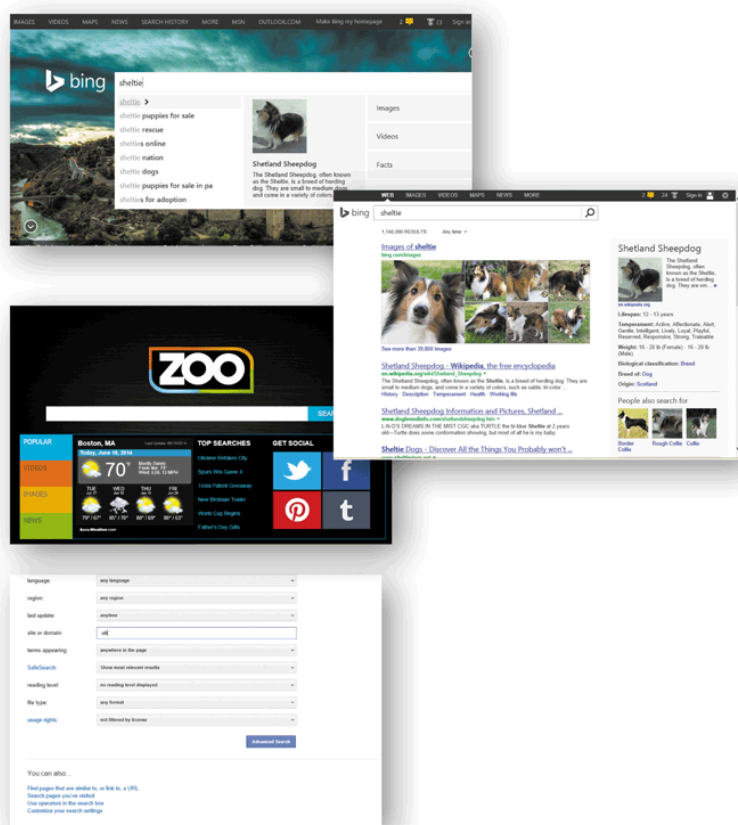


3

Searching the Web

(Lecture 5)



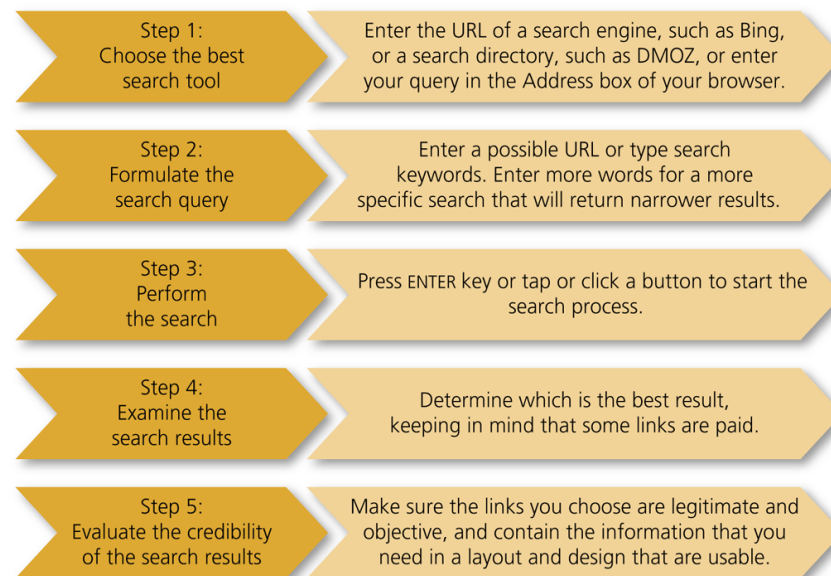
Discovering the Internet, 5th Edition

Objectives

- Describe how search engines work, and understand the search process
- Use different types of search tools and compare search results
- Apply search tool shortcuts and advanced features, including Boolean operators
- Perform searches using browser search features
- Identify and use specialized search tools

The Search Process

- Choose the best search tool
- Formulate the search query
- Perform the search
- Examine the search results
- Evaluate the credibility of the search results



The Search Process

- Choosing the Best Search Tool
 - Easy to use
 - Returns results quickly
 - Provides access to frequently updated large indexes and other web-based files
 - Presents most relevant search results returned
 - Clearly indicates paid or sponsored links

The Search Process

Search Engines

Content/Topic	Examples
Business	Business.com, GenieKnows, Justdial
Food/Recipes	RecipeBridge, Yummly
Job/Career	CareerBuilder, Hotjobs, Indeed, Glassdoor, Monster
Legal	Google Scholar, Lexis Nexis, WestLaw
Medical	Bing Health, Healthline, WebMD
Real Estate/Property	Realtor.com, Redfin, Trulia, Zillow
Maps	Google Maps, Mapquest, OpenStreetMap, Wikimapia
Question and Answer	Answers.com, eHow, wikiHow, WolframAlpha

