Classified Displays of Web Search Results
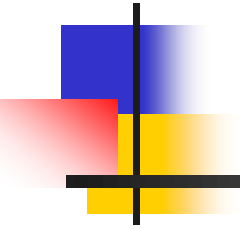
Susan Dumais, Ed Cutrell (Microsoft Research)

Hao Chen (UC Berkeley)

Organizing Search Results
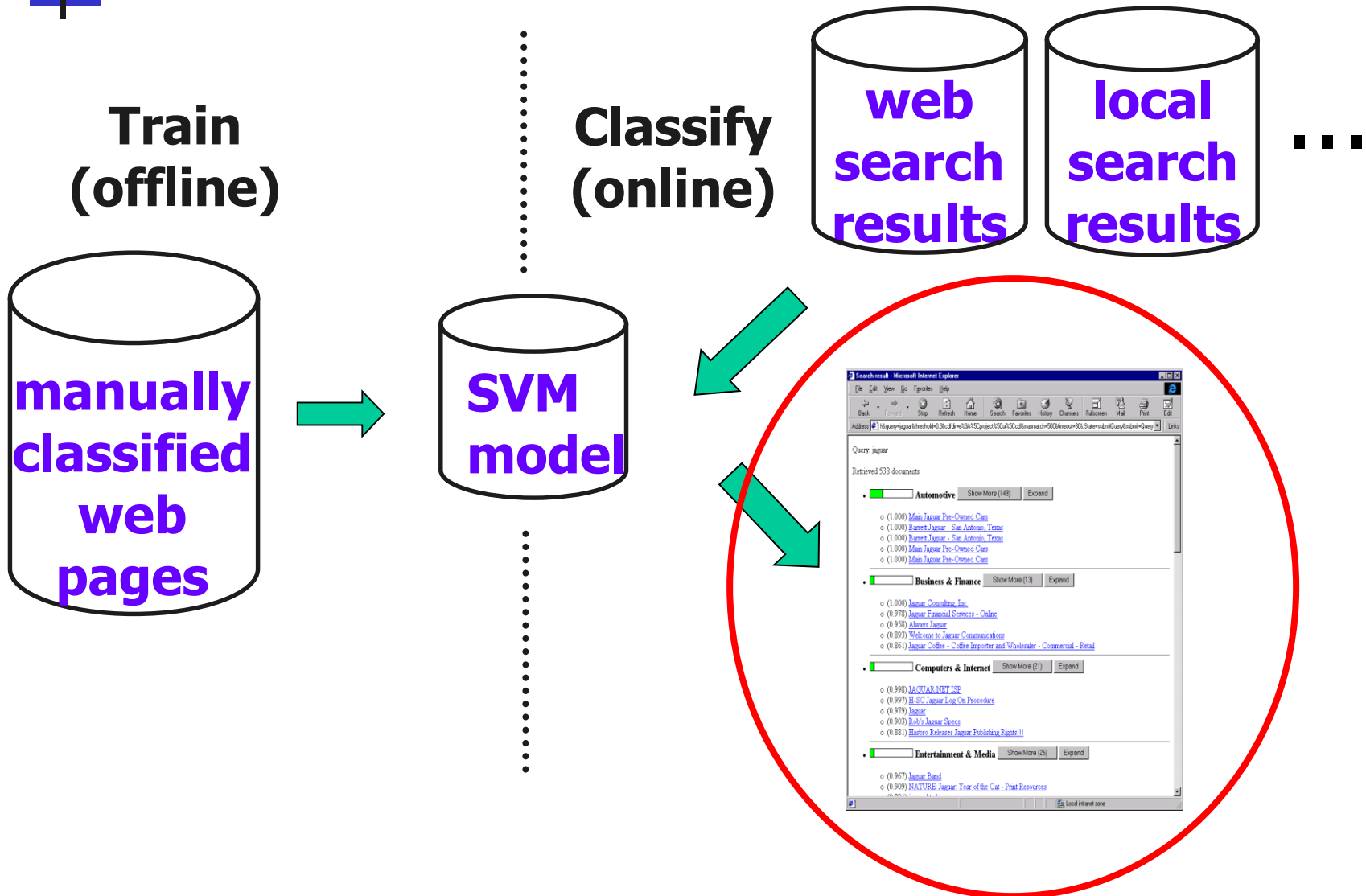
- Going beyond the ranked list

- Two natural approaches
  - Clustering – discover structure
    - e.g., Zamir & Etzioni, Hearst & Pedersen, Maarek, MSR
    - … But, slowish, and difficulties in understanding and labeling categories
  - Classification – map results to known structure
    - e.g., Chakrabarti et al., Stata et al., Northern Light
    - Fast, and easily understood category labels
    - Extends manually compiled directory structure

Classifying Search Results

- Combines the advantages of:
  - Manually compiled directory structure
  - Broad coverage from search engines
- Two main system components:
  - *Classification models*
    - Trained on manually classified pages (offline)
    - Classify search results on-the-fly (online)
  - *UI for integrating search and structure*
    - Present results in a useful and usable form

SWISH Architecture

Online Classification

Query: "jaguar"

Standard List Results ————→ Category Tree (SWISH)

automobile
->

automobile
->

animal
->

sports
...

Interface Design Considerations

- Problems
  - Large amount of information to display
    - Search results (focus)
    - Category structure (context)
  - Limited screen real estate
- Solutions
  - Information overlay
  - Distilled information display

Information Overlay

- Use tooltips (hover_text) to show
  - Summaries of web pages
  - Ca...

```
                          ->
Automoti                 Buy or Sell a Car
                         Chat
                         Shows & Museums
                         Finance & Insurance
○ (99) H.D. Rog          Trucks & Tractors    MG Triumph Renault Peugeot Ferrari Fiat
○ (94) Jaguar Clu        Magazines & Books
○ (85) Bauer Ja          Vintage & Classic    ign sports cars and Jaguar automob
○ (84) A&L Lux           Maintenance & Repair
                         Makes, Models & Clubs    dealership for
                         Motorcycles              owned
Entertainm               New Car Showrooms        es with
                         Off-Road, 4X4 & RVs
○ (83) Tom's Col         Other Auto Interests
```

Expansion of Category Structure

- ▮☐ **Automotive** [MainCateg]

  - ○ ▮☐ **Maintenance & Repair** [More (7)]

    - ■ (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
    - ■ (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob

  - ○ ▮☐ **Buy or Sell a Car** [More (6)]

    - ■ (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
    - ■ (84) A&L Luxury Car Center - Jaguar Main Page

  - ○ ▮☐ **Vintage & Classic** [More (2)]

    - ■ (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B

  - ○ ▮☐ **Makes, Models & Clubs** [More (1)]

    - ■ (94) Jaguar Club of Florida

Distilled Information Display

- Only show categories with matches
- Order categories by number of matches
- Show only best matches within each initially (others available with expansion)

Automotive [SubCateg] [More (16)]

- (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat
- (94) Jaguar Club of Florida
- (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
- (84) A&L Luxury Car Center - Jaguar Main Page

Entertainment & Media [SubCateg] [More (14)]

- (83) Tom's Collection of Jaguar Mark II Photos

Chen & Dumais (CHI'2000) Interface Conditions

Category Interface

SWISH Search result - Microsoft Internet Explorer

Query: Jaguar
Retrieved 100 documents

- ▮▭ **Computers & Internet** [SubCateg] [More (18)]
  - (95) Clan Jaguar Quake & Quake 2 Clan
  - (90) Atari Jaguar-System
  - (79) Atari - Jaguar Order Form
  - (69) Jaguar XK8 Screen Saver

- ▮▭ **Automotive** [SubCateg] [More (16)]
  - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - (94) Jaguar Club of Florida
  - (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
  - (84) A&L Luxury Car Center - Jaguar Main Page

- ▮▭ **Entertainment & Media** [SubCateg] [More (14)]
  - (83) Tom's Collection of Jaguar Mark II Photos
  - (82) MacJag's Jaguar Page
  - (80) The Jaguar Photo Gallery

- ▮▭ **Travel & Vacations** [SubCateg] [More (4)]
  - (92) Classic Car Source -- Welsh Jaguar Classic Car Museum

- ▮▭ **Business & Finance** [SubCateg] [More (2)]
  - (66) Jaguar Consulting, Inc.

- ▮▭ **Shopping & Services** [SubCateg] [More (2)]

Local intranet zone

List Interface

SWISH Search result - Microsoft Internet Explorer

Query: Jaguar
Retrieved 100 documents

- ▮▮ **Search Results** [More (80)]
  - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - (98) Jaguar Clubs of North America
  - (97) Welcome to Jaguar
  - (96) Terry's Jaguar Parts
  - (95) Clan Jaguar Quake & Quake 2 Clan
  - (94) Jaguar Club of Florida
  - (93) Jaguar Daimler Heritage Trust
  - (92) Classic Car Source -- Welsh Jaguar Classic Car Museum
  - (91) A.C. Imports Jaguar - Palatine IL
  - (90) Atari Jaguar-System
  - (89) Jaguar Underground Dox
  - (88) Jaguar Owners Web Ring
  - (87) Jaguar, new cars, used cars, and accessories
  - (86) Jaguar Sovereign
  - (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
  - (84) A&L Luxury Car Center - Jaguar Main Page
  - (83) Tom's Collection of Jaguar Mark II Photos
  - (82) MacJag's Jaguar Page
  - (81) Welcome to Jaguar Homes!
  - (80) The Jaguar Photo Gallery

Local intranet zone

Experimental Setup

User Study

- Participants:
  - 18 intermediate Web users
- Tasks
  - 30 search tasks
    e.g., *"Find home page for Seattle Art Museum"*
  - Search terms are fixed for each task
- Experimental Design
  - Category/List – within subjects
    - 15 search tasks with each interface
  - Order (Category/List First) – counterbalanced between subjects
- Both Subjective and Objective Measures

Subjective Results

- 7-point rating scale (1=disagree; 7=agree)
- Questions:

Question	Category	List	significance
It was easy to use this software.	6.4	3.9	p<.001
I liked using this software	6.7	4.3	p<.001
I prefer this to my usual Web Search engine	6.4	4.3	p<.001
It was easy to get a good sense of the range of alternatives	6.4	4.2	p<.001
I was confident that I could find information if it was there.	6.3	4.4	p<.001
The "More" button was useful	6.5	6.1	n.s.
The display of summaries was useful	6.5	6.4	n.s.

