

24. UIs for Search and IR

INFO 202 - 13 November 2012

Bob Glushko

Plan for Today's Lecture

The "Classical" Model of Search and the "Classical" UI for IR

Web-based Search

Best practices for UIs in query specification, query results, and query reformulation

Four big ideas for new Search UIs

An even bigger idea - "Watson-class" natural language UIs

A Range of Information Needs

Finding the citations to documents (so you could look for them in the physical library)

Answers to specific questions

Comparison of similar items

Familiarization (browsing and "building upon")

Knowledge discovery / data mining

... finding out what your friends are doing

The "Classical" IR System: Searching for Citations or Documents

Early in the digital era (1970s) large specialized bibliographic collections were created for academic articles, legal cases and opinions, news articles, etc.

The users of these systems were reference librarians, paralegals, journalists, and other professionals willing to be trained in using them

Since these systems predated the PC and graphical terminals, their user interfaces used "command lines"

The command languages were complex and powerful, typically supporting Boolean, adjacency, and term stemming operators

```
(PCR OR POLYMERASE(W)CHAIN(W)REACTION? OR DNA(W)SEQUENC?) AND  
(CANCER? OR PRECANCER? OR NEOPLASM? OR CARCINO?)
```

Query Specification in "Classic" Search UI

```
?s (biometric? and security)/TI,LP
    1621 BIOMETRIC?/TI,LP
    40268 SECURITY/TI,LP
S1      678 S (biometric? and security)/TI,LP
?s S1 and CS=(HARVARD AND MEDIC?)
S2      52 S1 and CS=(HARVARD AND MEDIC?)
?T S2/3/1-10
```

Challenges Posed by the "Classic" IR UI

Untrained users (e.g. with public access terminals in libraries) generally couldn't use these systems effectively on their own

But effective use is essential in narrow and specialized domains where high recall is mandated

Classic Query Specification "Webified" -- 1998

Search History
[Database Details](#)

Set	Term Searched	Items	
S1	MELANOMA OR SKIN (3N)CANCER?	1264	Display
S2	PREVENT? OR PROTECT?	28430	Display
S3	ULTRAVIOLET OR UV OR SUN OR SUNSCREEN	1886	Display

Format:
Number of Records:


Show Database Details for:


© 1998 The Dialog Corporation plc

Command | s s1 and s2 and s3

Classic Source and Field Selection "Webified" -- 1998

[↑ Top](#) : [↑ Medicine and Pharmaceuticals](#) : [↑ Research and Development](#) : [↑ Medicine](#) : [Articles by Author](#)

 **Targeted Search:** Articles by Author

▶ **Search Form** 

Sources BIOSIS® CAB Ei Compendex® EMBASE®
 INSPEC JICST MEDLINE® NTIS
 PASCAL SciSearch® All