The Search Process

➤ Formulating the Search Query

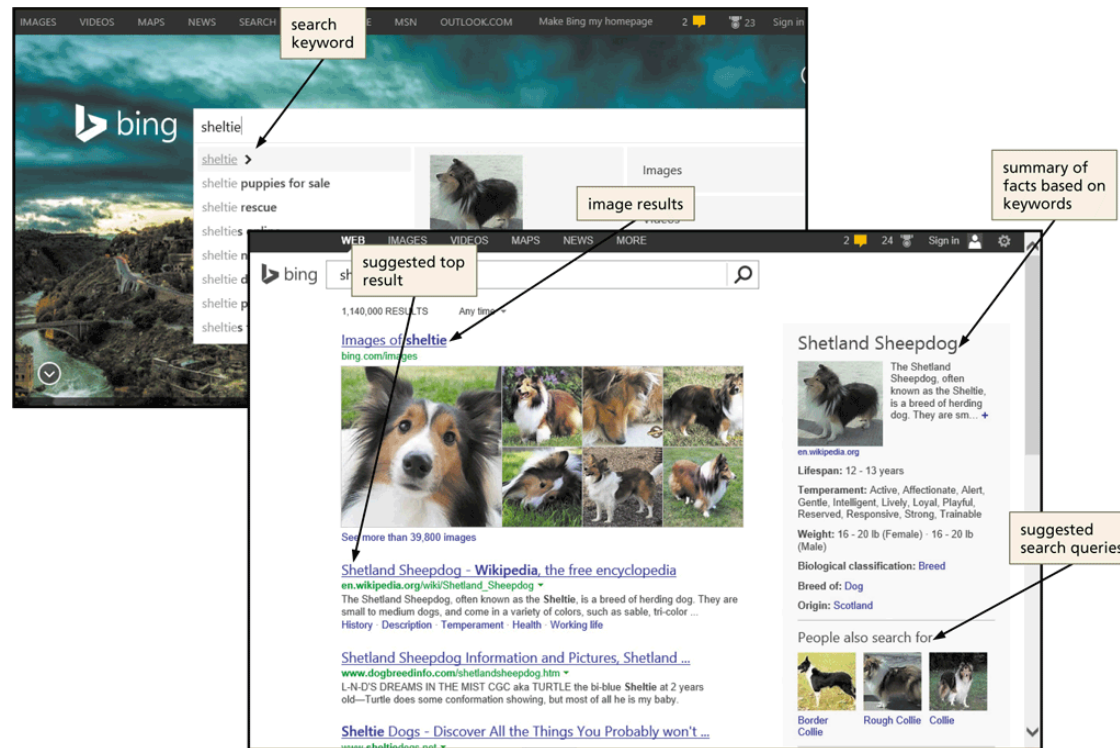
- The **search query** is the question that defines the information you seek
- A query should include at least one **keyword**, the specific word that describes that information
- Choose carefully, using specific rather than general keywords
- To perform a search, enter one or more keywords into the **search text box** or Address box

The Search Process

- Formulating the Search Query (continued)
 - A search tool then uses the keywords to identify relevant webpages and return a **search results list**
 - **Search results list** contains the URL, title, and description of and links to webpages
 - Each webpage item listed in search results list is a **hit**

The Search Process

- Formulating the Search Query (continued)
 - More keywords equal more focused results
 - Surround a phrase with quotation marks



The Search Process

- Formulating the Search Query (continued)
 - Ensure proper spelling
 - A **targeted search** seeks specific information using keyword combinations
 - An **open-ended search** seeks information on a broader scale using a simple keyword
 - A search that uses complete sentences is sometimes called a natural language search
 - Small, unimportant words in a natural language search query are called **stop words**, and are ignored

The Search Process

- Examining the Search Results
 - Different search tools return different search results for the same keyword
 - Be aware that paid or sponsored listing may appear at top of search results
 - People typically look only at the first 10 or 20 hits; search again with different keywords or search tool, if necessary

The Search Process

➤ Evaluating the Credibility of Search Results

– Authority

- Owner or author expertise
- Check top-level domains
- **Primary source** is any document, item, or other data that provides firsthand information about a topic

– Objectivity

- Fair
- Unbiased
- Not skewed toward commercial or political interests

The Search Process

➤ Evaluating the Credibility of Search Results (continued)

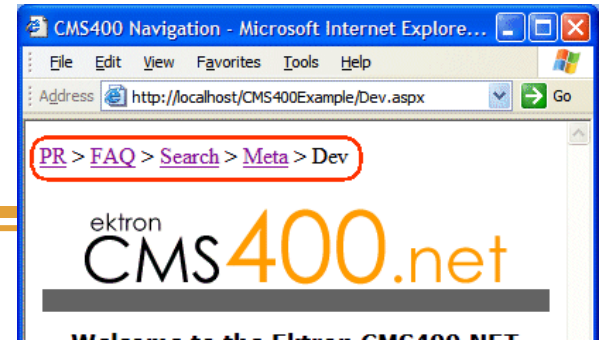
- Scope and Quality
 - Depth of coverage
 - Amount of detail provided
 - Accurate and up-do-date
 - Compares favorably with other pages on the same topic
 - Consider currency for breaking news
- Design and Functionality
 - Well-designed and easy-to-use page
 - **Broken links** indicate poor attention to detail

The Search Process

Evaluating the Credibility of Webpages

Area	Questions
Authority	<ul style="list-style-type: none">• Is this a primary source document?• Is the webpage's sponsoring organization or author a noted authority?• Are the webpages up to date?
Objectivity	<ul style="list-style-type: none">• Is the webpage objective?• Is any bias clearly stated?
Scope	<ul style="list-style-type: none">• What is the intended audience for this website?• How does the information on the webpage compare with others on the same topic?
Design and functionality	<ul style="list-style-type: none">• Does the webpage have a professional appearance?• Do all parts of the webpage work correctly?