Use of Interface Features

- Average Number of Uses of Feature per Task

Interface Features	Category	List	significance
Expansing / Collapsing Structure	0.78	0.48	p<.003
Viewing Summaries in Tooltips	2.99	4.60	p<.001
Viewing Web Pages	1.23	1.41	p<.053

Search Time
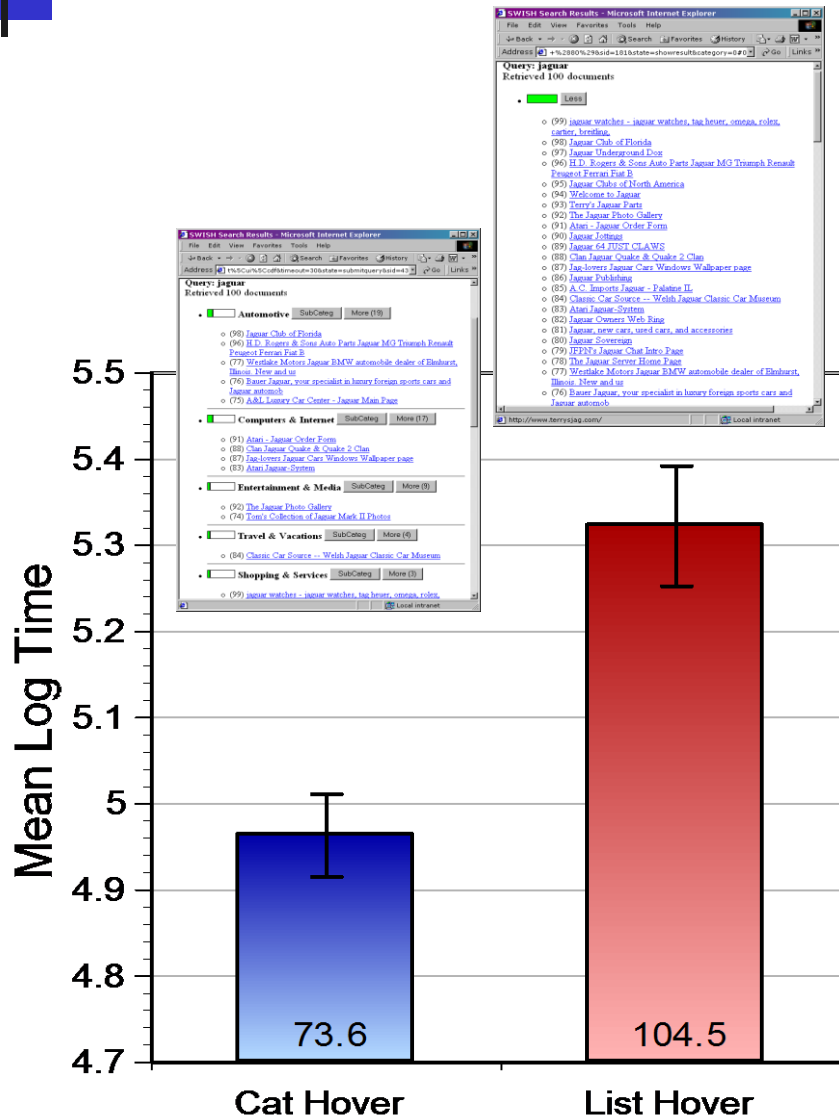
Category: 73 secs
List: 104 secs
p < .002

43% faster with Category interface

Further Category/List UI Experiments

- Enhance List
  - Inline Summary
  - Add Category Labels to ... (e)
- Decompose Category ...
  - Remove Page Titles (Br...
  - Remove Category Labels
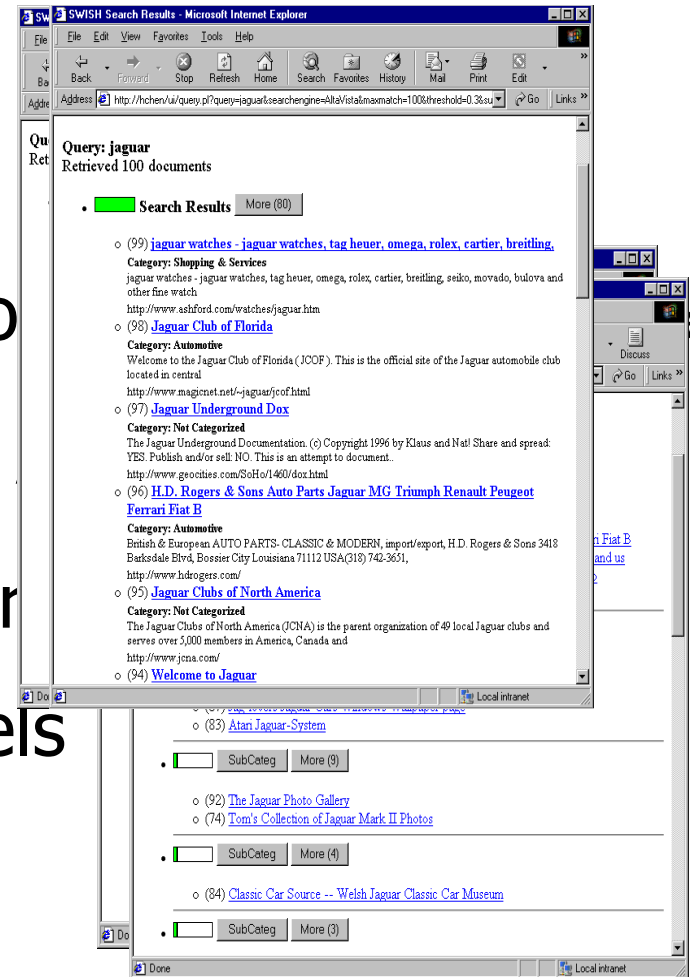
List/Categ: Hover Text vs. Inline
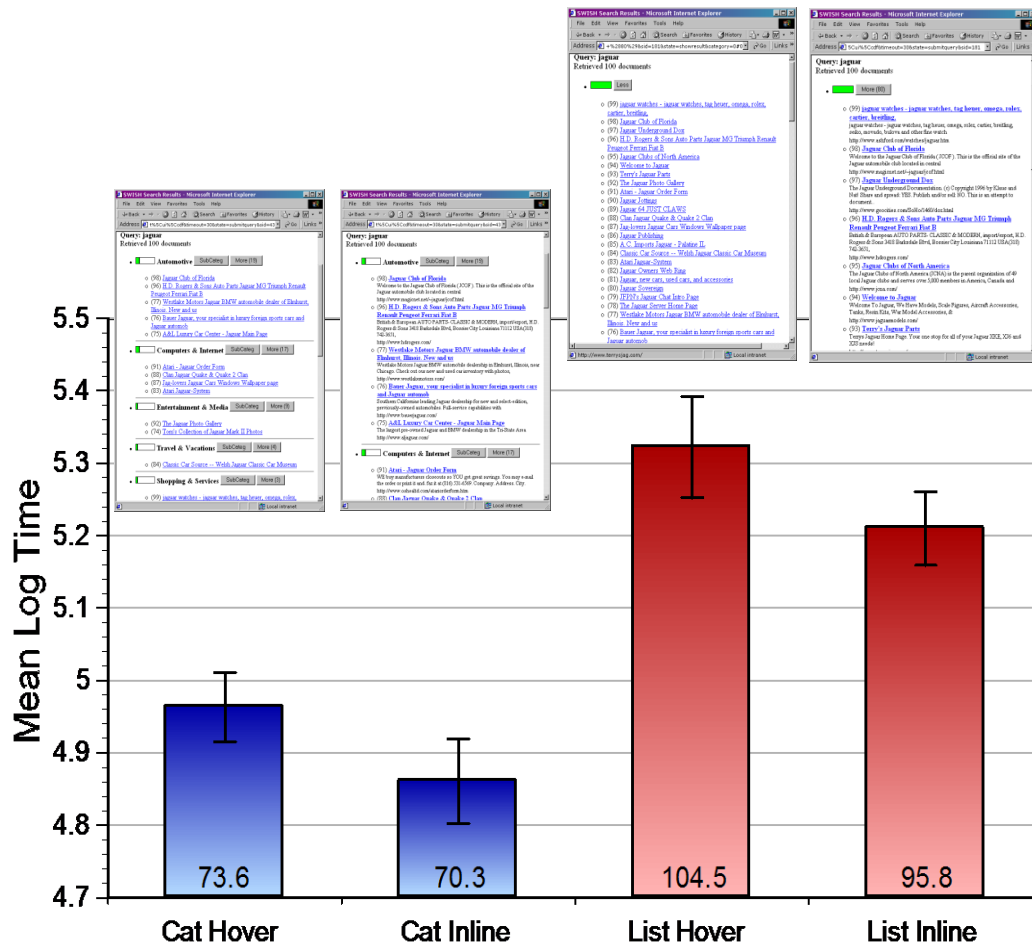
Inline faster than
    Hover Text, $p<0.06$
Category faster than
    List interface, $p<0.01$

No interaction
No gender effects

List: Hover, Inline, CatNames
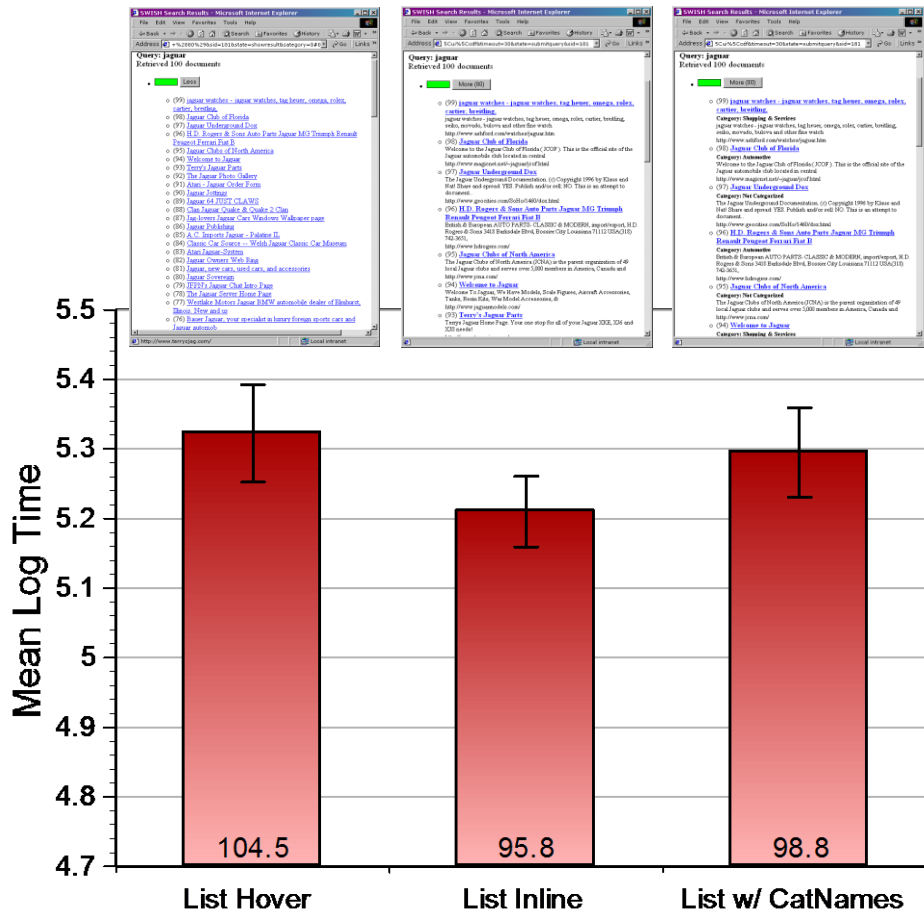
Adding Category Names to List results does not help

