



IT4853

# Tìm kiếm và trình diễn thông tin

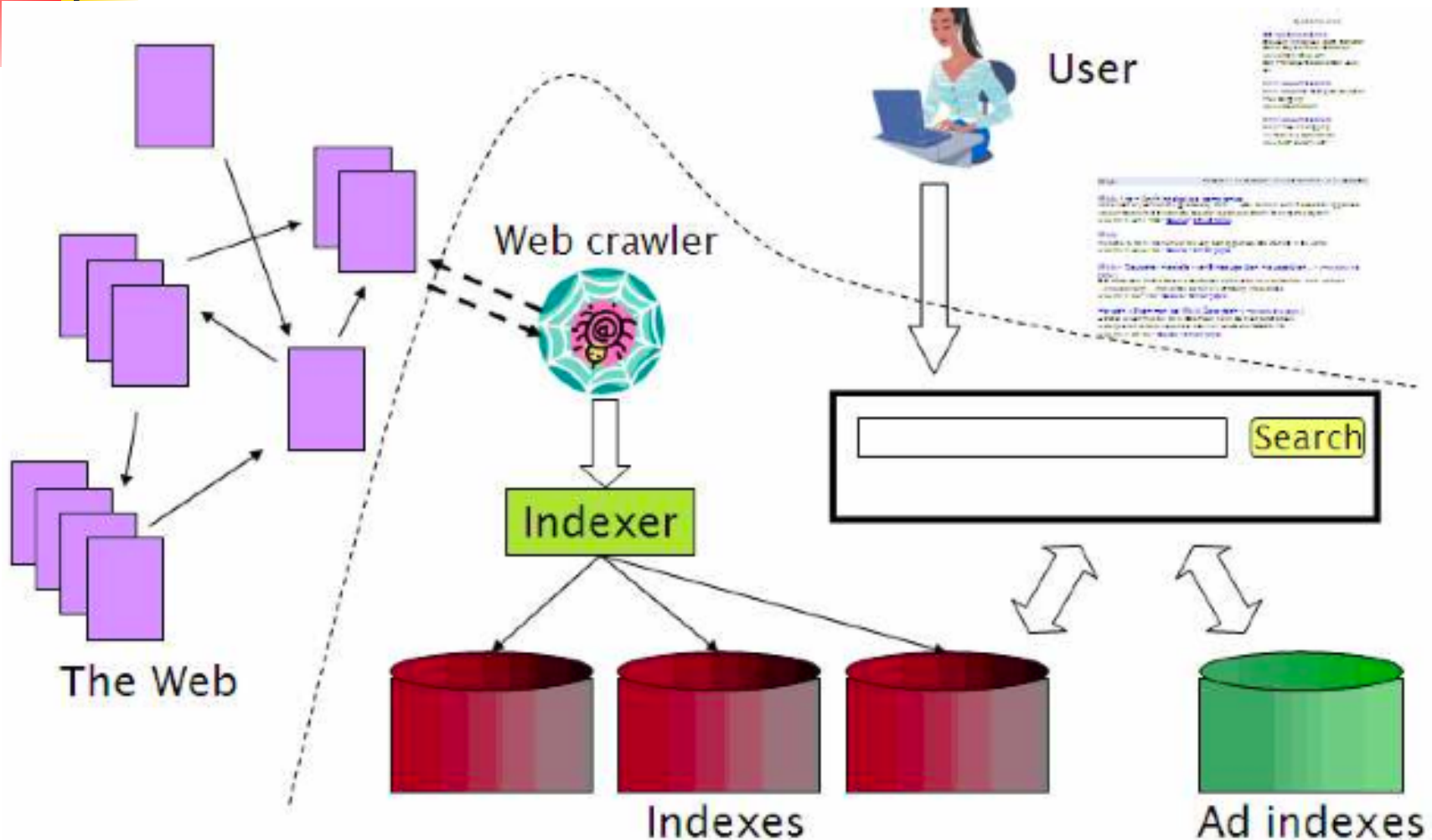
---

Bài 17. Quảng cáo và SPAM

IIR. C19. Web search basics

*Bộ môn Hệ thống thông tin  
Viện CNTT & TT*

# Tổng quan công cụ tìm kiếm trên Web





# Nội dung chính

---

- Quảng cáo trong tìm kiếm
- Tối ưu hóa nội dung hướng tìm kiếm

# Mô hình Goto (1996)

www.goto.com/d/search?;sessionid=1AQ42T4AAAH0R5QFIEF3QFUQ?type=home&tr=1&Keywords=Wilmington

Wilmington real estate.

