Classified Displays of Web Search Results

Susan Dumais, Ed Cutrell (Microsoft Research)  Hao Chen (UC Berkeley)

ASIS&T SIG/CR Classification Research Workshop, 11/12/00
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

Train (offline)

Classify (online)

manually classified web pages → SVM model → web search results → local search results...

...
Online Classification

Query: “jaguar”

Standard List Results

Automobile
Animal
Sports

Category Org (SWISH)

Category Tags, on the fly

Query: "jaguar"
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
  - Category hierarchy
Expansion of Category Structure

- Automotive

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

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

- Makes, Models & Clubs
  - (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
- (99) H.D. Rogers & Sons Auto Parts
- (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
- (83) Tom's Collection of Jaguar Mark II Photos
Chen & Dumais (CHI’2000)

Interface Conditions

Category Interface

Query: Jaguar
Retrieved 100 documents

- Computers & Internet
  - Clan Jaguar Quake & Quake 2 Clan
  - Atari Jaguar System
  - Atari - Jaguar Order Form
  - Jaguar XKS Screen Saver

- Automotive
  - H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - Jaguar Club of Florida
  - Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automobiles
  - A&L Luxury Car Center - Jaguar Main Page

- Entertainment & Media
  - Tony’s Collection of Jaguar Mark II Photos
  - MacTag’s Jaguar Page
  - The Jaguar Photo Gallery

- Travel & Vacations
  - Classic Car Source -- Welsh Jaguar Classic Car Museum

- Business & Finance
  - Jaguar Consulting, Inc.

- Shopping & Services

List Interface

Query: Jaguar
Retrieved 100 documents

- Search Results
  - H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - Jaguar Clubs of North America
  - Welcome to Jaguar
  - Terry’s Jaguar Parts
  - Clan Jaguar Quake & Quake 2 Clan
  - Jaguar Club of Florida
  - Jaguar Daimler Heritage Trust
  - Classic Car Source -- Welsh Jaguar Classic Car Museum
  - A.C. Imports Jaguar - Palatine IL
  - Atari Jaguar System
  - Jaguar Underground Depot
  - Jaguar Owners Web Ring
  - Jaguar, new cars, used cars, and accessories
  - Jaguar Sovereign
  - Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automobiles
  - A&L Luxury Car Center - Jaguar Main Page
  - Tony’s Collection of Jaguar Mark II Photos
  - MacTag’s Jaguar Page
  - Welcome to Jaguar Homes!
  - The Jaguar Photo Gallery
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:

<table>
<thead>
<tr>
<th>Question</th>
<th>Category</th>
<th>List</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was easy to use this software.</td>
<td>6.4</td>
<td>3.9</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I liked using this software</td>
<td>6.7</td>
<td>4.3</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I prefer this to my usual Web Search engine</td>
<td>6.4</td>
<td>4.3</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>It was easy to get a good sense of the range of alternatives</td>
<td>6.4</td>
<td>4.2</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I was confident that I could find information if it was there.</td>
<td>6.3</td>
<td>4.4</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>The &quot;More&quot; button was useful</td>
<td>6.5</td>
<td>6.1</td>
<td>n.s.</td>
</tr>
<tr>
<td>The display of summaries was useful</td>
<td>6.5</td>
<td>6.4</td>
<td>n.s.</td>
</tr>
</tbody>
</table>
## Use of Interface Features

- **Average Number of Uses of Feature per Task**

<table>
<thead>
<tr>
<th>Interface Features</th>
<th>Category</th>
<th>List</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expansing / Collapsing Structure</td>
<td>0.78</td>
<td>0.48</td>
<td>p&lt;.003</td>
</tr>
<tr>
<td>Viewing Summaries in Tooltips</td>
<td>2.99</td>
<td>4.60</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>Viewing Web Pages</td>
<td>1.23</td>
<td>1.41</td>
<td>p&lt;.053</td>
</tr>
</tbody>
</table>
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
- Decompose Category
  - Remove Page Titles (Browsing)
  - 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, $p<0.01$
Browsing slower than Category inline, $p<0.01$