Category: Remove Cat Names
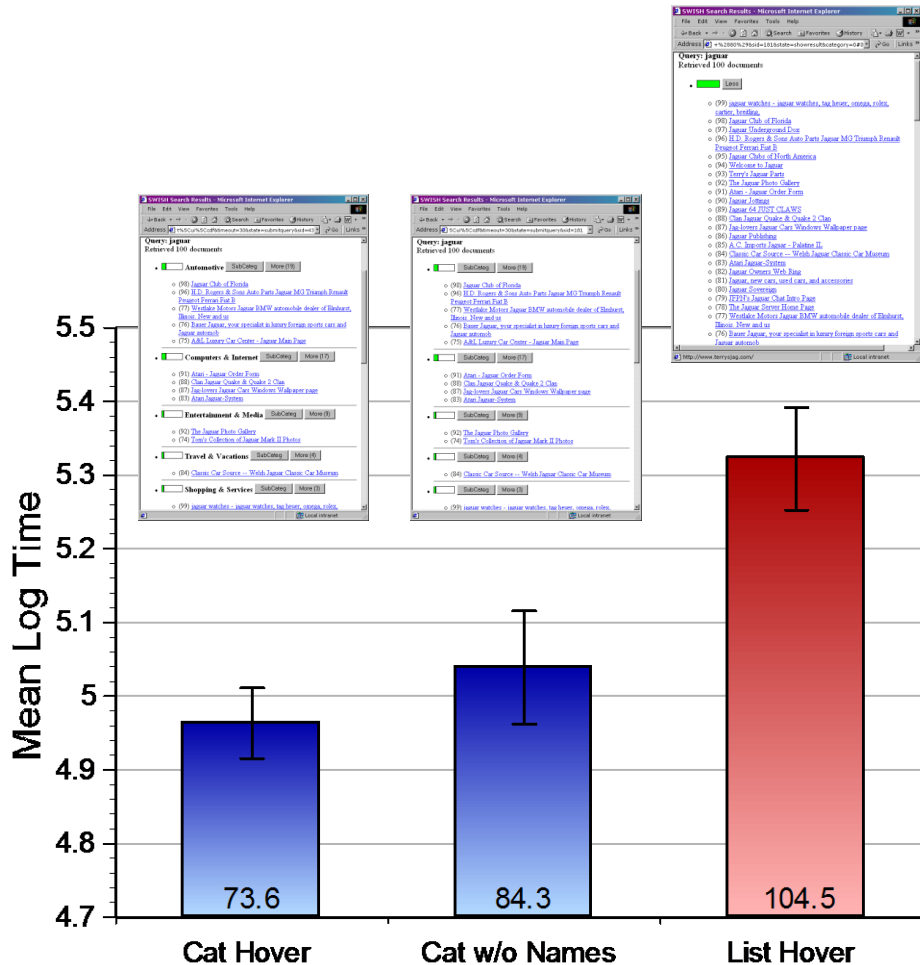
Removing Category Names no worse than Category with Names

Better than List hover, $p<0.01$

Category: Remove Page Titles (Browse)
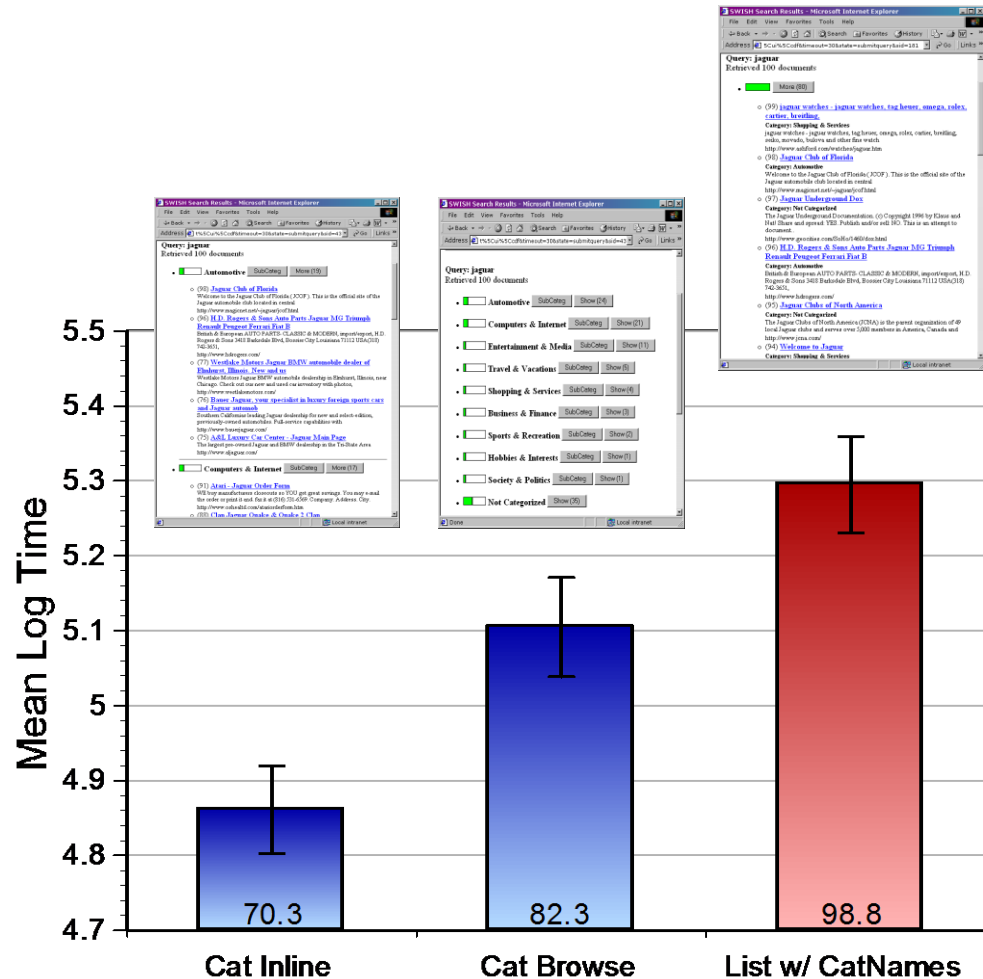
Browsing slower than Category inline, *p<0.01*

Browsing faster than List with CatNames, *p<0.02*

Mean Log Time

| 5.5 | 5.4 | 5.3 | 5.2 | 5.1 | 5 | 4.9 | 4.8 | 4.7 |

73.6 — Cat Hover
70.3 — Cat Inline
84.3 — Cat w/o Names
82.3 — Cat Browse
104.5 — List Hover
95.8 — List Inline
98.8 — List w/ CatNames
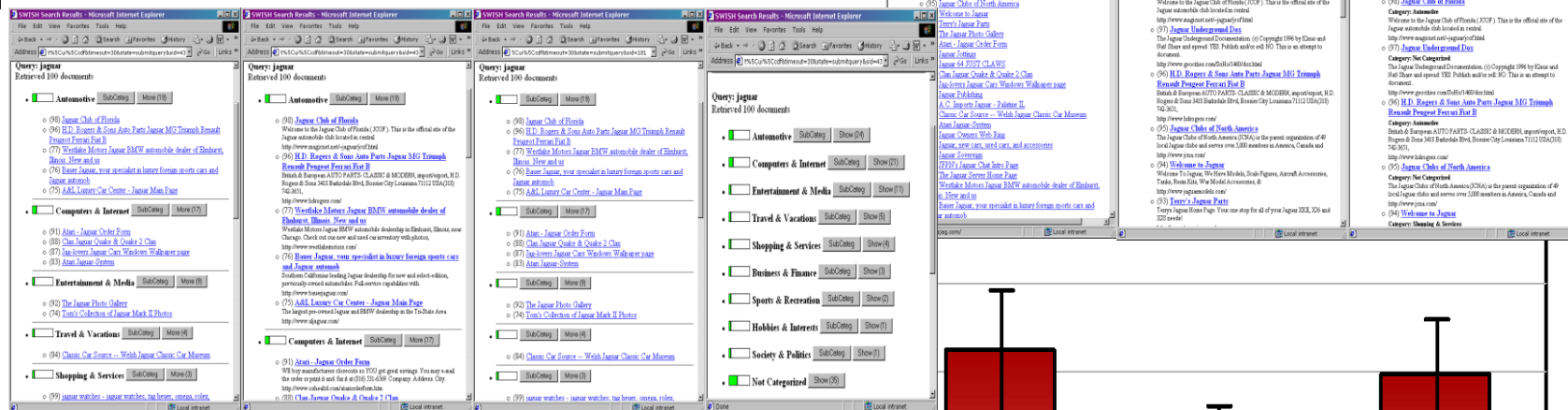
Summary

- Organizing Search Results
  - Learn accurate hierarchical classification models using SVMs
  - Classify new web pages on-the-fly
  - UI to expose classification is the key
- User Interface Design and Studies
  - Inline summaries better than Hover summaries
  - Category interfaces better than List interfaces
    - Strong preference and performance advantages for categorically organized presentation of search results
  - Best Category interface uses both category names (context) and page titles (focus)
    - Results "in" context; not just results "plus" context
    - Page Titles a little more useful than Category Names

Open Issues

- Improve Accuracy of Classification Algorithms
- Alternative User Interface Combinations
  - Heuristics for selecting categories and pages to display
    - Query_Match: rank of page, and sometimes match score
    - Categ_Match: p(category for each page)
  - Other methods for combining context and focus
  - Integration with non-content information
- Conduct End-to-end User Study
- More info: *http://research.microsoft.com/~sdumais*