Author Last First Initial

Author's Organization

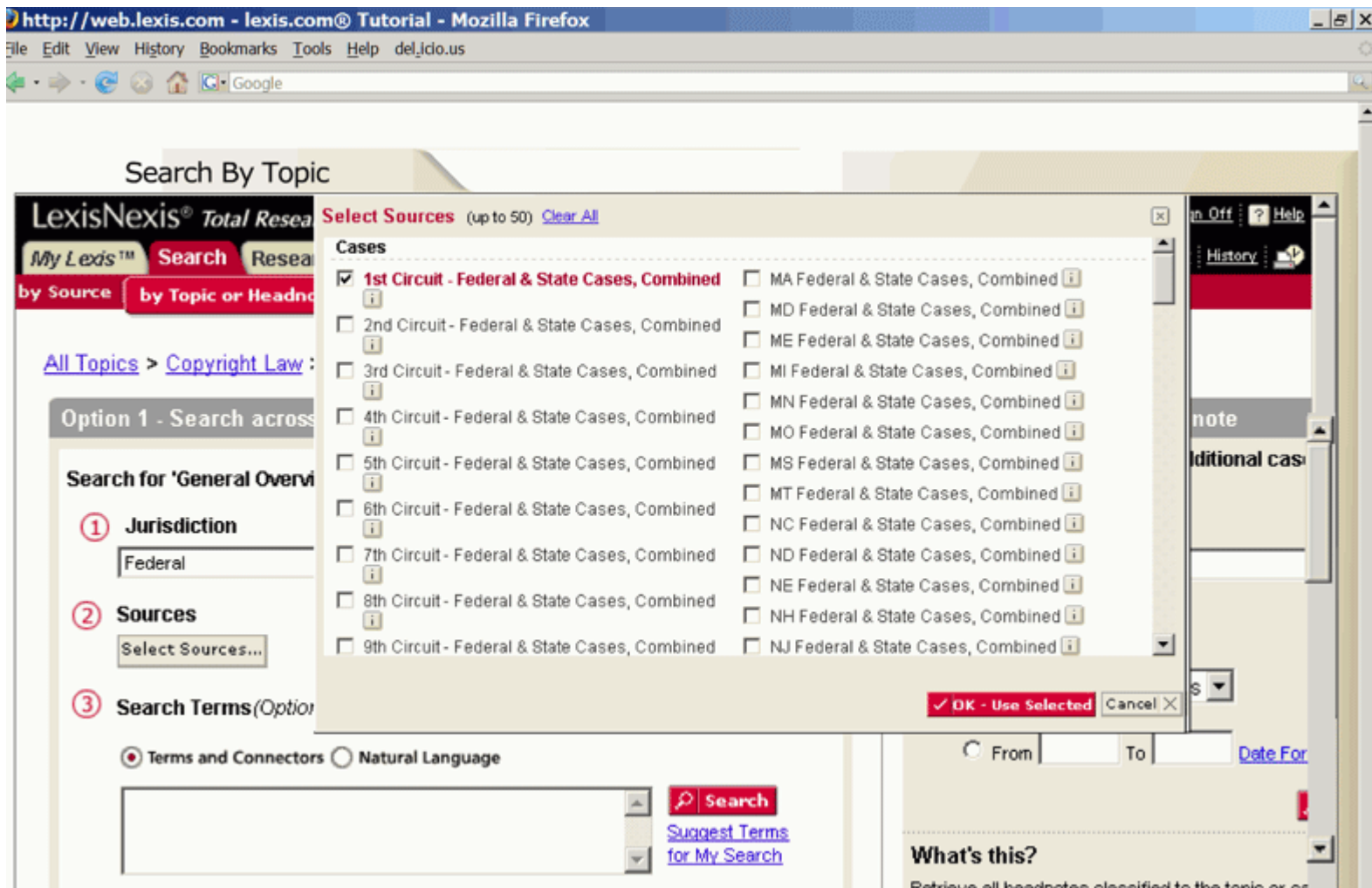
Words in Title

Publication

Language Any Language English Only

Year All Years From To

Classic Query Specification "Webified" -- 2007



LexisAcademic 2012

The screenshot displays the LexisNexis Academic website interface. At the top, the browser address bar shows the URL www.lexisnexis.com/hottopics/lnacademic/. The page title is "LexisNexis Academic".

General Searching

- » Easy Search™
- » Advanced Search

Tip: Click the headings below to view links to specialized search forms and other useful features.

Help LexisNexis
change the future of information and get rewarded
Learn more ▶

Navigation Menu: News, US Legal, International Legal, Companies, Country Information, Subject Areas, Sources, Guides & Resources, Mobile.

Easy Search™ Help Clear

Use of this service is subject to Terms and Conditions

Search the News (Covers 1980 to today.)
Search For: _____
By Source Type: Major World Publications
Or by Source Title: _____
Start typing a title like New York Times
» For more options like search by date or byline, use [Advanced Search](#) or [All News](#). **Go**

Look up a Legal Case (Covers US and state cases.)
By Citation like 347 U.S. 483
Or by Parties like *Mapp v. Ohio*
Or by Topic like Equal Opportunity
» Citations must be exact. [Click here](#) for citation help. For more options, try [Cases](#), [Law Reviews](#), or [Landmark Cases](#). **Go**

Get Company Info (Covers over 80 million companies.)
By Name like Microsoft
Or by Ticker like MSFT
» For more options like list companies by zip code or revenue, access the full [Company Dossier](#). **Go**

Hot Topics Links (Click the links below to see recent news articles.)
Today's Front Page News
Health Care
U.S. Presidential Election 2012
Syrian Uprising
Education
Immigration
Barack Obama
Mitt Romney
Human Trafficking
Hurricane Sandy
Environment

Research People (Covers public figures.)
Last Name (Required): _____
First Name (Optional): _____
Source: Biographical Reference
» Try also [People](#), or see [Browse Sources](#) for more People sources.

