SEARCH ENGINES

Internet Management

OUTLINE

- What are they and how do they work?

- Invisible Web

- Optimizing your searches
- How to read search results

- Mastering Google

WHAT ARE THEY

- **Web Search Engine**- A tool designed to search for information on the World Wide Web. The search results are usually presented in a list and are commonly called hits. The information may consist of web pages, images, information and other types of files.

HOW THEY WORK

Web Crawling • Indexing • Searching

- **Web Crawling**- Web search engines work by storing information about many web pages, which they retrieve from the WWW itself. These pages are retrieved by a Web crawler (sometimes also known as a spider) – An automated Web browser which follow every link it sees. The contents of each page are then analyzed to determine how it should be indexed (for example, words are extracted from the titles, headings, or special fields called meta tags.)

- **Indexing**- When a user enters a query into a search engine (typically by using key words), the engine examines its index and provides a list of best-matching web pages.

- **Searching-** The usefulness of a search engine depends on the relevance of the result set it gives back. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first.

HOW THEY WORK

- How a search engine decides which pages are the best matches, and what order the results should be shown in, *varies widely from one engine to another*. The methods also change over time as Internet usage changes and new techniques evolve

- Most Web search engines are supported by *advertising* revenue and, as a result, some employ the practice of allowing advertisers to pay money to have their listings ranked higher in search results. Those search engines which do not accept money for their search engine results make money by running search *related ads* alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads.

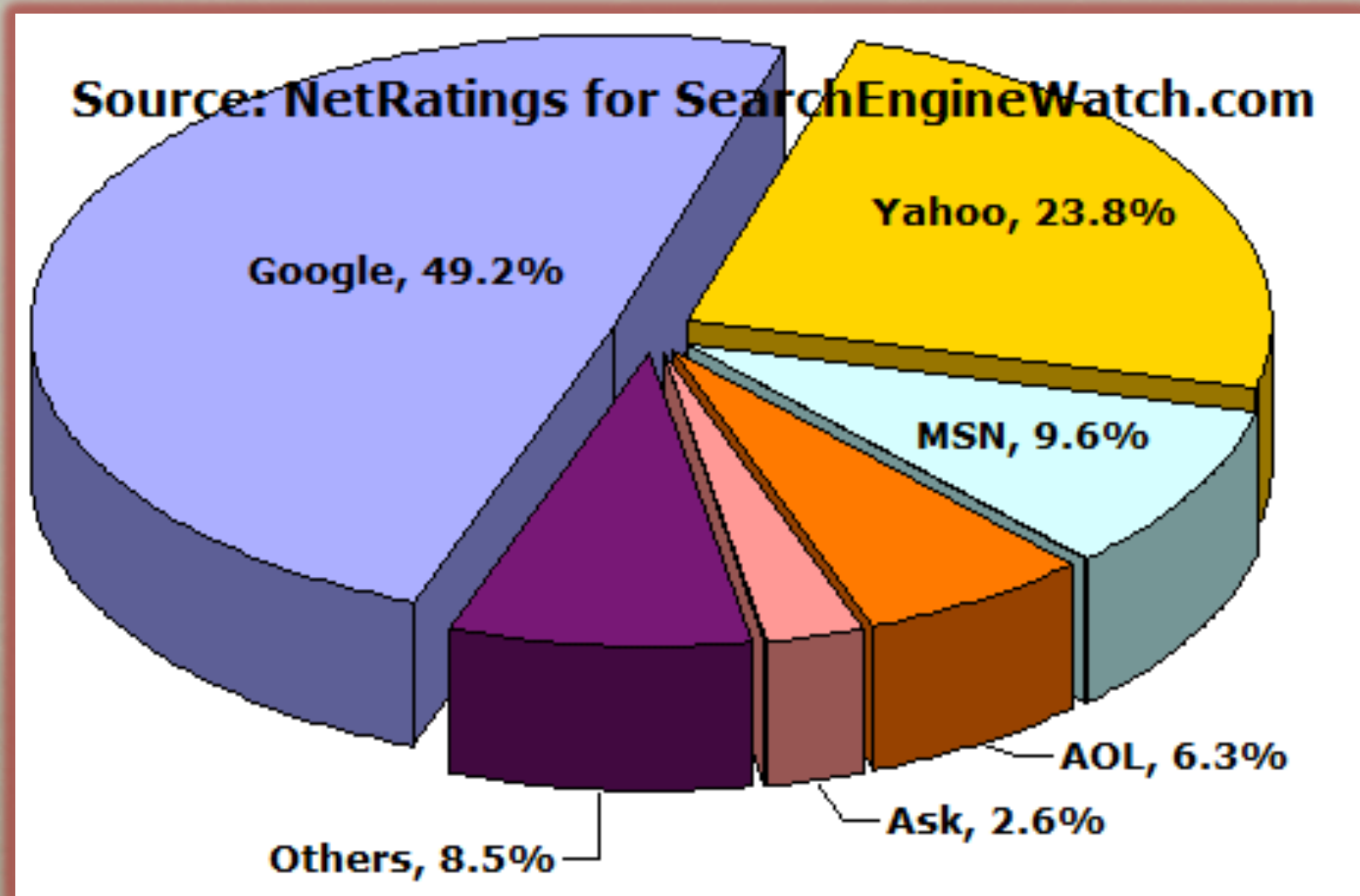# SEARCH ENGINE VS. SEARCH DIRECTORIES

- Search Engines use "spiders" (automated program) to reach thousands of pages per day, and index them for easy finding later.