Searching With Information Structured Hierarchically

SWISH
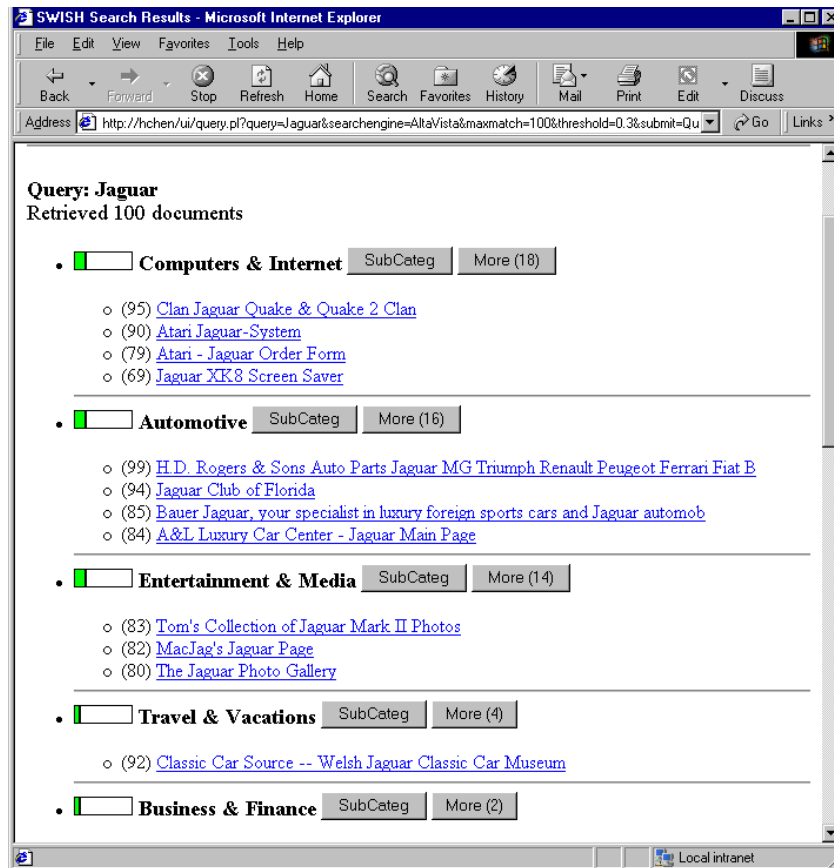
Did not try …

- Flat groupings a la Northern Light
- Separate category and list (tho probably close to category browse + list + switching time)

File   Edit   View   Favorites   Tools   Help

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address 🔲 http://search.yahoo.com/search?p=jaguar&hc=33&hs=453&h=c&b=21   ⮕ Go   Links »
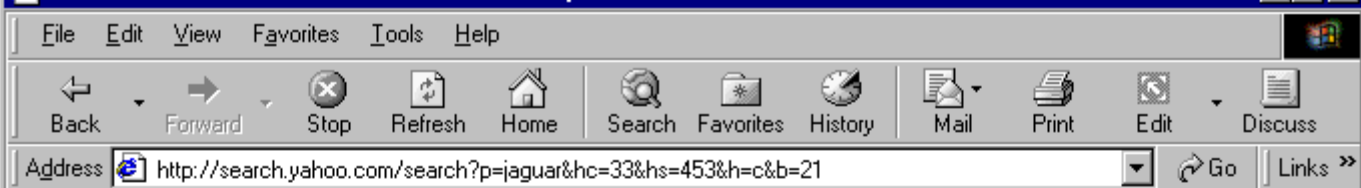
University of South Alabama · Athletics

Yahoo! Site Matches   (1 - 7 of 453)

Business and Economy > Companies > Automotive > Shopping and Services > Parts > Makes and Models > **Jaguar**

- Always **Jaguar** - online ordering of parts for all years and models, 1948 to the present.

Recreation > Sports > Auto Racing > Formula One > Teams

- **Jaguar** Racing - official site.

Business and Economy > Companies > Automotive > Shopping and Services > Makers > Vehicles > Ford Motor Company > Subsidiaries > **Jaguar** > Products and Services

- **Jaguar** S-Type

Business and Economy > Companies > Automotive > Business to Business > Parts > Makes and Models > **Jaguar**

- Always **Jaguar** - online ordering of parts for all years and models, 1948 to the present.

Business and Economy > Companies > Automotive > Shopping and Services > Makers > Vehicles > Ford Motor Company > Subsidiaries > **Jaguar** > Regional

- **Jaguar** Nederland C.V. - includes the official introduction of the new **Jaguar** XK8. Complete with screensavers, qtvr movies and pictures.

Business and Economy > Companies > Automotive > Shopping and Services > Dealers > Makes > **Jaguar**

🔲 Internet

Yahoo! Category Matches   (1 - 20 of 33)

Recreation > Automotive > Makes and Models > **Jaguar**

Business and Economy > Companies > Automotive > Shopping and Services > Dealers > Makes > **Jaguar**

Recreation > Games > Video Games > Systems > Atari > **Jaguar**

Business and Economy > Companies > Automotive > Shopping and Services > Parts > Makes and Models > **Jaguar**

Business and Economy > Companies > Automotive > Shopping and Services > Makers > Vehicles > Ford Motor Company > Subsidiaries > **Jaguar**

Business and Economy > Companies > Automotive > Business to Business > Parts > Makes and Models > **Jaguar**

Regional > Countries > United Kingdom > Business and Economy > Companies > Automotive > Shopping and Services > Parts > Makes and Models > **Jaguar**

Regional > U.S. States > Florida > Cities > Jacksonville > Business and Shopping > Shopping and Services > Automotive > Dealers > Makes > **Jaguar**

Regional > U.S. States > Virginia > Cities > Richmond > Business and Shopping > Shopping and Services > Automotive > Dealers > Makes > **Jaguar**

Regional > U.S. States > California > Cities > Long Beach > Business and Shopping > Shopping and Services > Automotive > Dealers > Makes > **Jaguar**

Northern Light Search: jaguar - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address http://www.northernlight.com/nlquery.fcg?cb=0&qr=jaguar&orl=   Go   Links »

Simple Search | Power | Business | Investext | Stock Quotes | Search News | Special Ec

Northern Light

Buy books at
BARNES&NOBLE

We're a new kind of data and Internet company.

294,652 items for:

jaguar   Search   Tips | Save Ale
Edit this search

Narrow Your Search with Custom Search Folders™

Your search returned 294,652 items which we have organized into the following Custom Search Folders:

📁 Search Current News
📁 Special Collection documents
📁 Jaguar cars
📁 Jacksonville Jaguars
📁 Investing
📁 Jaguar XJ6
📁 Fragrances
📁 Games & puzzles
📁 Cars & driving
📁 Carnivores
📁 Jaguar XK8
📁 Jaguar XJR
📁 Ferrari cars

📄 Documents that best match your search

1. Jaguar Cars Global Home Page
   72% - Company home page: 02/02/2000
   Commercial site: http://www.jaguarvehicles.com/
   📁 More results from this site

2. NFL.COM - Jaguars
   70% - Articles & General info: ESPN Network: DIVISION STANDINGS.
   3. 8- 8. 6- 10. 4- 12. 2- 14. Frank Gansz is one of the top special teams...
   03/08/2000
   Commercial site: http://www.nfl.com/jaguars/
   📁 More results from this site

3. Eurêka, c'est trouvé ! - L'annuaire de l'internet francophone - 85.000 sites .
   67% - Articles & General info: Thmes- Abbaye. Avocats. Baby-sitting. B
   C+ Cyclisme. Dactylographie. Dvd. Eau. Extreme. Faience. Fusain. Galer
   Gyncologie. Habilet. Hypnose. Icq. Iufm. Jaguar. Justice. Karaoke. Kosov
   Laboratoires... Date Not Available
   Commercial site: http://www.ctrouve.com/

Done   Internet

File   Edit   View   Favorites   Tools   Help

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address http://search.oingo.com/find.dll?s=jaguar&n=n   Go   Links »

The symbol 💡 indicates a "Meaning Hit" - the subject appears to be conceptually related to your search terms.
Click here to change Oingo Adult Material Filter settings...

Oingo Categories

💡 **Jaguar** (Regional / Europe / United Kingdom / Recreation and Sports / Motoring / Makes and Models / Jaguar)

   XK Club for **Jaguar** Sports Car Enthusiasts
   The International mecca for all Jaguar XK owners and enthusiasts.

💡 **Jaguar** (Recreation / Autos / Manufacturers / Jaguar)

   (16 links with word matches.)

💡 **Panthera onca** (Science / Biology / Taxonomy / Taxonomy Map / Animalia / Chordata / Mammalia / Carnivora / Felidae / Panthera / Panthera onca)

   (4 links with word matches.)

💡 **Panthera** (Science / Biology / Taxonomy / Taxonomy Map / Animalia / Chordata / Mammalia / Carnivora / Felidae / Panthera)

💡 **Product Reviews: Jaguar** (Home / Consumers / Information / Product Reviews / Autos / Car Manufacturers / Jaguar)

   (4 links with word matches.)

💡 **Jaguar Car Clubs** (Recreation / Autos / Clubs / Makes / Jaguar)

   (19 links with word matches.)

💡 **Epinions.com: New Jaguars** (Home / Consumers / Consumer Information / Product Reviews / Epinions.com / Autos / New / Jaguar)

Web Sites

- **Jaguar Cars Global Home Page**
  Jaguar Cars Official Web Site.

- **Jaguar Cars Global Home Page**
  Jaguar Cars Official Web Site....

- **Jaguar Cars Global Home Page**
  Jaguar Cars Official Web Site....

- **Jaguar Australia**
  ...

- **Jaguar Sovereign**
  Virtual Wheels Click icons to navigate: [Home] [Opinions] [Customline] [Sovereign] [Links] [Email us] Alexandra. a 1986 Jaguar XJ6 Sovereign 4.2. ...

- **Cole's Jaguar Page**
  This...

- **The Jaguar's Domain**
  Last updated 01/10/00. Visitors since 2/18/99 Service provided by Web-Counter. The Atari Jaguar is THE original 64-bit system. Atari is basically a...

- **Jaguar History**
  The History of Jaguar. By Mike Owens Pre-Jaguar Days: The Swallow Sidecar Company. On September 4, 1922, in Blackpool, England, two young motorcycle...

- **Jaguar Reef Lodge, Hopkins, Stann Creek, Belize, Central America**
  Jaguar Reef Lodge, Belize. A tropical resort between coral reef and

Internet

Ask.com

JEEVES SERVICES:

Personal Jeeves

Answer Point

Money

Travel

Health

Computers

Entertainment

Home & Family

Shopping

Ask Jeeves for Kids!

Visit our sponsors:
uBid Online Auction
Bid and sell at eBay

Ask Jeeves Live

Mother's Day Personal Shopper

How can Jeeves help me shop for:
· Books
· Computers
· Music
· Toys

○ Roulette   ○ Video P   review   CASINO ON THE

You asked: jaguar   **Ask!**

I have found answers to the following questions:

Ask! Where can I find information on the animal jaguar ?

Ask! Where can I find information on an exotic Jaguar XJ220 ?

Ask! Where can I find the Web site for the company Jaguar ?

Ask! Where can I find car reviews for Jaguars ?