Search Tools



➤ Directories

- **Directory** is a human-compiled, hierarchical list of webpages
- Yahoo! is one of the first directories
- Editors create an **index**, or list of webpages
- Organizes links in categories and subcategories
- Users increasingly **drilling down** through categories to find specific information
- A **breadcrumb trail** shows the hierarchical arrangement of categories and subcategories through which you have clicked

Search Tools

The image displays three sequential screenshots of the dmoz (Directory Mozilla) website, illustrating a search path for geocaching. Red arrows indicate the navigation flow from the main directory to the specific category.

Screenshot 1: Main Directory

- Arts
- Business
- Computers
- Games
- Health
- Home
- Kids and Teens
- News
- Regional
- Reference
- Science
- Shopping
- Society
- Sports
- World

Screenshot 2: Recreation -> Outdoors

Top: Recreation: Outdoors (11,206)

- [Chairs and Furniture](#) (13)
- [Directories](#) (44)
- [Equipment](#) (10)
- [Guides and Outfitters](#) (33)
- [Maps](#) (6)
- [News and Media](#) (46)
- [Organizations](#) (13)
- [Parks](#) (1)
- [Personal Pages](#) (16)
- [Schools and Education](#) (64)
- [Survival and Primitive Technology](#) (134)

Screenshot 3: Recreation -> Outdoors -> Geocaching

Top: Recreation: Outdoors: Geocaching (99)

- [Asia](#) (6)
- [Europe](#) (6)
- [North America](#) (13)
- [Oceania](#) (4)
- [Shopping](#) (16)
- [Software](#) (10)

See also:

- [Home: Consumer Information: Sports and Recreation: Outdoors: Global Positioning](#) (5)
- [Recreation: Outdoors: Letterboxing](#) (33)
- [Science: Earth Sciences: Geomatics: Global Positioning System](#) (33)
- [Sports: Orienteering](#) (48)

This category in other languages:

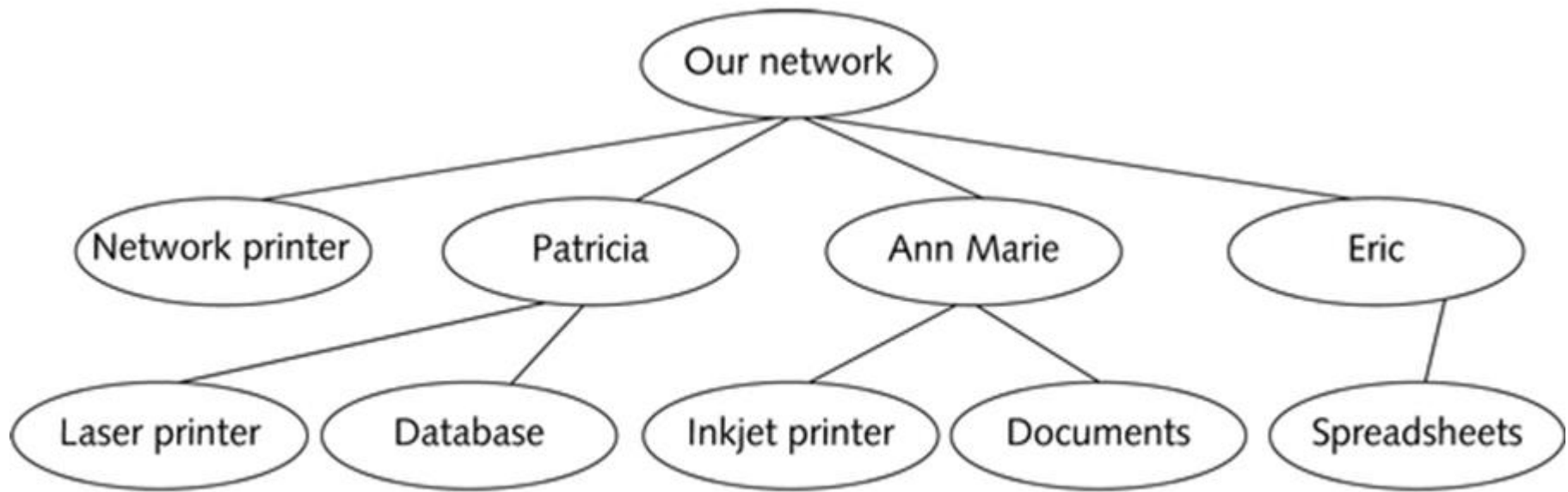
Czech (41)	Danish (5)
Dutch (8)	French (12)
German (147)	Italian (2)
Russian (1)	Slovakian (2)

Directory Server

- Stores information about people, hosts, and other resources on the network in directories and provides this information to computers on the network
- Requirements
 - To be able to handle a high volume of searches
 - To provide results for queries quickly