Access 75% of all users now!  
Premium Listings reach 75% of all  
Internet users. [Sign up](#) for Premium  
Listings today!

1. [Wilmington Real Estate - Buddy Blake](#)  
Wilmington's information and real estate guide. This is your on  
anything to do with Wilmington.  
[www.buddyblake.com](#) (Cost to advertiser: **\$0.38**)
2. [Coldwell Banker Sea Coast Realty](#)  
Wilmington's number one real estate company.  
[www.cbseacoast.com](#) (Cost to advertiser: **\$0.37**)
3. [Wilmington, NC Real Estate Becky Bullard](#)  
Everything you need to know about buying or selling a home c  
on my Web site!  
[www.iwwc.net](#) (Cost to advertiser: **\$0.35**)



## Mô hình Goto (1996)

---

- Quảng cáo được xếp hạng theo giá
  - Tối đa thu nhập cho Goto.
- Không phân biệt quảng cáo với kết quả tìm kiếm. Chỉ hiển thị một danh sách kết quả duy nhất!
- Không xếp hạng theo sự phù hợp, . . .
  - Bất tiện cho người dùng.

# Mô hình Google (2000/2001)

Web Images Maps News Shopping Gmail more Sign In

Google discount broker Search Advanced Search Preferences

Web Results 1 - 10 of about 807,000 for discount broker [definition]. (0.12 seconds)

**Discount Broker Reviews**  
Information on online discount brokers emphasizing rates, charges, and customer comments and complaints.  
[www.broker-reviews.us/](http://www.broker-reviews.us/) - 94k - Cached - Similar pages

**Discount Broker Rankings (2008 Broker Survey) at SmartMoney.com**  
Discount Brokers. Rank/ Brokerage/ Minimum to Open Account, Comments, Standard Commission\*, Reduced Commission, Account Fee Per Year (How to Avoid), Avg. ...  
[www.smartmoney.com/brokers/index.cfm?story=2004-discount-table](http://www.smartmoney.com/brokers/index.cfm?story=2004-discount-table) - 121k - Cached - Similar pages

**Stock Brokers | Discount Brokers | Online Brokers**  
Most Recommended. Top 5 Brokers headlines. 10. Don't Pay Your Broker for Free Funds  
May 15 at 3:39 PM. 5. Don't Discount the Discounters Apr 18 at 2:41 PM ...  
[www.foo.com/investing/brokers/index.aspx](http://www.foo.com/investing/brokers/index.aspx) - 44k - Cached - Similar pages

**Discount Broker**  
Discount Broker - Definition of Discount Broker on Investopedia - A stockbroker who carries out buy and sell orders at a reduced commission compared to a ...  
[www.investopedia.com/terms/d/discountbroker.asp](http://www.investopedia.com/terms/d/discountbroker.asp) - 31k - Cached - Similar pages

**Discount Brokerage and Online Trading for Smart Stock Market ...**  
Online stock broker SogoTrade offers the best in discount brokerage investing. Get stock market quotes from this Internet stock trading company.  
[www.sogotrade.com/](http://www.sogotrade.com/) - 39k - Cached - Similar pages

**15 questions to ask discount brokers - MSN Money**  
Jan 11, 2004 ... If you're not big on hand-holding when it comes to investing, a discount broker can be an economical way to go. Just be sure to ask these ...  
[moneycentral.msn.com/content/Investing/StartInvesting/P68171.asp](http://moneycentral.msn.com/content/Investing/StartInvesting/P68171.asp) - 34k - Cached - Similar pages

**Sponsored Links**

**Rated #1 Online Broker**  
No Minimums. No Inactivity Fee  
Transfer to Firsttrade for Free!  
[www.firsttrade.com](http://www.firsttrade.com)

**Discount Broker**  
Commission free trades for 30 days.  
No maintenance fees. Sign up now.  
TDAMERITRADE.com

**TradeKing - Online Broker**  
\$4.95 per Trade, Market or Limit  
SmartMoney Top Discount Broker 2001  
[www.TradeKing.com](http://www.TradeKing.com)