Ask! Where can I find information on the new Jaguar S-Type ?

exp Get your answer from an Expert on **Academic Subjects** at **EXP.com**

Ask! **Jaguar Cars Global Home Page**
Jaguar Cars Official Web Site.

Ask! **Jag-lovers - the Jaguar Enthusiasts' premier resource on the Internet with**
The Premier Jaguar Cars resource and meeting place homepage on the web for all enthusiasts

Ask! **Jaguar XJ6L**
Click Here and find big discounts on the Jaguar XJ6L.

More >>

Ask! 7 matches by AltaVista Jaguar Cars Global Home Page

Ask! 10 matches by WebCrawler AGH Museum – Atari Jaguar & Lynx

Done   Internet

File   Edit   View   Favorites   Tools   Help

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address http://research.microsoft.com/scripts/search/query.asp   Go   Links »

Microsoft Research Search

Our Recommendations:
MSR People
MSRIPv6 Download Page
Downloads

- home
- current research
- people
- **search**
- news
- publications
- partners
- downloads
- opportunities
- labs
- visit msr
- university programs

What do you want to find?   Search tips

[ ]   Search

SearchResults
Click to **show all results sorted by rank**

Search Results for **rashid**

People

1. Rick Rashid

Publications

1. The Tiger Video Fileserver
   Bolosky, William J. ; Barrera, Joseph S. III; Draves, Richard P.; Fitzgerald, Robert P.; Gibson, Garth A. ; Jones, Michael B. ; Levi, Steven P.; Myhrvold, Nathan P.; Rashid, Richard F.

News

1. Rick Rashid, Developer of Mach System, Joins Microsoft as Director of Research
2. Rashid Credited with Building a Highly Productive Research Organization from the Ground Up
3. Moving on up: Rick Rashid and Dan Ling promoted
4. Q&A with Rashid on NLP and the Demise of Moore's Law
5. Rick Rashid and Michael Cohen Speak to Moore's Law

Done   Internet

Our Recommendations:

MSR People

MSRIPv6 Download Page

Downloads

Microsoft

Research Search

home

current research

people

search

news

publications

partners

downloads

opportunities

labs

visit msr

university programs

What do you want to find?   Search tips

Search

SearchResults   Search Results for **rashid**

Click to **show all result sorted by category**

👤 = Person   📗 = News   🔷 = Publication

1. Rick Rashid
    Rick Rashid's Home Page

2. Rick Rashid
    Rick Rashid's Home Page

3. 🔷 The Tiger Video Fileserver
       Bolosky, William J. ; Barrera, Joseph S. III; Draves, Richard P.; Fitzgerald, Robert P.; Gibson, Garth A. ; Jones, Michael B. ; Levi, Steven P.; Myhrvold, Nathan P.; Rashid, Richard F.

4. 📗 Rick Rashid, Developer of Mach System, Joins Microsoft as Director of Research

5. 👤 Rick , Rashid
6. 📗 Rashid Credited with Building a Highly Productive Research Organization from the Ground Up

Done   Internet

File   Edit   View   Favorites   Tools   Help

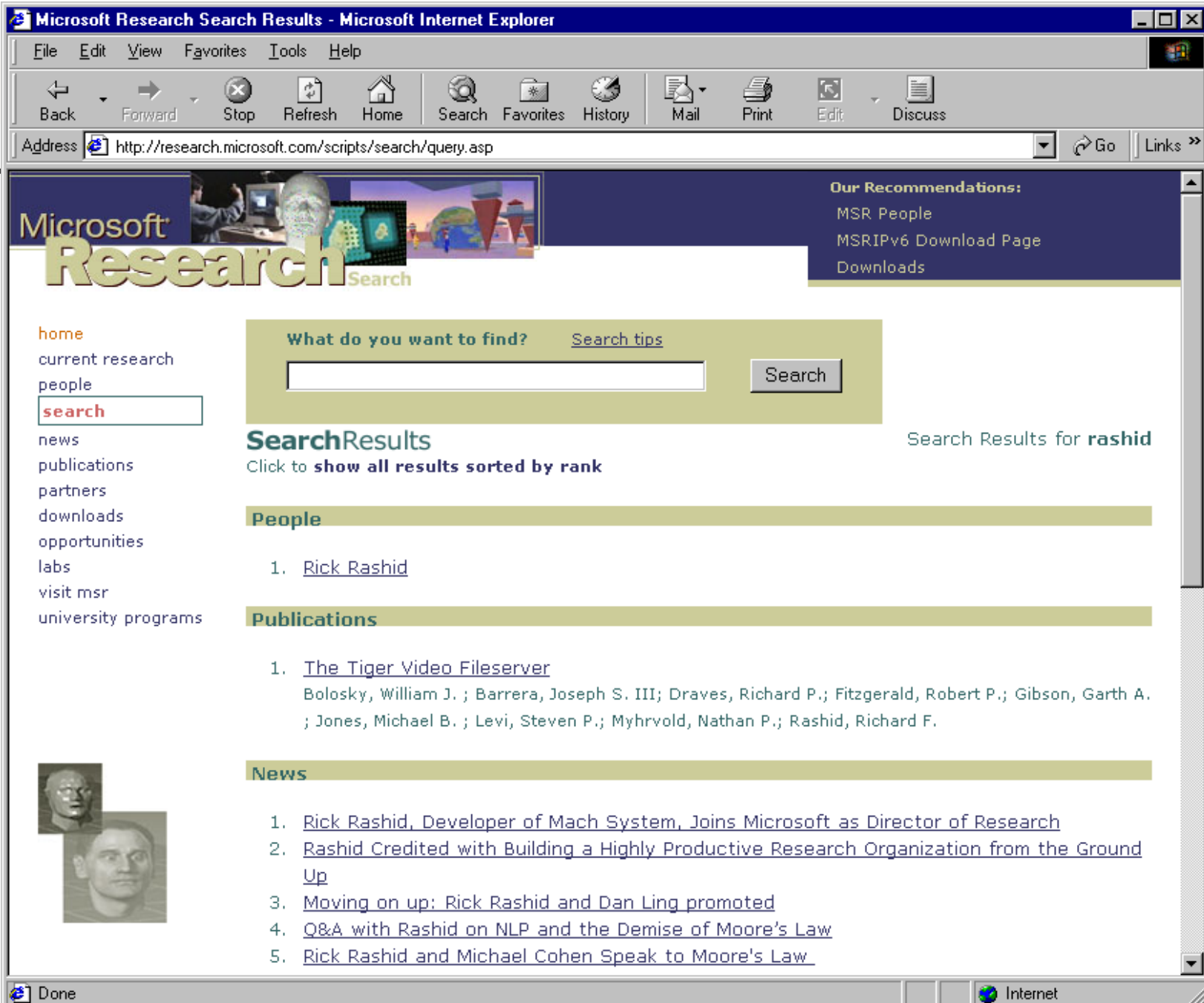Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss
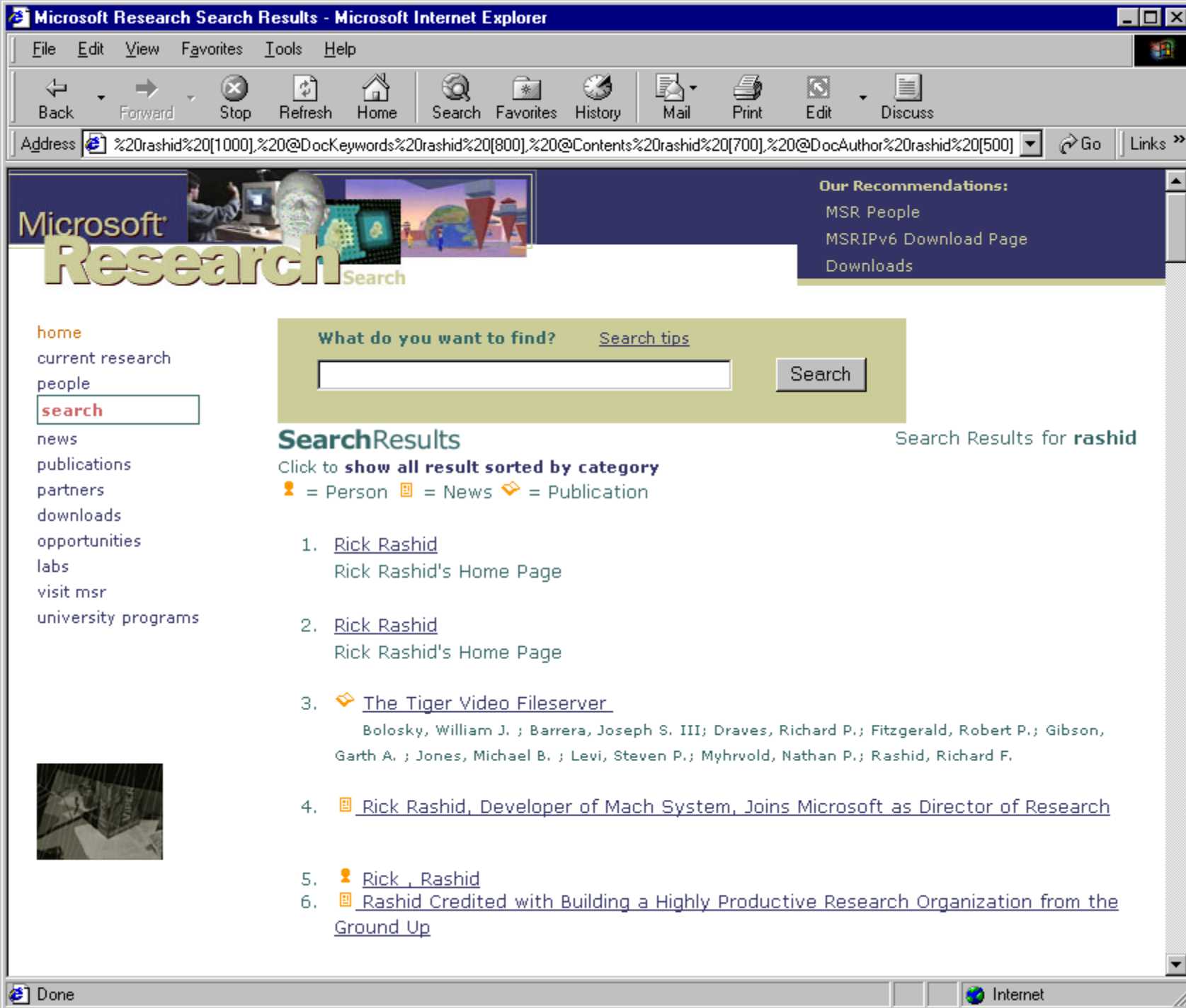
Address   umdoc=1&averagescore=0&cdfdir=e%3A%5Cproject%5Cui%5Ccdf&timeout=30&state=submitquery&sid=265   Go   Links »