How Directories Work

- By following an upside-down tree structure

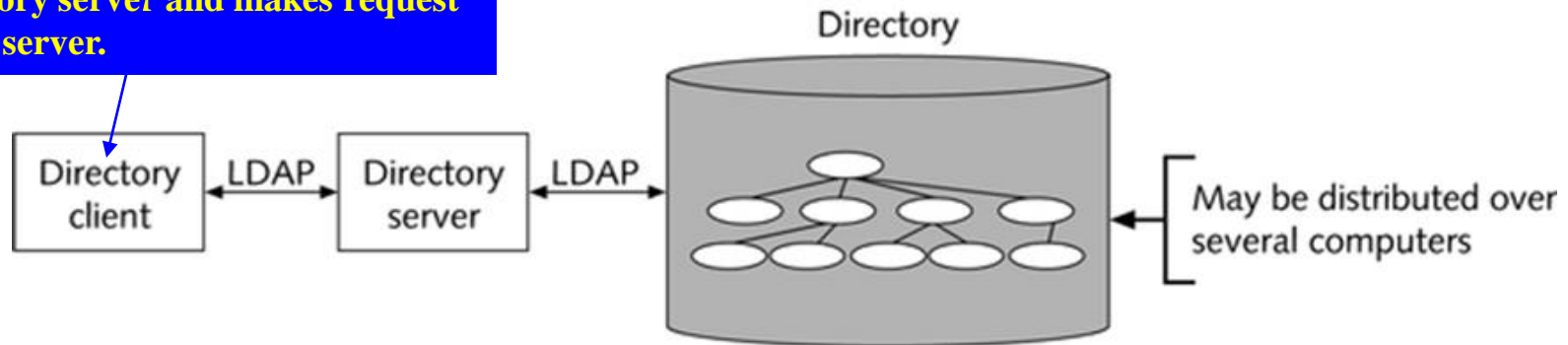


Information about resources on a network can be kept in a directory

How Directories Work

- Directory servers sometimes use a protocol called LDAP (Lightweight Directory Access Protocol) to access directories

When an application or user needs information from a directory, user's application acts as a client to the directory server and makes request to the server.



LDAP can be used by a program to query a directory

Using Directories

- Directories on the web are similar to search engines in the way they operate and provide information
- Example:

Search Tools

A spider, bot, or web crawler automatically browses the web going from link to link to add and update its searchable index

➤ Search Engines

- Use software called a **spider**, **bot**, or **web crawler**
- Typical information collected by a web crawler
 - Page title
 - URL
 - **Meta tag keywords**
 - Occurrence of keywords in page
 - All of the words on the page – **full-text searching**
 - Internal links
 - Number and relevancy of other webpages that link to the page

bot: short for robot

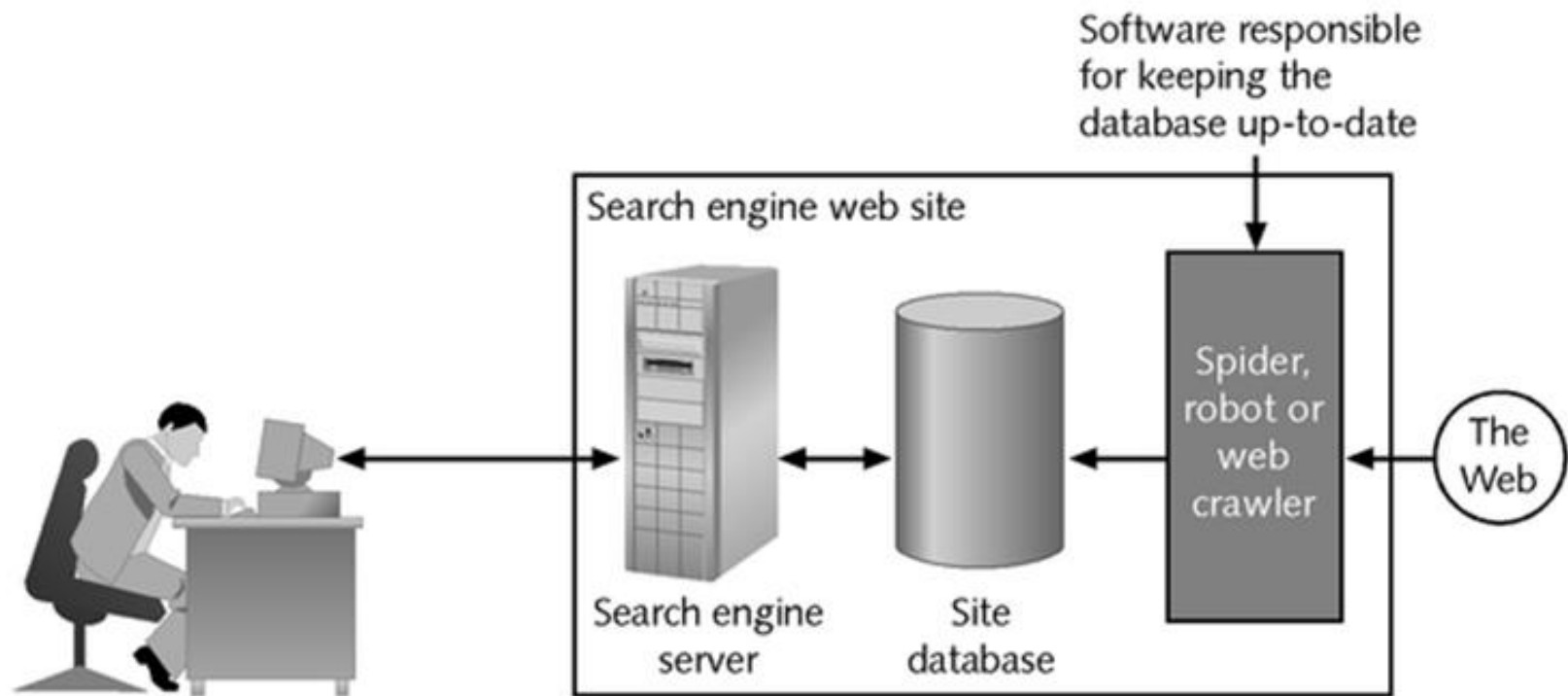
Search Tools

- Search Engines (continued)
 - Search engines use different algorithms to determine relevance of webpages in the search results
 - Most search engines also include paid or sponsored listings on the search results page
 - Many search tools today have become hybrids, basing results on indexes created by both web crawlers and human editors

