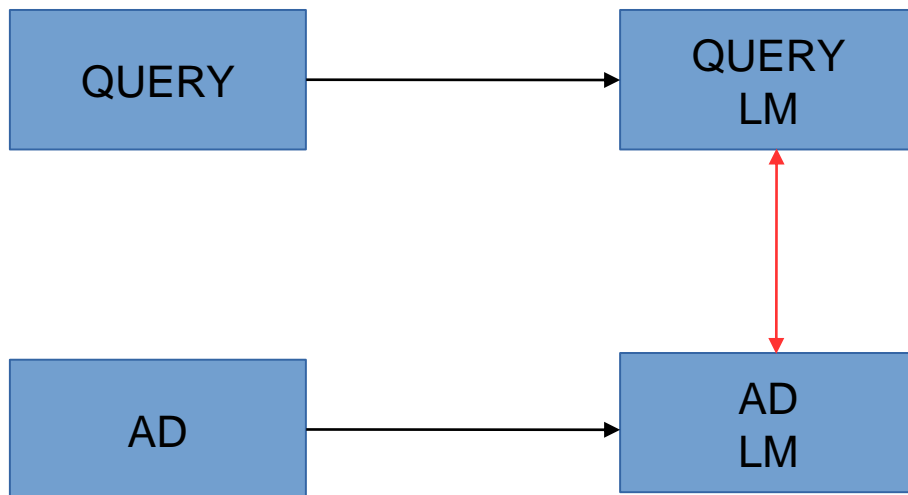
# Mô hình ngôn ngữ (tiếp)



$P(\text{query}|\text{ad LM})$



# Mô hình ngôn ngữ (tiếp)



$KL(ad\ LM; query\ LM)$

# Ưu, nhược điểm

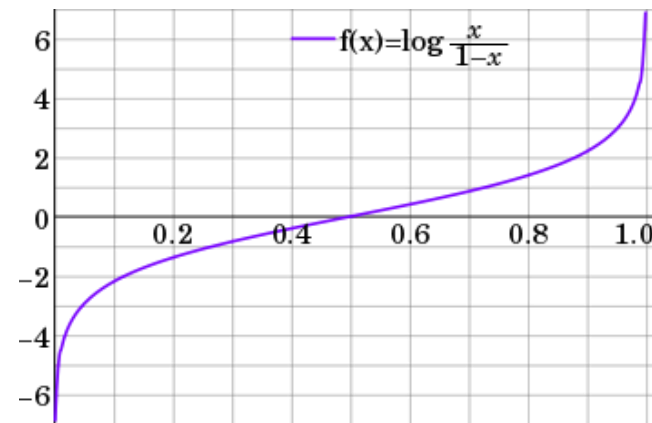
- Ưu điểm
  - Mô hình không phức tạp
  - Phù hợp với các câu truy vấn ngắn, phổ biến
- Nhược điểm:
  - Chưa xử lý được các câu truy vấn hiếm (long tail)
  - Chưa đáp ứng theo thời gian
  - Chưa tận dụng được phản hồi người dùng

# Tính điểm dựa trên phản hồi người dùng

- Tập các câu truy vấn  $Q$
- Tập các trang quảng cáo  $A$
- Với mỗi câu truy vấn  $q \in Q$  và trang quảng cáo  $a \in A$ , tính xác suất người dùng click vào trang quảng cáo  $\Pr(\text{click} | q, a)$
- Sử dụng phản hồi của người dùng để ước lượng xác suất này

# Hồi quy logistic

- Biểu diễn câu truy vấn và nội dung quảng cáo dưới dạng véc-tơ (túi từ)
- $\Pr(\text{click} | q, a) = f(\mathbf{q}, \mathbf{a}; \theta)$
- Hồi quy logistic:
  - Log-odds ( $\Pr(\text{click} | q, a)$ ) =  $\mathbf{q}' \mathbf{W} \mathbf{a}$
  - Cần phải ước lượng  $\mathbf{W}$  từ dữ liệu huấn luyện là phản hồi người dùng



from Wikipedia

# Lọc cộng tác

- Ma trận tương tác câu truy vấn, quảng cáo
- Sử dụng phản hồi ẩn của người dùng (click vào trang quảng cáo)
- Với một câu truy vấn  $q$  và trang quảng cáo  $a$ , dự đoán mức độ quan tâm của người dùng
- Lọc cộng tác
  - Dựa trên kNN
  - Biểu diễn quảng cáo theo câu truy vấn để tính độ tương tự