  Google.com • Dogpile.com • Yahoo.com • AskJeeves.com

- Search Directories use human editors and reader submissions to hand-pick their cataloged content. With human reading being much slower than robot spiders, you can expect search directories to be much smaller than search engines. The human editor element, however, does add the filter of human judgment, which can help cut down the information you have to sift through when searching.

  Yahoo's Directoy • DMOZ.org • About.com

WHICH ONE SHOULD YOU USE?

Source: NetRatings for SearchEngineWatch.com

Google, 49.2%

Yahoo, 23.8%

MSN, 9.6%

AOL, 6.3%

Ask, 2.6%

Others, 8.5%

WHICH ONE SHOULD YOU USE?

- There are over 300 Internet search engines available for free use. All of them have their individual merits and they each use different techniques for how to catalog the billions of web pages around the world.

- "Which search tool do you personally prefer?"- Google.com has the least advertising on its screen, and the most indexed content of all the search engines. DMOZ is slower to load, but it has excellent depth of content. About.com has lots of advertising, but has amazing subject matter expertise…

  Whichever you personally prefer, every one of these search tools contains more content than you or I could ever read in a lifetime! The smart choice, accordingly, would be to test and compare these major search tools for yourself.

GOOD PRACTICES

- Compare searches on Search Engines to Search Directories. Combine the high volume of spidered content, and the hand-picked reviewing of human editors.

- Avoid the rut of relying on only one search engine

- Rotate your search tools

- Have perseverance and patience, and you will get good results.

- Visit sites which recommend specific search engines for specific topics. Copy and paste this link into a web browser: http://www.usa-people-search.com/content-how-to-choose-a-search-engine-or-directory.aspx

INVISIBLE WEB

- As powerful as search engines and directories are they do not index everything on the World Wide Web. They index less than 10% of the entire web! That remaining 90% is called the "Invisible Web".

  This is massive content that is publicly available, but hidden from regular search engines.

- These invisible Web pages are not the regular web pages you and I can make. Rather, these are dynamic database reports that exist only when called from large databases. (e.g.- Government discussions, Online library catalogues.)

  To search the Invisible Web use directories (Turbo10) and databases.

MOST COMMON INVISIBLE WEB PAGES

- Pages in non-html format (pdf, Word, Excel, PowerPoint), now converted into HTML.

- Script-based pages, whose URLs contain a ? Or other script coding

- Pages generated dynamically by other types of database software. These can be indexed if there is a stable URL somewhere that search engine crawlers can find.

Google current indexes only 12.5 billion out of the 220+ billion pages out there...less than 6% of all available internet content.

OPTIMIZING YOUR SEARCHES

Basic Facts:

- Every word matters. Generally, all the words you put in the query will be used. With some exceptions. "A" "for" "the"

- Search is always case insensitive. new york times = New York Time

- Wit some exceptions, punctuation is ignored (!#$%^&*)

Guidelines for a Better Search:

- Keep it simple!

- Use words that are most likely to appear on the page.

- Describe what you need with as few terms as possible.

- Use descriptive words

GOOD OR BAD?

- Zack's Pizza, salad and subs Randolph, MA **vs.**
  Zack's pizza, Randolph, MA

- My head hurts **vs.**
  Headache

- In what country are bats considered an omen of good luck **vs.**
  Bats good luck

- Weather report for Cancun Mexico **vs.**
  Weather Cancun

- Celebrity sounds **vs.**
  Celebrity ringtones

Keep it simple

On the page

On the page

Few terms

Descriptive words

OPTIMIZING SEARCH RESULTS

- **Advanced Search**

OPTIMIZING SEARCH RESULTS

- **Advanced Search on any search engine**

Winter **or** wheat

"Winter Wheat"

Wheat*****

Sink******

Winter **near** wheat

winter **and** wheat

Wheat **and not** barley

"Wheat and barley"

**To complete the Optimizing Your Search worksheet, copy and paste this address into a web browser:

http://www.sarc.montana.edu/search_help.html

HOW TO READ SEARCH RESULTS

1. Title

2. Snippet

3. URL

4. Cached Link

MASTERING GOOGLE

- Web
- Images
- Videos
- Maps
- News
- Shopping

- Gmail
- Calendar
- Documents
- Photos
- Reader
- Sites

Goooooooooogle ▶
1 2 3 4 5 6 7 8 9 10   Next
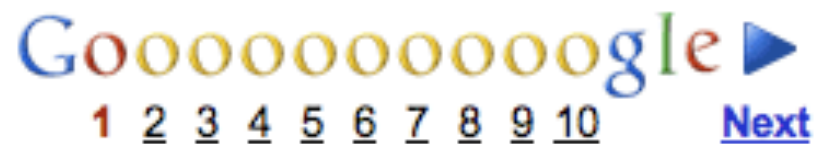
REVIEW

- Search Engine

- Search Directory

- Web Searching → Indexing → Searching

- How they make money

- Which one should you use

- Good practices

- Invisible Web

- Optimizing your searches

- How to read results

- Mastering Google