Search Engine Web Sites

- Help users find information anywhere on the web
- Keep databases containing keywords and the URLs where the keywords are located
- Ways in which databases are created and maintained
 - A person enters keywords in a subject directory
 - Web crawlers, spiders, or web robots look for new information

Using a Search Engine Web Site



When you use a search engine web site to locate something on the Web, the search engine server searches the site's database, not the entire Web

Three-step process of Search Engines

➤ Three-step process

1. Crawling

- use software called web crawler, bot or spider to collect information from webpages

2. Indexing

- Index the data coming in from crawlers by using keywords

3. The search process

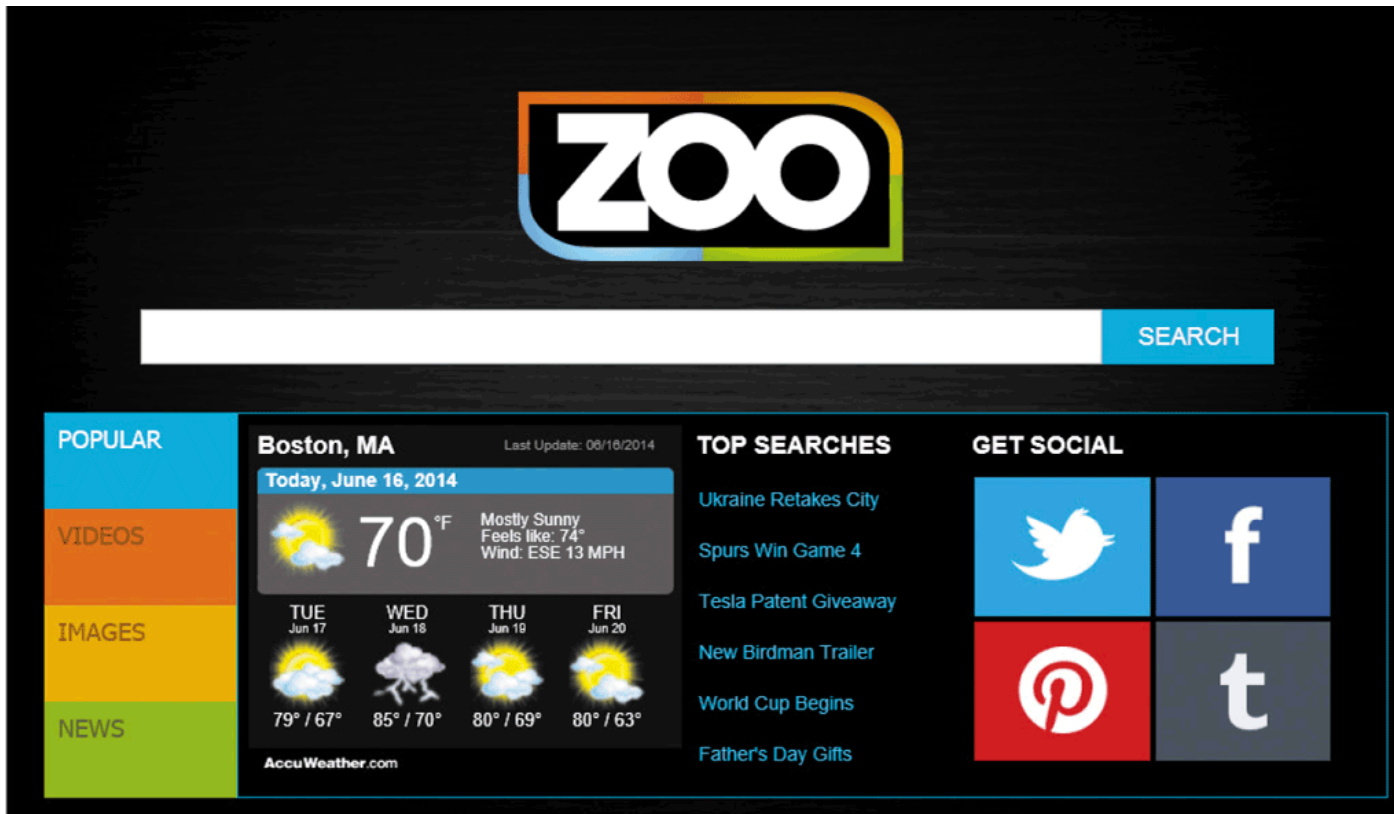
- Use the index created to look up the search term

Search Tools

➤ Metasearch Engines

- A **metasearch engine** compiles the search results from multiple search engines into a single search results list
 - Some eliminate duplicates
 - Watch for paid or sponsored listings in search results list
 - Dogpile
 - Mamma.com
 - Zoo
 - Ixquick