# Lọc cộng tác (tiếp)

Top quảng cáo  
tương tự  
quảng cáo  $a$

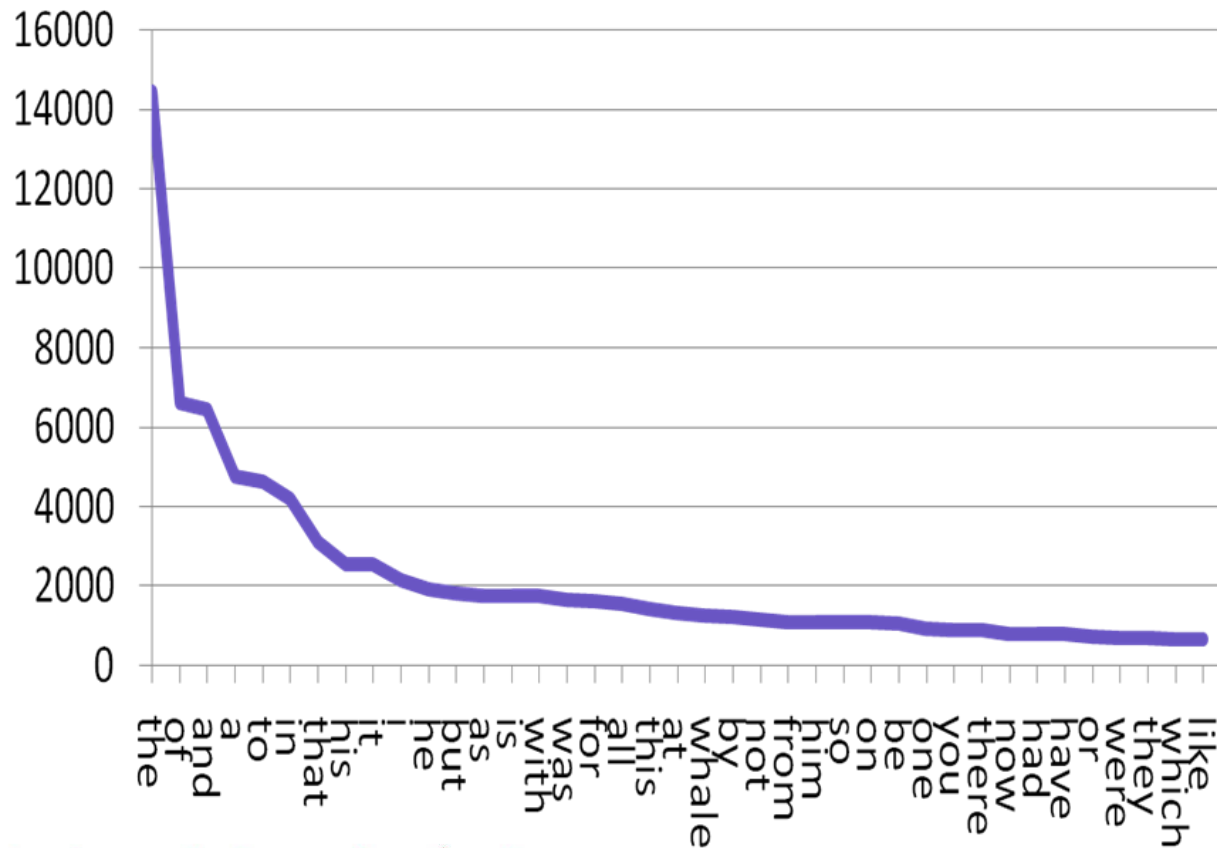
$$r_{qa} = \frac{\sum_{a' \in N(a)} \text{sim}(a, a') r_{qa'}}{\sum_{a' \in N(a)} \text{sim}(a, a')}$$

Mức độ quan tâm  
đến quảng cáo  $a$   
đ/v truy vấn  $q$

Ma trận tương tự  
quảng cáo - quảng cáo

# 3. Khai phá truy vấn

- Google: 40,000 truy vấn/s



# Đặc điểm truy vấn

- Một truy vấn chứa trung bình 2.4 từ
- 21% truy cập internet xuất phát từ máy tìm kiếm
- Phản hồi người dùng
  - 50% click vào kết quả đầu tiên
  - Người dùng hầu như chỉ sử dụng hai trang kết quả đầu tiên



# Đặc điểm truy vấn (tiếp)

- Người dùng thường tinh chỉnh câu truy vấn
- Xu hướng tìm kiếm dịch chuyển từ giải trí sang thương mại điện tử, trong đó tìm kiếm sản phẩm chiếm 1/5
- Phân bố từ vựng trên câu truy vấn và trên nội dung trang web khác nhau → thứ người dùng tìm kiếm khác với những nội dung có trên internet