Query: jaguar
Retrieved 100 documents

- **Automotive**   SubCateg   More (19)

    - (98) Jaguar Club of Florida
    - (96) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
    - (77) Westlake Motors Jaguar BMW automobile dealer of Elmhurst, Illinois. New and us
    - (76) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
    - (75) A&L Luxury Car Center - Jaguar Main Page

- **Computers & Internet**   SubCateg   More (17)

    - (91) Atari - Jaguar Order Form
    - (88) Clan Jaguar Quake & Quake 2 Clan
    - (87) Jag-lovers Jaguar Cars Windows Wallpaper page
    - (83) Atari Jaguar-System

- **Entertainment & Media**   SubCateg   More (9)

    - (92) The Jaguar Photo Gallery
    - (74) Tom's Collection of Jaguar Mark II Photos

- **Travel & Vacations**   SubCateg   More (4)

    - (84) Classic Car Source -- Welsh Jaguar Classic Car Museum

- **Shopping & Services**   SubCateg   More (3)

Done   Local intranet

Query: jaguar
Retrieved 100 documents

- Search Results More (80)

  - (99) jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling,
  - (98) Jaguar Club of Florida
  - (97) Jaguar Underground Dox
  - (96) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - (95) Jaguar Clubs of North America
  - (94) Welcome to Jaguar
  - (93) Terry's Jaguar Parts
  - (92) The Jaguar Photo Gallery
  - (91) Atari - Jaguar Order Form
  - (90) Jaguar Jottings
  - (89) Jaguar 64 JUST CLAWS
  - (88) Clan Jaguar Quake & Quake 2 Clan
  - (87) Jag-lovers Jaguar Cars Windows Wallpaper page
  - (86) Jaguar Publishing
  - (85) A.C. Imports Jaguar - Palatine IL
  - (84) Classic Car Source -- Welsh Jaguar Classic Car Museum
  - (83) Atari Jaguar-System
  - (82) Jaguar Owners Web Ring
  - (81) Jaguar, new cars, used cars, and accessories
  - (80) Jaguar Sovereign

Query: jaguar
Retrieved 100 documents

- ▐ ▭ **Automotive**   SubCateg   More (19)

  - (98) **Jaguar Club of Florida**
    Welcome to the Jaguar Club of Florida (JCOF). This is the official site of the Jaguar automobile club located in central
    http://www.magicnet.net/~jaguar/jcof.html

  - (96) **H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B**
    British & European AUTO PARTS- CLASSIC & MODERN, import/export, H.D. Rogers & Sons 3418 Barksdale Blvd, Bossier City Louisiana 71112 USA(318) 742-3651,
    http://www.hdrogers.com/

  - (77) **Westlake Motors Jaguar BMW automobile dealer of Elmhurst, Illinois. New and us**
    Westlake Motors Jaguar BMW automobile dealership in Elmhurst, Illinois, near Chicago. Check out our new and used car inventory with photos,
    http://www.westlakemotors.com/

  - (76) **Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob**
    Southern Californias leading Jaguar dealership for new and select-edition, previously-owned automobiles. Full-service capabilities with
    http://www.bauerjaguar.com/

  - (75) **A&L Luxury Car Center - Jaguar Main Page**
    The largest pre-owned Jaguar and BMW dealership in the Tri-State Area
    http://www.aljaguar.com/

- ▐ ▭ **Computers & Internet**   SubCateg   More (17)

  - (91) **Atari - Jaguar Order Form**
    WE buy manufacturers closeouts so YOU get great savings. You may e-mail the order or print it and. fax it at (816) 531-6569. Company. Address. City.
    http://www.ogbooltd.com/ctoriorderform.htm

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address   umdoc=1&averagescore=0&cdfdir=e%3A%5Cproject%5Cui%5Ccdf&timeout=30&state=submitquery&sid=265   Go   Links »

Query: jaguar
Retrieved 100 documents

- ⬜ **Search Results**   More (80)

  - (99) **jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling,**
    jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling, seiko, movado, bulova and other fine watch

    http://www.ashford.com/watches/jaguar.htm

  - (98) **Jaguar Club of Florida**
    Welcome to the Jaguar Club of Florida (JCOF). This is the official site of the Jaguar automobile club located in central

    http://www.magicnet.net/~jaguar/jcof.html

  - (97) **Jaguar Underground Dox**
    The Jaguar Underground Documentation. (c) Copyright 1996 by Klaus and Nat! Share and spread: YES. Publish and/or sell: NO. This is an attempt to document..

    http://www.geocities.com/SoHo/1460/dox.html

  - (96) **H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B**
    British & European AUTO PARTS- CLASSIC & MODERN, import/export, H.D. Rogers & Sons 3418 Barksdale Blvd, Bossier City Louisiana 71112 USA(318) 742-3651,

    http://www.hdrogers.com/

  - (95) **Jaguar Clubs of North America**
    The Jaguar Clubs of North America (JCNA) is the parent organization of 49 local Jaguar clubs and serves over 5,000 members in America, Canada and

    http://www.jcna.com/

  - (94) **Welcome to Jaguar**
    Welcome To Jaguar, We Have Models, Scale Figures, Aircraft Accessories, Tanks, Resin Kits, War Model Accessories, &

    http://www.jaguarmodels.com/

  - (93) **Terry's Jaguar Parts**
    Terrys Jaguar Home Page. Your one stop for all of your Jaguar XKE, XJ6 and XJS needs!

Done   Local intranet

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit   Discuss

Address   umdoc=1&averagescore=0&cdfdir=e%3A%5Cproject%5Cui%5Ccdf&timeout=30&state=submitquery&sid=265   Go   Links »

Query: jaguar
Retrieved 100 documents

- [SubCateg]   [More (19)]

    - (98) Jaguar Club of Florida
    - (96) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
    - (77) Westlake Motors Jaguar BMW automobile dealer of Elmhurst, Illinois. New and us
    - (76) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
    - (75) A&L Luxury Car Center - Jaguar Main Page

- [SubCateg]   [More (17)]

    - (91) Atari - Jaguar Order Form
    - (88) Clan Jaguar Quake & Quake 2 Clan
    - (87) Jag-lovers Jaguar Cars Windows Wallpaper page
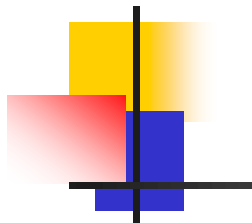   - (83) Atari Jaguar-System

- [SubCateg]   [More (9)]

    - (92) The Jaguar Photo Gallery
    - (74) Tom's Collection of Jaguar Mark II Photos

- [SubCateg]   [More (4)]

    - (84) Classic Car Source -- Welsh Jaguar Classic Car Museum

- [SubCateg]   [More (3)]

Done   Local intranet

Back   Forward   Stop   Refresh   Home   Search   Favorites   History   Mail   Print   Edit
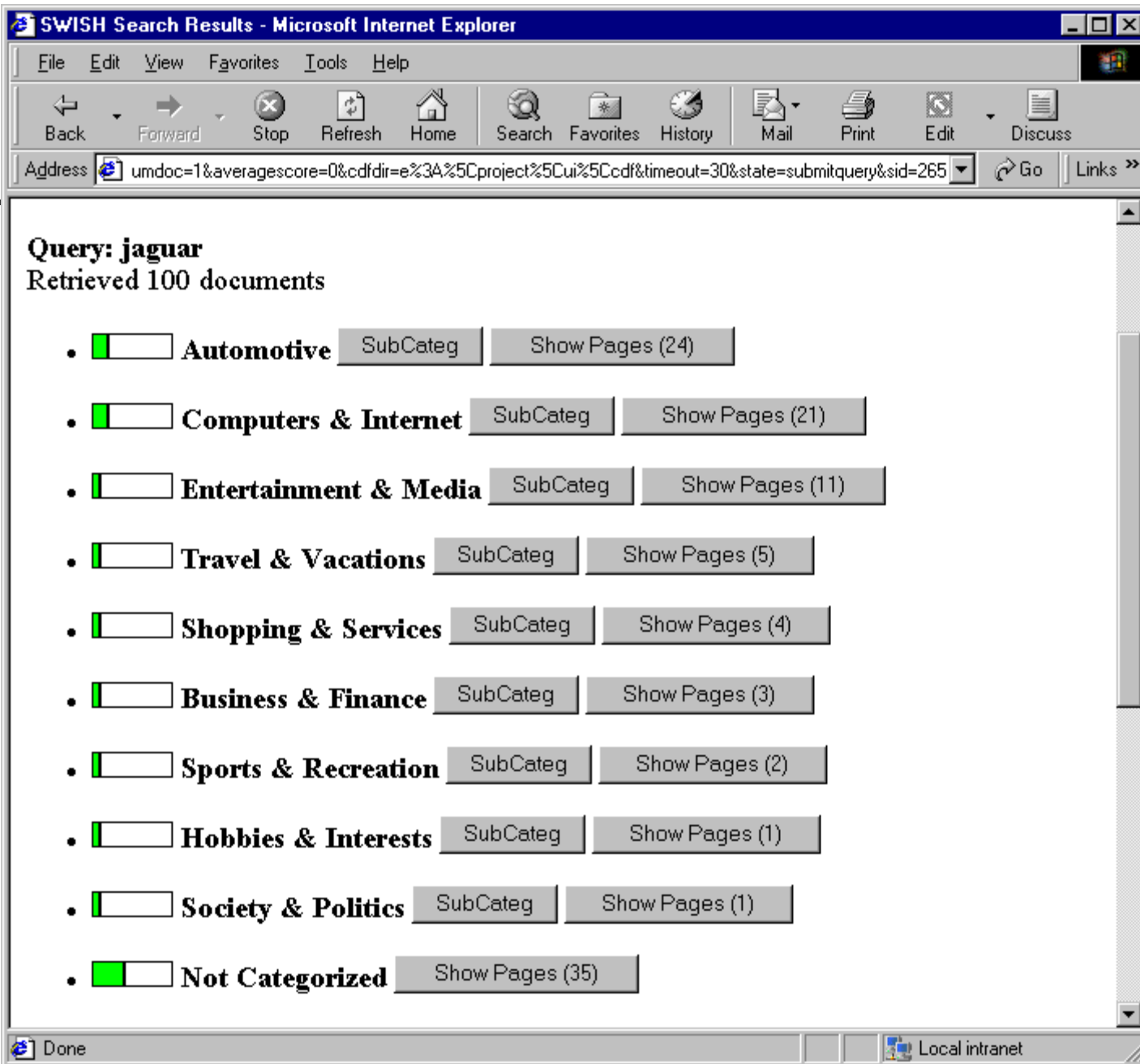
Address   http://hchen/ui/query.pl?query=jaguar&searchengine=AltaVista&maxmatch=100&threshold=0.3&su   Go   Links »
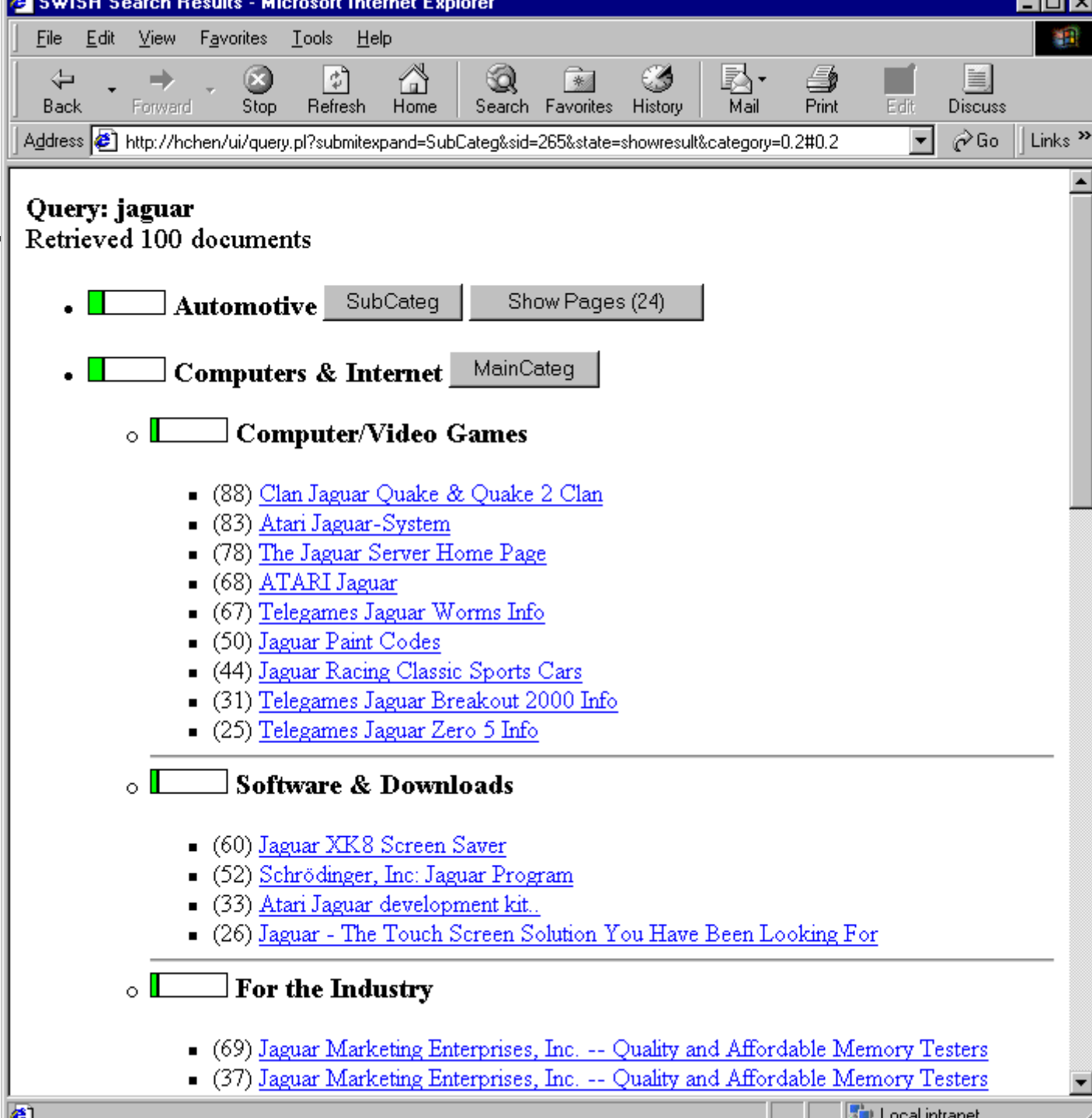
Query: jaguar
Retrieved 100 documents