Search Tools



Advanced Search Techniques

➤ Refining Searches

- A **Boolean operator** specifies which keywords should be included or excluded from search results

Boolean Operators

Task	Procedure
Search for all the words in any order.	Type AND between keywords. Example: Canada AND nickel AND mines
Search for at least one of the words.	Type OR between keywords. Example: ocean OR sea
Search for a phrase in the given order.	Surround the phrase with quotation marks. Example: "Catalina yachts"
Exclude a concept from the search results.	Type NOT before the excluded word. Examples: orange NOT Florida or sometimes: kayak AND NOT inflatable

Advanced Search Techniques

➤ Refining Searches

- Boolean operators include:
 - AND
 - Keyword must be included
 - OR
 - At least one of multiple keywords must be included
 - NOT
 - Keyword must be excluded
- Major search engines offer searching shortcuts

Advanced Search Techniques

- Advanced Search Forms
 - Used to structure complex search queries
 - Prompts you to specify Boolean operators and other criteria
 - Can filter results by:
 - Language
 - File type
 - Domain
 - Great way to learn more about how to use specific search engine shortcuts and Boolean operators

Browser Searches

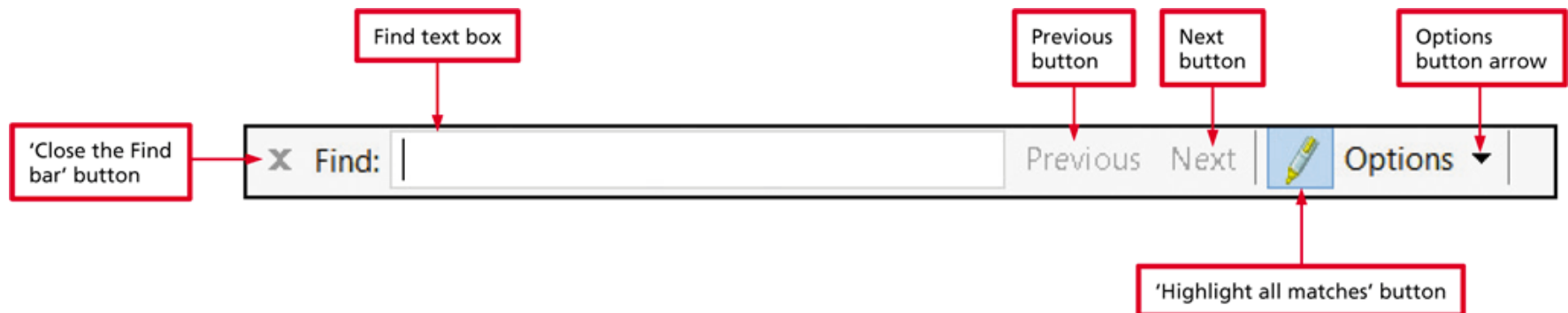
➤ Address Bar Searches

- Enter keywords in the Address box on the Address bar
- Press ENTER key or tap or click appropriate button to open search results page in the current tab

Browser Searches

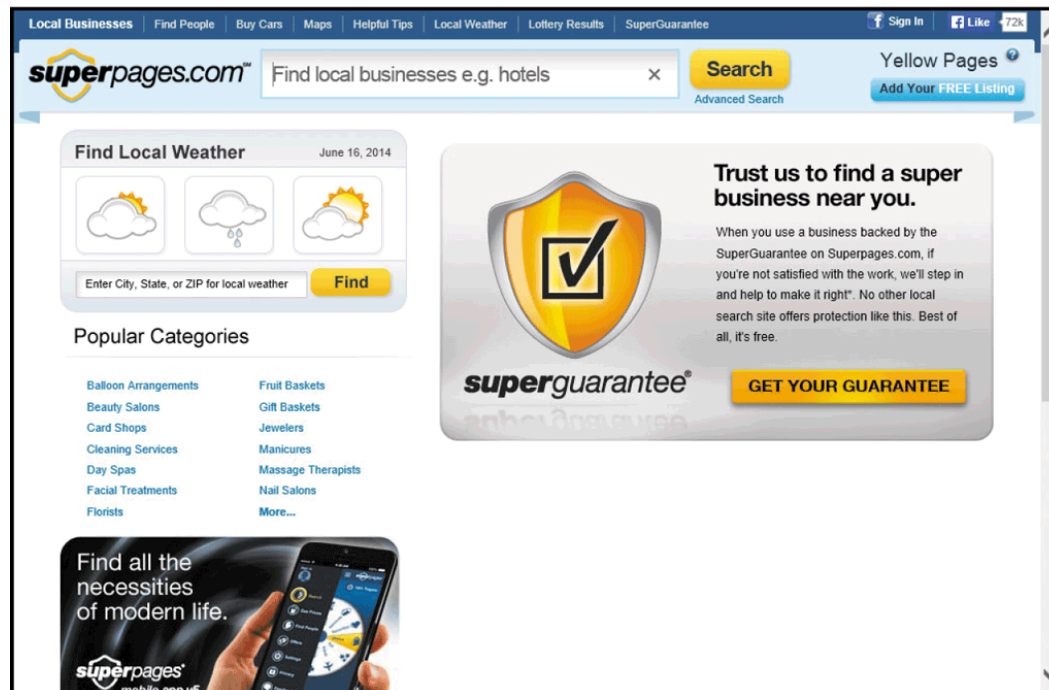
➤ Webpage Searches

- Find bar allows you to search the contents of a webpage
- Press CTRL+F to display the Find Bar