# Log truy vấn

- Thông tin máy khách
- Nội dung truy vấn
- Danh sách văn bản liên quan
- Danh sách văn bản được chọn bởi người dùng

# Tiền xử lý truy vấn

- Xác định phiên truy vấn
- Lọc truy vấn tự động
- Chuẩn hóa truy vấn

# Xác định phiên truy vấn

- Phân loại cặp câu truy vấn liên tiếp vào các lớp:
  - Cùng nội dung truy vấn nhưng thay đổi phạm vi tìm kiếm
  - Tổng quát hóa câu truy vấn
  - Tinh chỉnh truy vấn nhằm đưa ra câu truy vấn chính xác hơn
  - Chi tiết hóa câu truy vấn
  - Nội dung truy vấn mới


# Lọc truy vấn tự động

- Truy vấn được sinh ra tự động sinh ra bởi bot nhằm thu thập kết quả của máy tìm kiếm
- Các câu truy vấn tự động thường có nội dung trùng lặp nhau
- Tốc độ truy vấn cao bất thường hoặc/và có tần suất truy vấn định kỳ

# Chuẩn hóa truy vấn

- Lọc từ dừng
- Chuyển thành viết thường
- Chuẩn hóa chữ số
- Stemming
- Với tiếng Việt
  - Khôi phục dấu
  - Tách từ

# Ứng dụng 1: Gợi ý truy vấn

Gmail Images  Sign in



Đại học Bách

đại học bách **khoa**

đại học bách **khoa** – Ho Chi Minh City University of Technology, College in Ho Chi Minh City, Vietnam

đại học bách **khoa** – Hanoi University of Science and Technology, University in Hanoi, Vietnam

đại học bách **khoa** – Da Nang University of Technology, University in Da Nang, Vietnam

đại học bách **khoa điểm chuẩn**

đại học bách **khoa thủ đức**

đại học bách **khoa cơ sở 2**

đại học bách **khoa lý thường kiệt**

đại học bách **khoa tô hiến thành**

đại học bách **khoa tiếng anh**

đại học bách **khoa cs2**

đại học bách **khoa paris**

đại học bách **khoa wikipedia**

Vietnam

Advertising Business About

Google Search

I'm Feeling Lucky

Privacy Terms Settings Use Google.com

Report inappropriate predictions

# Mô hình ngôn ngữ

- Học mô hình ngôn ngữ trên tập các truy vấn  
$$\operatorname{argmax}_w P(w/w_0, w_1, \dots, w_{n-1}, w_n)$$
- Yêu cầu tập truy vấn đủ lớn
- Đơn vị cơ bản của mô hình ngôn ngữ
  - Từ (yêu cầu tách từ)
  - Âm tiết
  - Bán âm tiết ('ch', 'ang')
  - Kí tự