Browsing faster than List with CatNames, $p<0.02$
<table>
<thead>
<tr>
<th>Category</th>
<th>Mean Log Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cat Hover</td>
<td>5.27</td>
</tr>
<tr>
<td>Cat Inline</td>
<td>5.10</td>
</tr>
<tr>
<td>Cat w/o Names</td>
<td>5.55</td>
</tr>
<tr>
<td>Cat Browse</td>
<td>5.08</td>
</tr>
<tr>
<td>List Hover</td>
<td>5.70</td>
</tr>
<tr>
<td>List Inline</td>
<td>5.55</td>
</tr>
<tr>
<td>List w/ CatNames</td>
<td>5.98</td>
</tr>
</tbody>
</table>

The table above shows the mean log time for different categories of interaction with the interface. The categories include Cat Hover, Cat Inline, Cat w/o Names, Cat Browse, List Hover, List Inline, and 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: $p(\text{category for each page})$
  - 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)
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.
We're a new kind of data and Internet company.

294,652 items for:

Jaguar

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
   Commercial site: http://www.ctrouve.com/
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 WebCounter. 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
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

You asked: jaguar

I have found answers to the following questions:

- Where can I find information on the animal Jaguar?
- Where can I find information on an exotic Jaguar XJ220?
- Where can I find the Web site for the company Jaguar?
- Where can I find car reviews for Jaguars?
- Where can I find information on the new Jaguar S-Type?

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

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

- Jag-lovers - the Jaguar Enthusiasts' premier resource on the Internet with Jaguar XJ6L
  The Premier Jaguar Cars resource and meeting place homepage on the web for all enthusiasts

More >>

- 7 matches by AltaVista Jaguar Cars Global Home Page
- 10 matches by WebCrawler AGH Museum – Atari Jaguar & Lynx
Search Results
Click to show all results sorted by rank

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
What do you want to find?   Search tips

Search Results
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
Query: jaguar
Retrieved 100 documents

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

- Travel & Vacations
  - (84) Classic Car Source -- Welsh Jaguar Classic Car Museum

- Shopping & Services
  - SubCateg
  - (More (3))
Query: jaguar
Retrieved 100 documents

- Search Results
  - jaguar watches - jaguar watches, tag heuer, omega, rolex, cartier, breitling
  - Jaguar Club of Florida
  - Jaguar Underground Dox
  - H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - Jaguar Clubs of North America
  - Welcome to Jaguar
  - Terry's Jaguar Parts
  - The Jaguar Photo Gallery
  - Atari - Jaguar Order Form
  - Jaguar Jottings
  - Jaguar 64 JUST CLAWS
  - Clan Jaguar Quake & Quake 2 Clan
  - Jag-lovers Jaguar Cars Windows Wallpaper page
  - Jaguar Publishing
  - A.C. Imports Jaguar - Palatine IL
  - Classic Car Source -- Welsh Jaguar Classic Car Museum
  - Atari Jaguar-System
  - Jaguar Owners Web Ring
  - Jaguar, new cars, used cars, and accessories
  - Jaguar Sovereign
Query: jaguar
Retrieved 100 documents

- Automotive
  - (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
    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 California's 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
  - (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.atareltd.com/storjordrform.htm
Query: jaguar
Retrieved 100 documents

- Search Results
  - (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
    British & European AUTO PARTS- CLASSIC & MODERN, import/export, H.D. Rogers & Sons 3418 Barkdale Blvd, Bossier City Louisiana 71112 USA(318) 742-3651,
    http://www.hd Rogers.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
    Terry's Jaguar Home Page. Your one stop for all of your Jaguar XKE, XJ6 and XJS needs!
    http://www.terry3.com/jaguar.html
Query: jaguar
Retrieved 100 documents

- (98) Jaguar Club of Florida
- (96) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat
- (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

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

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

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

- (84) Classic Car Source -- Welsh Jaguar Classic Car Museum
Query: jaguar
Retrieved 100 documents

- Search Results

- (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.hdrosers.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
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)
Query: jaguar
Retrieved 100 documents