Specialized Searches

- People and Business Search
 - Use online white pages directory or online yellow pages directory



Specialized Searches

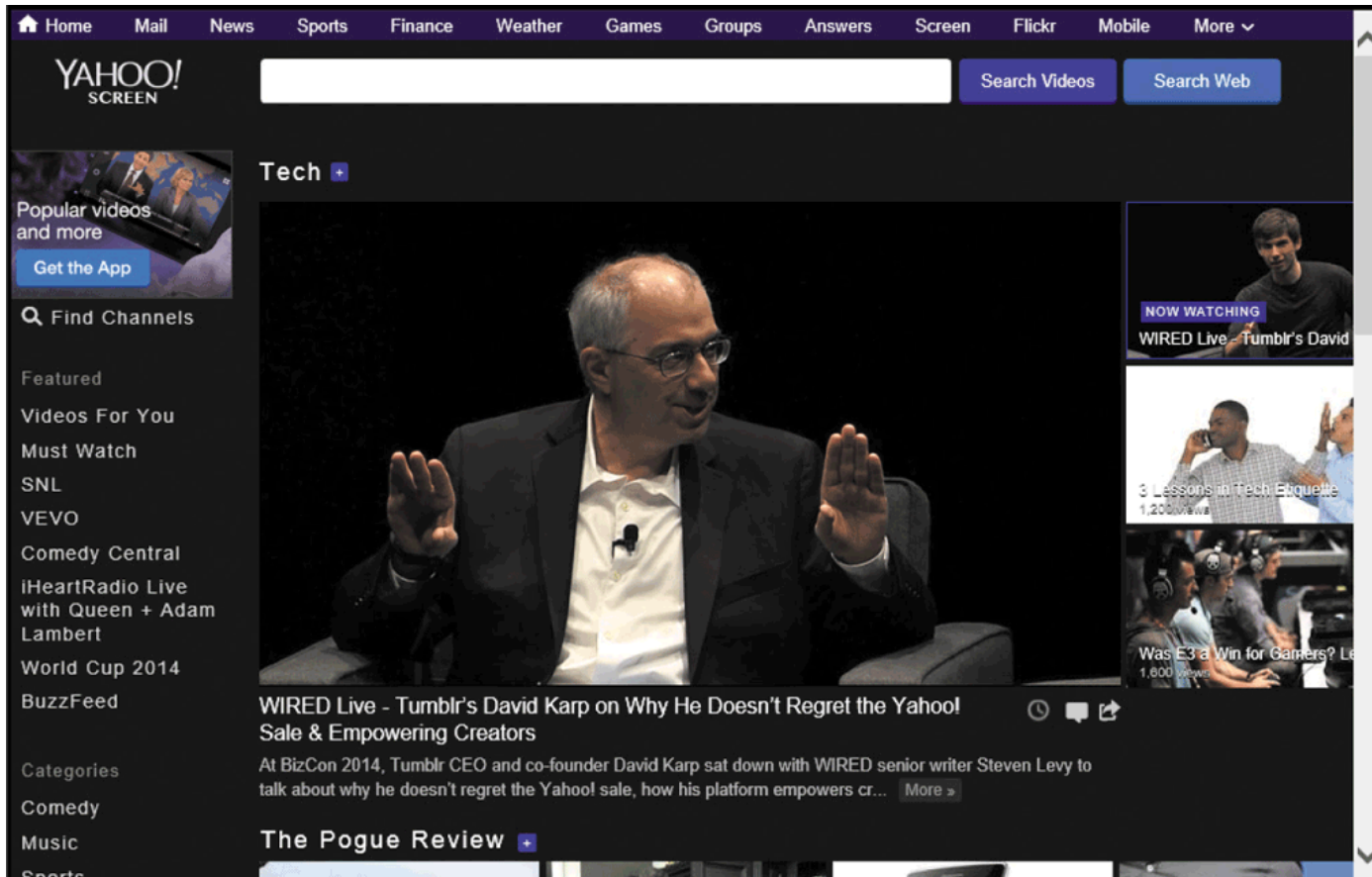
- People and Business Search (continued)
 - Free resources
 - Pipl
 - Wink
 - Fee-based (pay for additional information)
 - WhitePages
 - Superpages

Specialized Searches

- News and Current Event Search
 - MSN
 - Yahoo!
 - Associated Press (AP)
 - United Press International (UPI)
 - Reuters