# Mô hình ngôn ngữ n-gram

- Unigram

$$P(w) = (\text{count}(w)+1) / (\sum_w \text{count}(w')+V)$$

- Bigram

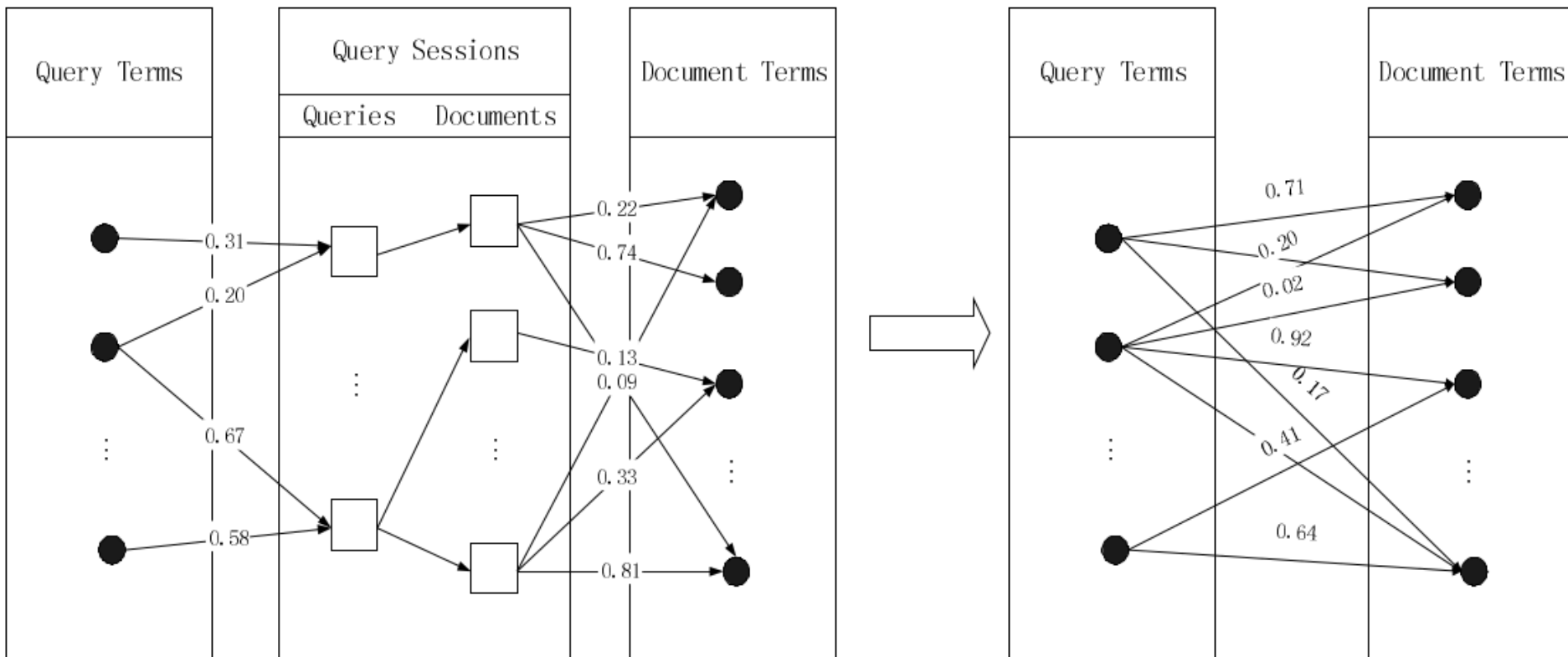
$$P(w_0, w_1) = P(w_1|w_0) * P(w_0)$$

$$P(w_1|w_0) = (\text{count}(w_0, w_1)+1) / (\sum_w \text{count}(w_0, w')+V)$$

# Ứng dụng 2: Mở rộng truy vấn

- Câu truy vấn của người dùng thường không chứa đủ thông tin
- Mở rộng truy vấn chỉ dựa trên nội dung văn bản có thể chưa đáp ứng đúng nhu cầu người dùng
  - Sử dụng phản hồi người dùng
- Giả thiết: Nếu truy vấn chứa một từ khóa dẫn đến các văn bản liên quan chứa một từ khóa khác thì nhiều khả năng hai từ khóa này có liên quan đến nhau

# Mô hình mở rộng truy vấn



# Mô hình mở rộng truy vấn (tiếp)

$$\begin{aligned} P(w_j^{(d)} | w_i^{(q)}) &= \frac{P(w_j^{(d)}, w_i^{(q)})}{P(w_i^{(q)})} \\ &= \frac{\sum_{\forall D_k \in S} P(w_j^{(d)}, w_i^{(q)}, D_k)}{P(w_i^{(q)})} \\ &= \frac{\sum_{\forall D_k \in S} P(w_j^{(d)} | w_i^{(q)}, D_k) \times P(w_i^{(q)}, D_k)}{P(w_i^{(q)})} \end{aligned}$$

# Mô hình mở rộng truy vấn (tiếp)

$$P(w_j^{(d)} | w_i^{(q)}, D_k) = P(w_j^{(d)} | D_k)$$

$$\begin{aligned} P(w_j^{(d)} | w_i^{(q)}) &= \frac{\sum_{\forall D_k \in \mathcal{S}} P(w_j^{(d)} | D_k) \times P(D_k | w_i^{(q)}) \times P(w_i^{(q)})}{P(w_i^{(q)})} \\ &= \sum_{\forall D_k \in \mathcal{S}} P(w_j^{(d)} | D_k) \times P(D_k | w_i^{(q)}) \end{aligned}$$

$P(w_j^{(d)} | D_k)$  : Xác suất xuất hiện  $w_j^{(d)}$  với điều kiện  $D_k$  được chọn

$P(D_k | w_i^{(q)})$  : Xác suất  $D_k$  được chọn nếu  $w_i^{(q)}$  xuất hiện trong câu truy vấn