**Scottrade Brokerage**  
\$7 Trades, No Share Limit. In-Depth Research. Start Trading Online Now!  
[www.Scottrade.com](http://www.Scottrade.com)

**Stock trades \$1.99-\$3**  
100 free trades, up to \$100 back for transfer costs, \$500 minimum  
[www.sogotrade.com](http://www.sogotrade.com)

**\$3.95 Online Stock Trades**  
Market/Limit Orders, No Share Limit and No Inactivity Fees  
[www.Marsco.com](http://www.Marsco.com)

**INGDIRECT | ShareBuilder**  
No Commission on Trades. No Inactivity Fees

SogoTrade xuất hiện trong kết quả tìm kiếm

SogoTrade trong mục quảng cáo

Công cụ tìm kiếm có ưu tiên nội dung được quảng cáo trong xếp hạng?

Đa phần các công cụ tìm kiếm hàng đầu đều tuyên bố không.



# Xếp hạng quảng cáo

---

- **Đấu giá:**
  - Nhà quảng cáo đặt giá cho từ khóa;
  - Bất kỳ ai cũng có thể tham gia đặt giá cho bất kỳ từ khóa nào.
- Nhà quảng cáo chỉ phải trả tiền khi ai đó mở quảng cáo.

Xếp hạng quảng cáo, và giá phải trả khi người dùng mở quảng cáo?



## Xếp hạng quảng cáo (2)

---

- Vấn đề: Nếu xếp hạng theo giá được trả như trong Goto, truy vấn và quảng cáo có thể không khớp về nội dung;
- Cải tiến: Xếp hạng dựa trên giá và sự phù hợp
  - Tham số chính để đánh giá sự phù hợp là tỉ lệ mở liên kết
  - Các tiêu chí xếp hạng khác: địa điểm, thời gian, chất lượng và tốc độ tải trang web, v.v.





## Xếp hạng quảng cáo (3)

---

- Kết quả: Quảng cáo kém phù hợp sẽ bị xếp hạng thấp hơn
  - Mong muốn: Người dùng luôn tìm được thông tin hữu ích.
  - Sự hài lòng của người dùng là cực đại => doanh thu của hệ thống là cực đại.

# Mô hình đấu giá với giá thứ hai của Google

advertiser	bid	CTR	ad rank	rank	paid
A	\$4.00	0.01	0.04	4	(minimum)
B	\$3.00	0.03	0.09	2	\$2.68
C	\$2.00	0.06	0.12	1	\$1.51
D	\$1.00	0.08	0.08	3	\$0.51

- **bid**: giá tối đa được trả cho một lần mở quảng cáo
- **CTR**: tỉ lệ bấm quảng cáo: (số lần bấm quảng cáo / số lần hiển thị quảng cáo). **CTR thể hiện sự phù hợp**.
- **Xếp hạng quảng cáo**:  $\text{bid} \times \text{CTR}$ : tổng hợp (i) giá được trả (ii) mức độ phù hợp của quảng cáo
- **rank**: xếp hạng
- **paid**: giá thực trả

# Mô hình đấu giá với giá thứ hai của Google (2)

advertiser	bid	CTR	ad rank	rank	paid
A	\$4.00	0.01	0.04	4	(minimum)
B	\$3.00	0.03	0.09	2	\$2.68
C	\$2.00	0.06	0.12	1	\$1.51
D	\$1.00	0.08	0.08	3	\$0.51

Giá thứ hai là chi phí tối thiểu để duy trì vị trí của họ trong đấu giá và thêm 1 xu (0.01\$).

$$\text{price}_1 \times \text{CTR}_1 = \text{bid}_2 \times \text{CTR}_2 \text{ (kết quả là rank}_1 = \text{rank}_2)$$

$$\text{price}_1 = \text{bid}_2 \times \text{CTR}_2 / \text{CTR}_1$$

$$\text{price}_1 = \text{bid}_2 \times \text{CTR}_2 / \text{CTR}_1 = 3.00 \times 0.03 / 0.06 = 1.50$$

$$\text{price}_2 = \text{bid}_3 \times \text{CTR}_3 / \text{CTR}_2 = 1.00 \times 0.08 / 0.03 = 2.67$$

$$\text{price}_3 = \text{bid}_4 \times \text{CTR}_4 / \text{CTR}_3 = 4.00 \times 0.01 / 0.08 = 0.50$$

$$\text{paid} = \text{price} + 0.01$$



# Mô hình đấu giá theo giá thứ hai của Google (2)

---

- Google:
  - Thu thêm **một xu** từ mỗi quảng cáo sẽ **bổ xung hàng tỉ** lợi nhuận cho Google.
  - Mô hình quảng cáo có thể là lĩnh vực nghiên cứu quan trọng bậc nhất đối với máy tìm kiếm (computational advertising).