Combined Search (Search multiple kinds of content.)
Search For: 770 F. Supp.2d 666
Date: Previous 2 years
Major US & World news
Company profiles
SEC Filings
US & State Legal Cases
Law Reviews
» Try also [Advanced Search](#), [Sources](#), and [Help](#)

LexisAcademic 2012 - Reults

9 of 9 Shepardize®

770 F. Supp. 2d 666, *2011 U.S. Dist....

UNITED STATES DISTRICT COURT FOR THE SOUTHERN DISTRICT OF NEW YORK

770 F. Supp. 2d 666; 2011 U.S. Dist. LEXIS 29126; 79 Fed. R. Serv. 3d (Callaghan) 114; 98 U.S.P.Q.2D (BNA) 1229; Copy. L. Rep. (CCH) P30,057; 2011-1 Trade Cas. (CCH) P77,387

March 22, 2011, Decided
March 22, 2011, Filed

SUBSEQUENT HISTORY: Motion denied by, Class certification granted by [Authors Guild v. Google, Inc.](#), 2012 U.S. Dist. LEXIS 76080 (S.D.N.Y., May 31, 2012)

PRIOR HISTORY: [Authors Guild v. Google, Inc.](#), 2009 U.S. Dist. LEXIS 116175 (S.D.N.Y., Dec. 1, 2009)

CASE SUMMARY

OVERVIEW: In a lawsuit by plaintiff authors and publishers alleging copyright infringement from an Internet search engine's acts of digitally copying books and others writings without authorization, a [Fed. R. Civ. P. 23\(e\)](#) class action settlement was not approved because it would give the search engine a significant competitive advantage and reward it for engaging in wholesale copying of copyrighted works without permission while releasing claims well beyond those presented in the underlying case.

OUTCOME: Motion denied.

Searching the Web

The size and scope of the Web is vastly greater than any "classic" bibliographic or document collection

But the scope of what people search for is also vastly greater than in classical IR systems

People expect to get information and documents, rather than just citations

Challenges Posed by Searching the Web

UIs must accommodate differences among people in:

- Knowledge / life experience
- Cultural background and expectations
- Reading / scanning ability and style
- Methods of organizing and looking for things

UI functionality is constrained by heterogeneity of content (e.g., can't assume complete and consistent metadata or structure)

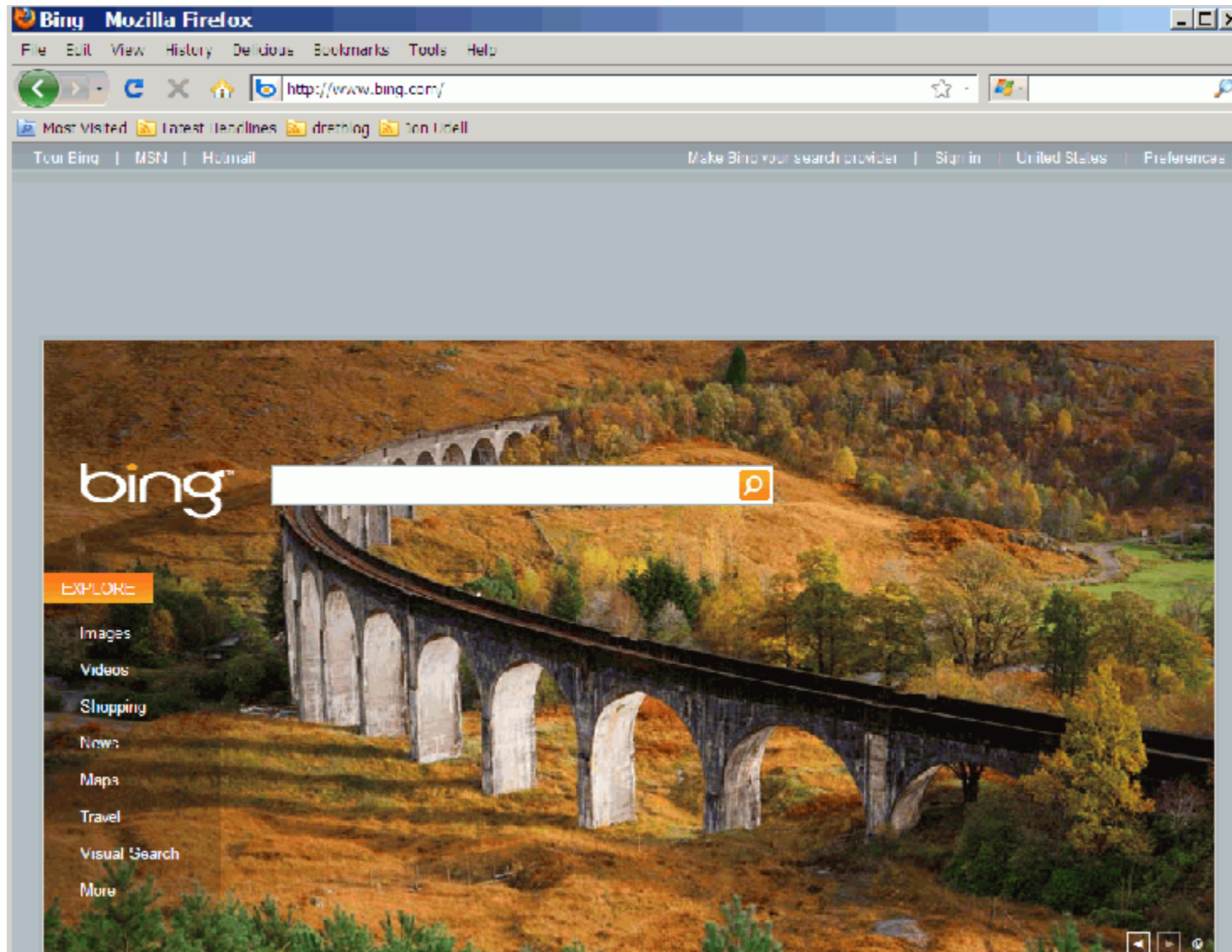
So what this means in practice is that the default web search UI is the simplest possible one: just a search box...