- Automotive
- Computers & Internet
  - Computer/Video Games
    - Clan Jaguar Quake & Quake 2 Clan
    - Atari Jaguar-System
    - The Jaguar Server Home Page
    - ATARI Jaguar
    - Telegames Jaguar Worms Info
    - Jaguar Paint Codes
    - Jaguar Racing Classic Sports Cars
    - Telegames Jaguar Breakout 2000 Info
    - Telegames Jaguar Zero 5 Info
  - Software & Downloads
    - Jaguar XK8 Screen Saver
    - Schrödinger, Inc. Jaguar Program
    - Atari Jaguar development kit
    - Jaguar - The Touch Screen Solution You Have Been Looking For
  - For the Industry
    - Jaguar Marketing Enterprises, Inc. -- Quality and Affordable Memory Testers
    - 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
  - \( P(L_2) > \text{threshold} \)

- Hierarchical
  - 13 top-level; 150 second-level classifiers
  - Use for both feature selection and model building
  - Combine classifiers from L1 and L2 using:
    - \( P(L_1) \times P(L_2 | L_1) > \text{threshold} \) \text{(multiplicative)}
    - \( P(L_1) > \text{threshold} \) \&\& \( P(L_2 | L_1) > \text{threshold} \) \text{(Boolean)}
### Best/Worst Categories

<table>
<thead>
<tr>
<th>Category Name</th>
<th>Best F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health &amp; Fitness/Drugs &amp; Medicines</td>
<td>0.841</td>
</tr>
<tr>
<td>Home &amp; Family/Real Estate</td>
<td>0.797</td>
</tr>
<tr>
<td>Reference &amp; Education/K-12 Education</td>
<td>0.781</td>
</tr>
<tr>
<td>Sports &amp; Recreation/Fishing</td>
<td>0.750</td>
</tr>
<tr>
<td>Reference &amp; Education/Higher &amp; Cont. Ed.</td>
<td>0.741</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category Name</th>
<th>Worst F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Society &amp; Politics/World Cultures</td>
<td>0.034</td>
</tr>
<tr>
<td>Home &amp; Family/For Kids</td>
<td>0.088</td>
</tr>
<tr>
<td>Computers &amp; Internet/News &amp; Magazines</td>
<td>0.122</td>
</tr>
<tr>
<td>Computers &amp; Internet/Internet &amp; the Web</td>
<td>0.131</td>
</tr>
<tr>
<td>Business &amp; Finance/Business Professions</td>
<td>0.133</td>
</tr>
</tbody>
</table>
### Results

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Efficiency</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F1</td>
<td>training (sec)</td>
</tr>
<tr>
<td><strong>L1:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical, L1</td>
<td>0.649</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>L2:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hierarchical</td>
<td>0.476</td>
<td>729</td>
<td></td>
</tr>
<tr>
<td>Hierarchical, P(L1)*P(L2</td>
<td>L1)</td>
<td>0.495</td>
<td>1258 + 128</td>
</tr>
<tr>
<td>Hierarchical, P(L1)&amp;&amp;P(L2)</td>
<td>0.497</td>
<td>1258 + 128</td>
<td>13 + 11 categ</td>
</tr>
<tr>
<td>Hierarchical, P(L1)=1</td>
<td>0.711</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 -&gt; (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: 
    \[
    \min_{\|\mathbf{w}\|^2} \quad \mathbf{w} \cdot \mathbf{x} - b \leq 1, \quad \mathbf{w} \cdot \mathbf{x} - b \geq -1
    \]
  - Classify new items using: 
    \[
    f(\mathbf{w} \cdot \mathbf{x})
    \]
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
  Business & Finance
  Computers & Internet
  Entertainment & Media
  Health & Fitness
  Hobbies & Interests
  Home & Family
  People & Chat
  Reference & Education
  Shopping & Services
  Society & Politics
  Sports & Recreation
  Travel & Vacations

  Buy or Sell a Car
  Chat
  Finance & Insurance
  Magazines & Books
  Maintenance & Repair
  Makes, Models & Clubs
  Motorcycles
  New Car Showrooms
  Off-Road, 4X4 & RVs
  Other Auto Interests
  Shows & Museums
  Trucks & Tractors
  Vintage & Classic
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