# Mô hình mở rộng truy vấn (tiếp)

$$P(D_k | w_i^{(q)}) = \frac{f_{ik}^{(q)}(w_i^{(q)}, D_k)}{f^{(q)}(w_i^{(q)})}$$

$$P(w_j^{(d)} | D_k) = \frac{W_{jk}^{(d)}}{\max_{\forall t \in D_k} (W_{tk}^{(d)})}$$

$$P(w_j^{(d)} | w_i^{(q)}) = \sum_{\forall D_k \in S} (P(w_j^{(d)} | D_k) \times \frac{f_{ik}^{(q)}(w_i^{(q)}, D_k)}{f^{(q)}(w_i^{(q)})})$$

$f_{ik}^{(q)}(w_i^{(q)}, D_k)$  : Số phiên truy vấn mà câu truy vấn chứa  $w_i^{(q)}$  mà  $D_k$  được chọn

$f^{(q)}(w_i^{(q)})$  : Số phiên truy vấn mà câu truy vấn chứa  $w_i^{(q)}$

$W_{jk}^{(d)}$  : Trọng số của  $w_j^{(d)}$  trong văn bản  $D_k$

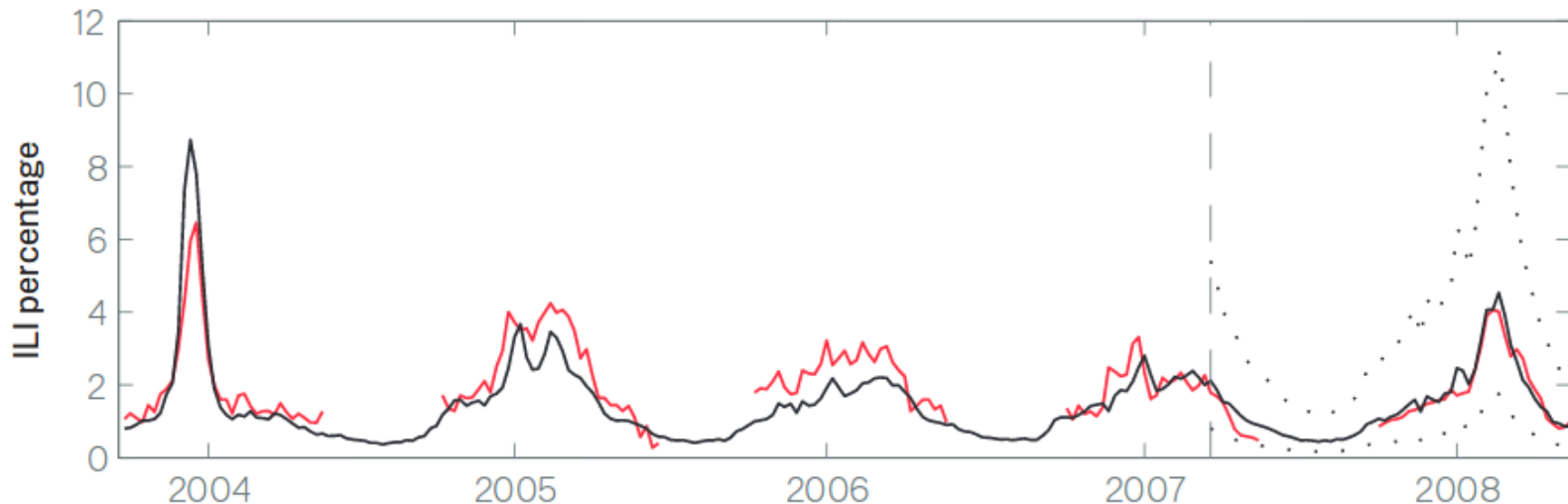
# Mô hình mở rộng truy vấn (tiếp)

$$CoWeight_Q(w_j^{(d)}) = \ln\left(\frac{1}{\sum_{w_t^{(q)} \in Q} (P(w_j^{(d)} | w_t^{(q)})) + 1}\right)$$

1. Trích xuất các term trong câu truy vấn Q
2. Tìm tất cả các văn bản liên quan đến bất kỳ term nào
3. Với mỗi term trong mỗi văn bản, sử dụng công thức trên xác định mức độ liên quan đến câu truy vấn Q
4. Sử dụng n term có điểm cao nhất để tạo thành câu truy vấn Q'
5. Tìm kiếm với câu truy vấn Q'

# Ứng dụng 3: Cảnh báo dịch bệnh

- <https://www.google.org/flutrends>
- Dựa trên các truy vấn liên quan
- Số người tìm thông tin về bệnh tỉ lệ thuận với số người bị bệnh







25 YEARS ANNIVERSARY  
**SOICT**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

**Thank you for  
your attentions!**