# Những từ khóa được định giá cao

Theo <http://www.cwire.org/highest-paying-search-terms/>

- \$69.1 mesothelioma treatment options
- \$65.9 personal injury lawyer michigan
- \$62.6 student loans consolidation
- \$61.4 car accident attorney los angeles
- \$59.4 online car insurance quotes
- \$59.4 arizona dui lawyer
- \$46.4 asbestos cancer
- \$40.1 home equity line of credit
- \$39.8 life insurance quotes
- \$39.2 refinancing
- \$38.7 equity line of credit
- \$38.0 lasik eye surgery new york city
- \$37.0 2nd mortgage
- \$35.9 free car insurance quote



# Quảng cáo trong tìm kiếm: ba bên đều có lợi?

---

- Công cụ tìm kiếm thu lợi mỗi khi ai đó mở quảng cáo.
- Người dùng chỉ mở quảng cáo nếu họ thực sự quan tâm.
  - Công cụ tìm kiếm sẽ *phạt* những quảng cáo không phù hợp.
  - ... để người dùng hài lòng với những gì họ tìm thấy sau khi mở quảng cáo.
- Nhà quảng cáo tìm thấy những khách hàng mới với chi phí tối thiểu.



Có thể đánh lừa hệ thống?

---



# Có thể đánh lừa hệ thống? Đầu cơ từ khóa

---

- Mua một từ khóa trên Google
- Sau đó chuyển hướng lưu lượng tới bên thứ ba (đối tác trả giá cao hơn giá bạn phải trả Google)
  - Ví dụ, chuyển tới một trang chứa rất nhiều quảng cáo
- v.v.





# Có thể đánh lừa hệ thống? Vi phạm thương hiệu

---

- Ví dụ: Geico
- Một thời gian trong năm 2005: Từ khóa tìm kiếm “Geico” trên Google bị mua bởi những đối thủ.
  - Geico đã thất bại trong trường hợp này ở Mỹ.
- Louis Vuitton cũng đã gặp tình huống tương tự ở châu Âu.
- Tham khảo [http://google.com/tm\\_complaint.html](http://google.com/tm_complaint.html)



# Nội dung chính

---

- Quảng cáo trong tìm kiếm
- Tối ưu hóa nội dung hướng tìm kiếm



# Tối ưu nội dung hướng tìm kiếm

---

- SEO - Search Engine Optimization
- Tùy chỉnh nội dung trang web để chiếm thứ hạng cao trong tìm kiếm
- Mục đích:
  - Thương mại, chính trị, tôn giáo
  - Tài trợ bởi quỹ quảng cáo
- Thực hiện dưới dạng dịch vụ bởi các công ty, webmasters, search engine optimizers, v.v.

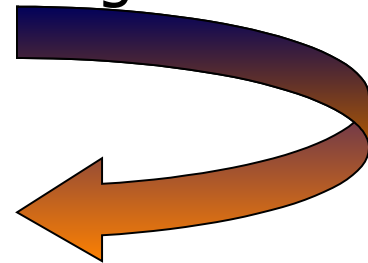


# Kỹ thuật đơn giản nhất

---

- Kích tf/idf
  - Đối với những mô hình xếp hạng cổ điển, những trang hàng đầu chứa nhiều từ truy vấn.
- SEOs sử dụng các từ được lựa chọn với mật độ dày đặc
  - Thông thường các từ được lặp lại có màu trùng màu nền: Vẫn được đưa vào chỉ mục, nhưng vô hình với người dùng