- ▮ **Search Results** More (80)

  - (99) **jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling,**
    Category: Shopping & Services
    jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling, seiko, movado, bulova and other fine watch
    http://www.ashford.com/watches/jaguar.htm

  - (98) **Jaguar Club of Florida**
    Category: Automotive
    Welcome to the Jaguar Club of Florida (JCOF). This is the official site of the Jaguar automobile club located in central
    http://www.magicnet.net/~jaguar/jcof.html

  - (97) **Jaguar Underground Dox**
    Category: Not Categorized
    The Jaguar Underground Documentation. (c) Copyright 1996 by Klaus and Nat! Share and spread: YES. Publish and/or sell: NO. This is an attempt to document..
    http://www.geocities.com/SoHo/1460/dox.html

  - (96) **H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B**
    Category: Automotive
    British & European AUTO PARTS- CLASSIC & MODERN, import/export, H.D. Rogers & Sons 3418 Barksdale Blvd, Bossier City Louisiana 71112 USA(318) 742-3651,
    http://www.hdrogers.com/

  - (95) **Jaguar Clubs of North America**
    Category: Not Categorized
    The Jaguar Clubs of North America (JCNA) is the parent organization of 49 local Jaguar clubs and serves over 5,000 members in America, Canada and
    http://www.jcna.com/

  - (94) **Welcome to Jaguar**

Local intranet

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss

Address umdoc=1&averagescore=0&cdfdir=e%3A%5Cproject%5Cui%5Ccdf&timeout=30&state=submitquery&sid=265 Go Links »

Query: jaguar
Retrieved 100 documents

- **Automotive** SubCateg Show Pages (24)

- **Computers & Internet** SubCateg Show Pages (21)

- **Entertainment & Media** SubCateg Show Pages (11)

- **Travel & Vacations** SubCateg Show Pages (5)

- **Shopping & Services** SubCateg Show Pages (4)

- **Business & Finance** SubCateg Show Pages (3)

- **Sports & Recreation** SubCateg Show Pages (2)

- **Hobbies & Interests** SubCateg Show Pages (1)

- **Society & Politics** SubCateg Show Pages (1)

- **Not Categorized** Show Pages (35)

Done Local intranet

Query: jaguar
Retrieved 100 documents

- ▮▭ **Automotive**   SubCateg   Show Pages (24)

- ▮▭ **Computers & Internet**   MainCateg

  - ▮▭ **Computer/Video Games**

    - (88) Clan Jaguar Quake & Quake 2 Clan
    - (83) Atari Jaguar-System
    - (78) The Jaguar Server Home Page
    - (68) ATARI Jaguar
    - (67) Telegames Jaguar Worms Info
    - (50) Jaguar Paint Codes
    - (44) Jaguar Racing Classic Sports Cars
    - (31) Telegames Jaguar Breakout 2000 Info
    - (25) Telegames Jaguar Zero 5 Info

  - ▮▭ **Software & Downloads**

    - (60) Jaguar XK8 Screen Saver
    - (52) Schrödinger, Inc: Jaguar Program
    - (33) Atari Jaguar development kit..
    - (26) Jaguar - The Touch Screen Solution You Have Been Looking For

  - ▮▭ **For the Industry**

    - (69) Jaguar Marketing Enterprises, Inc. -- Quality and Affordable Memory Testers
    - (37) Jaguar Marketing Enterprises, Inc. -- Quality and Affordable Memory Testers

Variability

- Across subjects
  - 37 secs/task -> 142 secs/task
- Across queries
  - category: 24 secs -> 158 secs
  - list: 18 secs -> 217 secs
  - fastest - *"Ford's theater"*
  - slowest - *"Books about the author, Dylan Thomas"*

Training Data: LookSmart Web Dir

- Random sample
  - Pages can be in more than one category
- Training set
  - 50k pages
  - Top-level (13 categories; 578-11163 pages per category)
  - Second-level (150 categories; 3-3141 pages per category)
- Testing set
  - 10k pages
  - Compute $f(\vec{w} \cdot \vec{x})$
  - Compare to threshold

Learned Classifiers

- Non-Hierarchical (Flat)
  - 150 second-level classifiers
  - P(L2) > threshold
- Hierarchical
  - 13 top-level; 150 second-level classifiers
  - Use for both feature selection and model building
  - Combine classifiers from L1 and L2 using:
    - P(L1)*P(L2|L1) > threshold   (*multiplicative*)
    - P(L1)>threshold && P(L2|L1)>threshold (*Boolean*)

Best/Worst Categories

Category Name	
Best F1	
0.841	*Health & Fitness/Drugs & Medicines*
0.797	*Home & Family/Real Estate*
0.781	*Reference & Education/K-12 Education*
0.750	*Sports & Recreation/Fishing*
0.741	*Reference & Education/Higher & Cont. Ed.*
Worst F1	
0.034	*Society & Politics/World Cultures*
0.088	*Home & Family/For Kids*
0.122	*Computers & Internet/News & Magazines*
0.131	*Computers & Internet/Internet & the Web*
0.133	*Business & Finance/Business Professions*

Results

	Accuracy	Efficiency	Efficiency
	F1	training (sec)	runtime
L1:			
Hierarchical, L1	0.649		
L2:			
Non-Hierarchical	0.476	729	150 categ
Hierarchical, P(L1)*P(L2|L1)	0.495	1258 + 128	13 + 150 categ
Hierarchical, P(L1)&&P(L2)	0.497	1258 + 128	13 + 11 categ
Hierarchical, P(L1)=1	0.711		

Text Classification

- *Text Classification*: assign objects to one or more of a predefined set of categories using text features
  - E.g., News feeds, OHSUMED, Email - spam/no-spam, Web data
- Approaches:
  - Human classification (e.g., LCSH, MeSH, Yahoo!, CyberPatrol)
  - Hand-crafted knowledge engineered systems (e.g., CONSTRUE)
  - Inductive learning methods
   - Learn models from examples -> (Semi-) automatic classification

Inductive Learning Methods

- ## Supervised learning from *examples*
  - Examples are easy for domain experts to provide
  - Models easy to learn, update, and customize
- ## Example learning algorithms
  - Rocchio-style Relevance Feedback
  - Naïve Bayes
  - Bayes Nets
  - Decision Trees
  - Linear Discriminant Analysis (LDA)
  - Language Models
  - Support Vector Machines (SVMs)

Support Vector Machine (SVM)

