Web Search Basics
Introduction to Information Retrieval
INF 141/ CS 121
Donald J. Patterson

Content adapted from Hinrich Schütze
http://www.informationretrieval.org
• I’m married and have four kids and live in UniHills
• I’m pretty open about my story
My Background

- B.S., Computer Science from Cornell (NY)
- M.Eng, Electrical Engineering
- STRIKE Officer
- USS CURTIS WILBUR, DDG-54, Japan
- Operations Officer
- USS SIMON LAKE, AS-33, Italy
- Ph.D in Computer Science and Engineering at UW
- Director of the Laboratory for Ubiquitous Computing and Interaction
- Co-founder of quub, whisper.fm, swayr, waitscout, audia
My Research

- Artificial Intelligence + Collapse Informatics + Ubiquitous Computing
- Example: I made a sensor system for babies in the NICU to detect cerebral palsy
Overview

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
  - Size of the Web
- Web Users
- Spam
Search use ...
(iProspect Survey, 4/04,
http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf)
### Who uses search?

<table>
<thead>
<tr>
<th>% of online adults in each group who use search engines</th>
<th>% of each group who ever use search engines</th>
<th>% of each group who used a search engine yesterday</th>
</tr>
</thead>
<tbody>
<tr>
<td>All online adults</td>
<td>91%</td>
<td>59%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>90</td>
<td>59</td>
</tr>
<tr>
<td>Female</td>
<td>92</td>
<td>60</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>93*</td>
<td>63*</td>
</tr>
<tr>
<td>African American</td>
<td>89*</td>
<td>44</td>
</tr>
<tr>
<td>Hispanic</td>
<td>79</td>
<td>44</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>96</td>
<td>66*</td>
</tr>
<tr>
<td>30-49</td>
<td>91</td>
<td>65*</td>
</tr>
<tr>
<td>50-64</td>
<td>92</td>
<td>52*</td>
</tr>
<tr>
<td>65+</td>
<td>80</td>
<td>38</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>78</td>
<td>34</td>
</tr>
<tr>
<td>High school</td>
<td>88*</td>
<td>45*</td>
</tr>
<tr>
<td>Some college</td>
<td>94*</td>
<td>65*</td>
</tr>
<tr>
<td>College graduate</td>
<td>95*</td>
<td>74*</td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $30,000</td>
<td>84</td>
<td>45</td>
</tr>
<tr>
<td>$30,000 - $49,999</td>
<td>93*</td>
<td>54*</td>
</tr>
<tr>
<td>$50,000 - $74,999</td>
<td>97*</td>
<td>66*</td>
</tr>
<tr>
<td>$75,000+</td>
<td>95*</td>
<td>76*</td>
</tr>
</tbody>
</table>
Google is far and away the search engine of choice, preferred by 83% of search users.

% of search users who answered the question: Which search engine do you use MOST OFTEN?

- None/DK: 8%
- Other: 19%
- Yahoo: 26%
- Google: 47%
- None/DK: 5%
- Other: 6%
- Yahoo: 6%
- Google: 83%

Some search users’ experiences vary by age

% of each group who have experienced each of the following...

- Found a really obscure fact or piece of information you didn’t think you’d be able to find:
  - 18-29 [n=314]: 48%
  - 30-49 [n=508]: 46%
  - 50+ [n=756]: 55%*

- Got conflicting or contradictory information in results and could not figure out what was correct:
  - 18-29 [n=314]: 51%
  - 30-49 [n=508]: 41%
  - 50+ [n=756]: 34%

- Got so much information in a set of search results that you felt overwhelmed:
  - 18-29 [n=314]: 35%
  - 30-49 [n=508]: 37%
  - 50+ [n=756]: 42%*

- Discovered really critical information was missing from search results:
  - 18-29 [n=314]: 37%
  - 30-49 [n=508]: 35%
  - 50+ [n=756]: 29%*

Source: The Pew Research Center’s Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults, age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. An asterisk (*) indicates a significant difference at the 95% confidence level.
Views of targeted advertising vary by age and income

Which of the following statements comes closest to how you, personally, feel about TARGETED ADVERTISING being used online – even if neither is exactly right?

As asked of adult internet users [n=1,729]

<table>
<thead>
<tr>
<th>Age Group</th>
<th>I’m NOT OKAY</th>
<th>I’m OKAY</th>
<th>Neither (VOL)</th>
<th>DK/Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-29</td>
<td>59%</td>
<td>36%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>30-49</td>
<td>65%</td>
<td>32%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>50-64</td>
<td>78%</td>
<td>19%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>65+</td>
<td>72%</td>
<td>21%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>&lt;30K</td>
<td>58%</td>
<td>39%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>30K to &lt;50K</td>
<td>68%</td>
<td>29%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>50K to &lt;75K</td>
<td>74%</td>
<td>23%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>75K+</td>
<td>68%</td>
<td>29%</td>
<td>2%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Pew Internet 2012
Help your students become better searchers

Web search can be a remarkable tool for students, and a bit of instruction in how to search for academic sources will help your students become critical thinkers and independent learners.

With the materials on this site, you can help your students become skilled searchers—whether they’re just starting out with search, or ready for more advanced training.

Lesson Plans & Activities
Download lesson plans to develop your students’ search literacy skills.
Browse lesson plans

Power Searching
Improve your search skills and learn advanced tips with online lessons and activities.
Start now

A Google a Day Challenges
Put your students’ search skills to the test with these trivia challenges.
Browse challenges

Live Trainings
Join us for live search trainings or watch past trainings from search experts here at Google.
Start training