Mật độ từ là một tín hiệu xếp hạng kém tin cậy





## Các phương pháp kích từ khóa

---

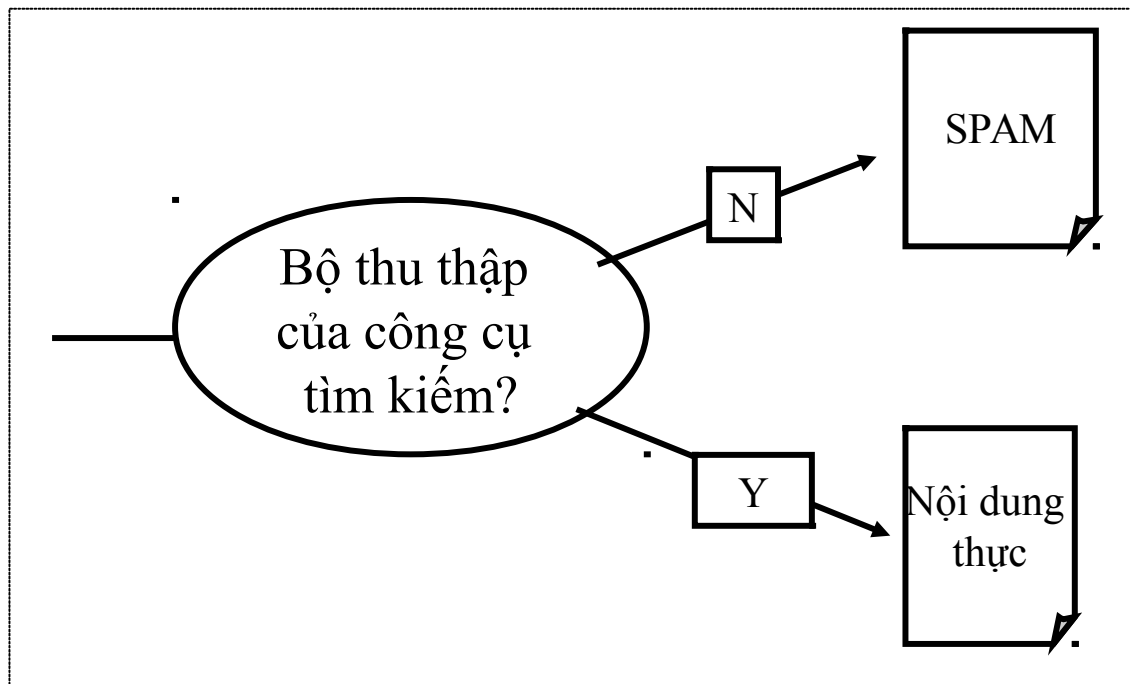
- Sử dụng thẻ Meta
- Văn bản ẩn, màu sắc, CSS, v.v.

**Meta-Tags =**

“... London hotels, hotel, holiday inn, hilton, discount, booking, reservation, sex, mp3, britney spears, viagra, ...”

# Tàng hình

- Cung cấp nội dung giả cho bộ thu thập
- DNS: Hoán đổi IP.





## Những kỹ thuật khác

---

- Doorway pages
  - Trang với nội dung tối ưu cho một từ khóa, điều hướng tới các trang đích
- Link pages
  - Hidden links,
  - Domain flooding.
- Robots
  - Luồng truy vấn ảo



## Đối phó với spam

---

- Các tín hiệu chất lượng:
  - Dựa trên đánh giá của các tác giả và người dùng;
- Chính sách tiếp nhận URL:
  - Phát hiện robot
- Hạn chế từ khóa meta
- Phân tích liên kết
- Phát hiện SPAM bằng học máy
- v.v.





# Tài liệu tham khảo

---

- Chính sách của các công cụ tìm kiếm trên Web đối với SEO (chấp nhận/từ chối).
  - <http://help.yahoo.com/help/us/ysearch/index.html>
  - <http://www.google.com/intl/en/webmasters/>
- Quảng cáo trong tìm kiếm: Cuộc chiến kỹ thuật giữa SEOs và công cụ tìm kiếm
  - <http://airweb.cse.lehigh.edu/>



## Bài tập 22.1

---

Hệ thống Goto xếp hạng văn bản theo bid: các quảng cáo có bid cao nhất đứng đầu danh sách. Điều gì sẽ xảy ra nếu quảng cáo có bid cao nhất không phù hợp với truy vấn? Vì sao nhà quảng cáo muốn trả bid cao cho quảng cáo không phù theo cách này?



## Bài tập 22.2

---

Giả sử chúng ta có thêm CTR được khai thác từ dữ liệu lịch sử của hệ thống quảng cáo. Hãy đề xuất một thay đổi cho phương pháp xếp hạng của Goto để giải quyết vấn đề của bài tập 22.1.