- ## Optimization Problem
  - Find hyperplane, h, separating positive and negative examples
  - Optimization for maximum margin: $\min \|\vec{w}\|^2, \vec{w} \cdot \vec{x} - b \geq 1, \vec{w} \cdot \vec{x} - b \leq -1$
  - Classify new items using: $f(\vec{w} \cdot \vec{x})$

support vectors
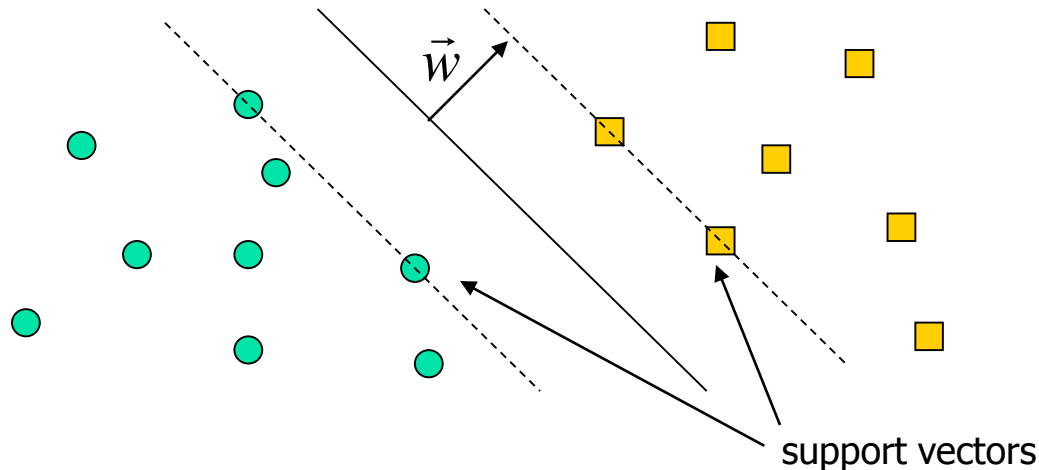
SVMs for Text Classification

- Can also handle
  - Non-separable problems
  - Non-linear problems
- Accurate and efficient for text classification
  - Dumais et al. (1998); Joachims (1998); Yang and Lui (1999)
  - Platt's SMO algorithm for efficiency
- Text representation
  - Large vector of features (words, phrases, hand-crafted)

SVMs for Text Classification

- ## Learned Model, $\vec{w}$
  - Weighted vector of words
    - "Automobile" = motorcycle, vehicle, parts, automobile, harley, car, auto, honda, porsche …
    - "Computers & Internet" = rfc, software, provider, windows, user, users, pc, hosting, os, downloads …

- ## Classify using, $f(\vec{w} \cdot \vec{x})$
  - Sigmoid to produce posterior probabilities

- ## Hierarchical Models
  - 1 model for N top level categories
  - N models for second level categories
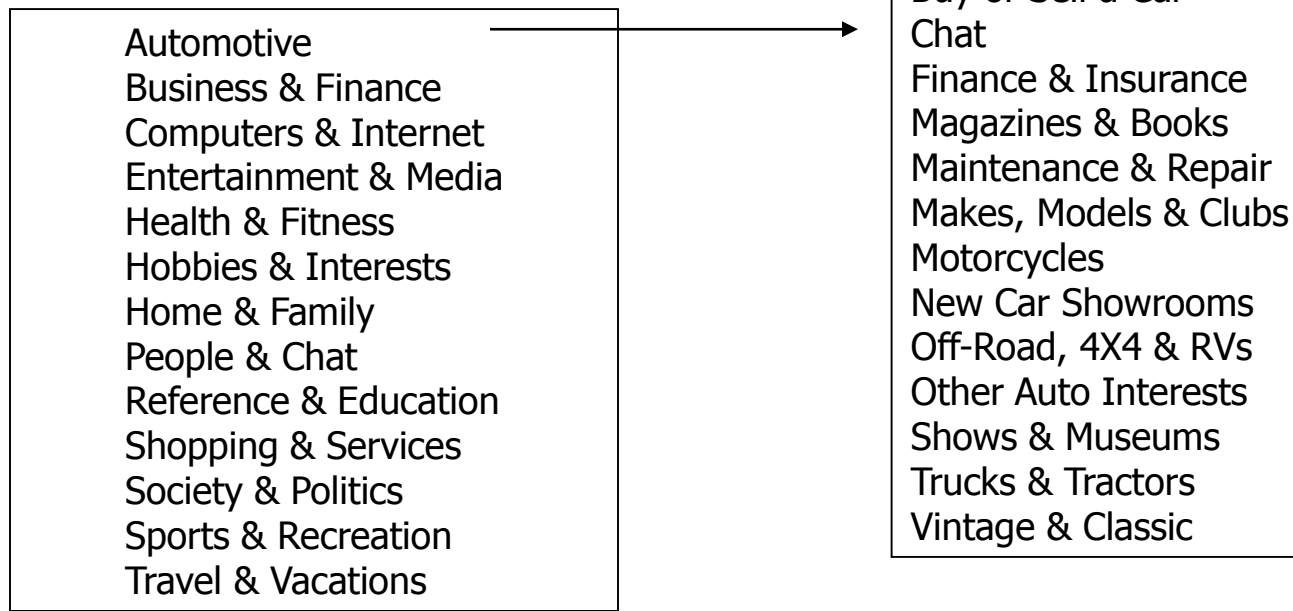 - Very useful in conjunction w/ user interaction

Hierarchical Classification
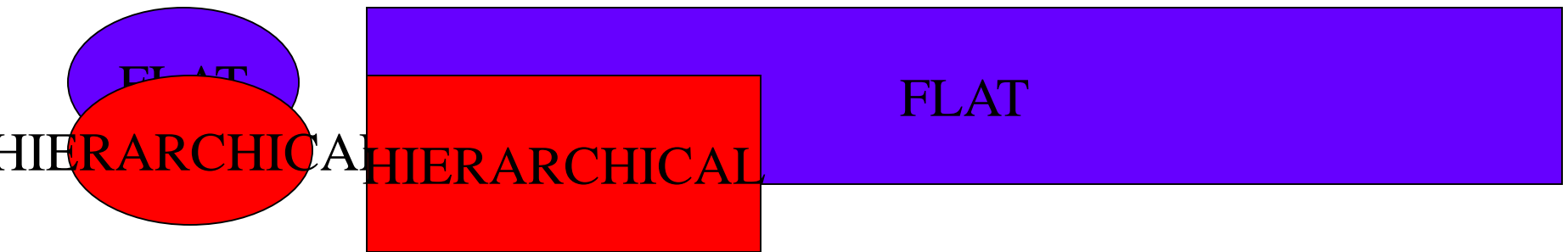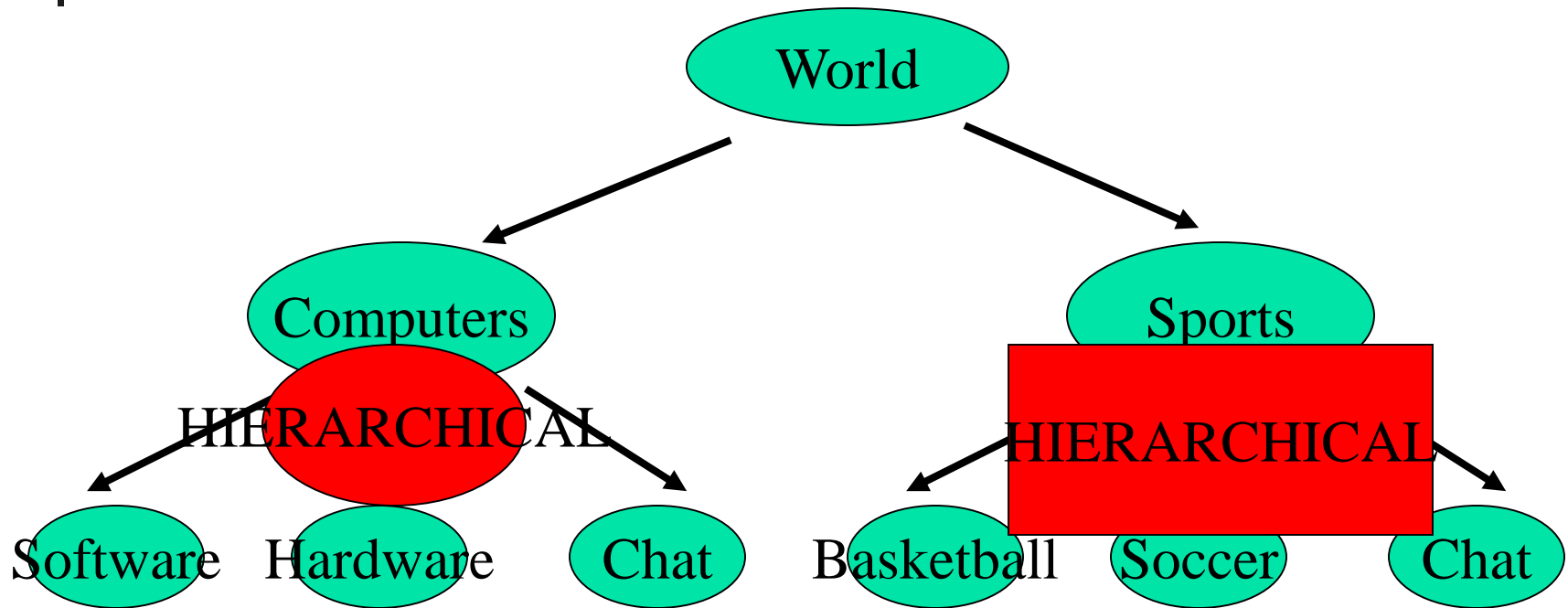
Training Set: LookSmart Web Directory

- **LookSmart Directory Structure (spring 99)**
  - 371k pages; 17k categories; 7 levels
  - 13 top-level categories; 150 second-level categories
  - *Training Set:* ~50k pages; chosen randomly from all cats

- **Example Categories**

Automotive	Buy or Sell a Car
Business & Finance	Chat
Computers & Internet	Finance & Insurance
Entertainment & Media	Magazines & Books
Health & Fitness	Maintenance & Repair
Hobbies & Interests	Makes, Models & Clubs
Home & Family	Motorcycles
People & Chat	New Car Showrooms
Reference & Education	Off-Road, 4X4 & RVs
Shopping & Services	Other Auto Interests
Society & Politics	Shows & Museums
Sports & Recreation	Trucks & Tractors
Travel & Vacations	Vintage & Classic

Training Classifiers - Expts

- **Experiments to build accurate classifiers**
  - Model: SVM vs LDA … SVM slightly better; more efficient
  - Number of features/category (500-1500) … small diffs
  - Other parametric variations (c, p) … small diffs
  - Binary vs. rate-based representation … small diffs
  - Length of document (full vs. summary)
    - Train full - Test full   71% avg break-even
    - Train full - Test sum   51% avg break-even
    - Train sum - Test sum   61% avg break-even
  - Hierarchical vs. flat category structure
    - Hierarchical – small accuracy advantages; large efficiency advantages

-> Hierarchical SVM, 1000 binary features/class, summaries