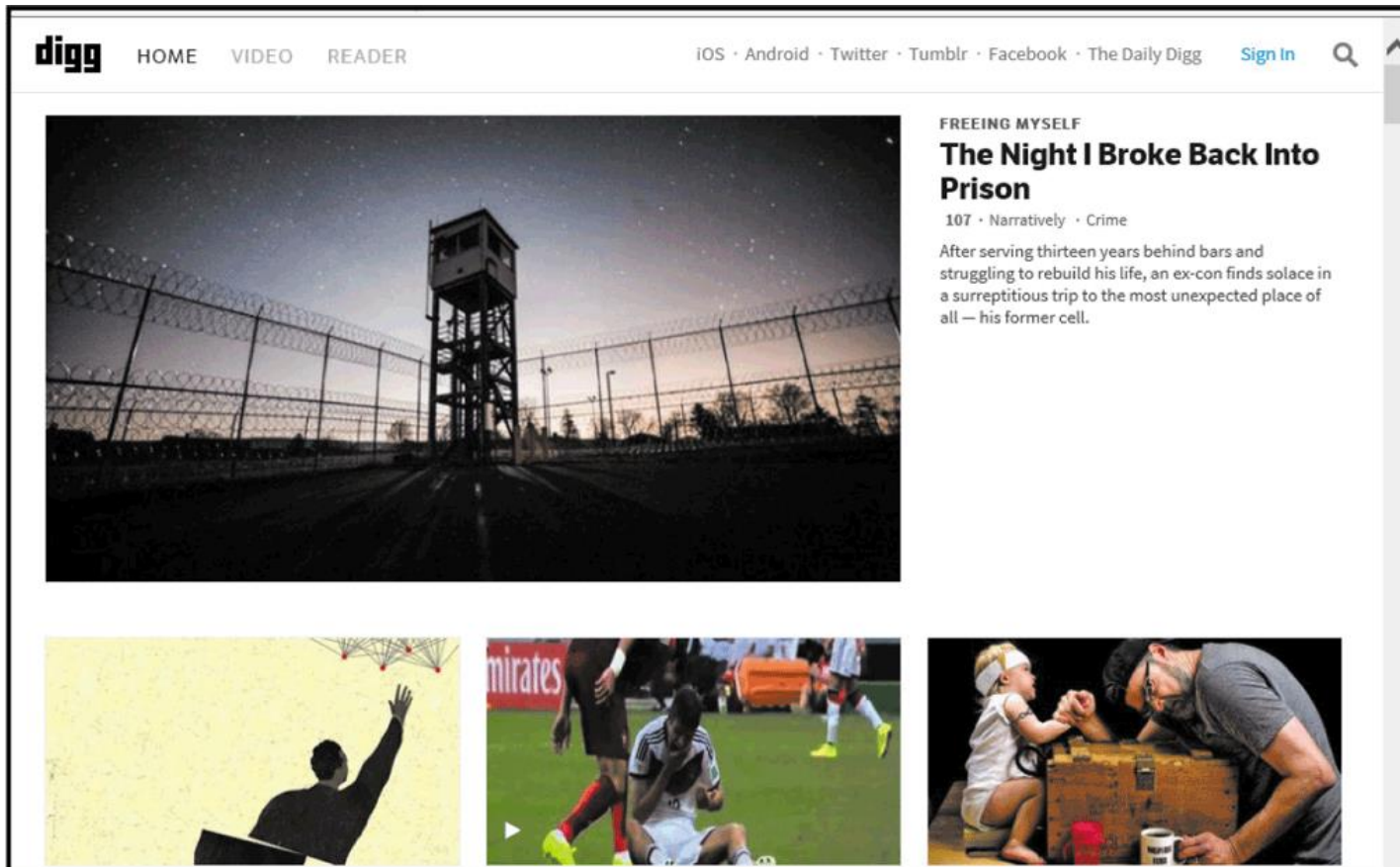
Specialized Searches

➤ Video Search



Specialized Searches

➤ Social Media Search



Specialized Searches

➤ Shopping Search

The screenshot displays the DealTime website interface. At the top, the DealTime logo is accompanied by the tagline "millions of deals on sale". A search bar with an orange "Search" button is positioned to the right. Below the logo, a navigation bar lists categories: Clothing & Accessories, Electronics, Home & Garden, Computers & Software, Sports & Outdoors, and More. A Facebook "Like" button with a count of 2.2k is also visible. The main section is titled "Deals of the Day" and features a large deal for a "Chroma Haiku 4-in. Paring Knife" with a -19% discount, priced at \$58.95 (down from 73.0) with free shipping. A "See Deal" button and a countdown timer showing "Time Left: 10:02:17" are present. To the right, a sidebar lists other deals: a Samsung UN55F7500 TV (29% off, \$1,997.99), a Chroma Haiku 4-in. Paring Knife (19% off, \$58.95), a Bataleon Feelbetter Snowboard (35% off, \$240.90), and a Speed Stacks 30 Set Sport Pack (15% off, \$569.99). At the bottom, a section titled "Most Popular Deals by Price" includes filters for "Under \$15", "\$15 - \$50", and "Over \$50", along with several deal cards showing discounts like -50%, -43%, -54%, -67%, and -47%.

Research Alternatives to Search Engines

- LexisNexis
 - Abstracts and articles on business, news, government, medical, and legal topics
- ingentaconnect
 - Online articles from academic and professional publications
- Hoover's (Dun & Bradstreet)
 - Profiles on public and private businesses, along with financial statements and analyses and other information
- FindLaw
 - Legal resources portal

Research Alternatives to Search Engines

- Proquest
 - Provides information services to the business, scientific, engineering, financial, and legal communities over the Internet or an internal intranet
- Government resources
 - Federal Citizen Information Center
 - Internal Revenue Service
 - U.S. Census Bureau
 - U.S. Bureau of Labor Statistics
 - THOMAS
 - U.S. Small Business Administration
 - U.S. National Park Service
 - USA.gov