Additional challenges arise because people have multiple devices on which they want to search the Web and many have limited capabilities for UIs

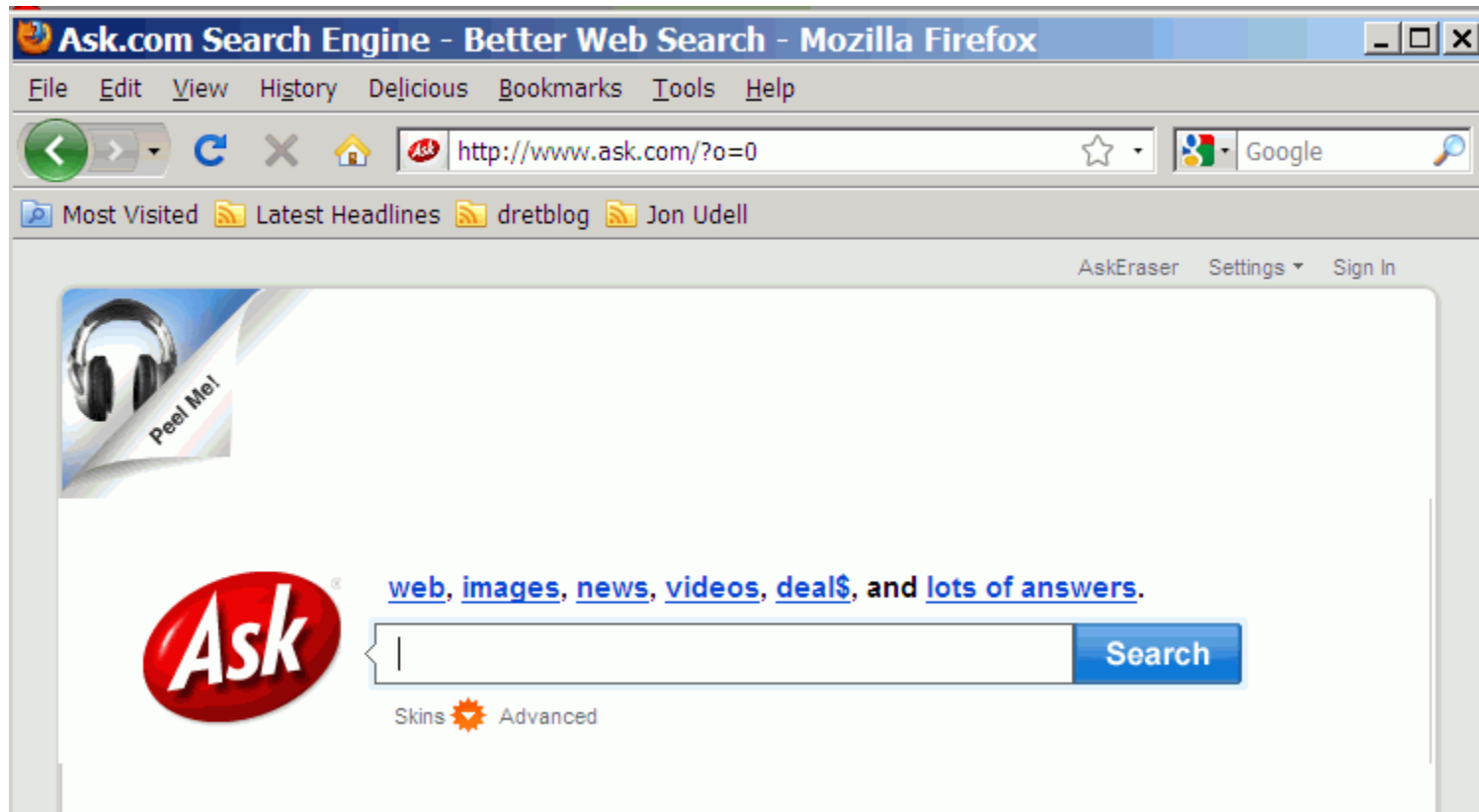
Google's Default Search UI



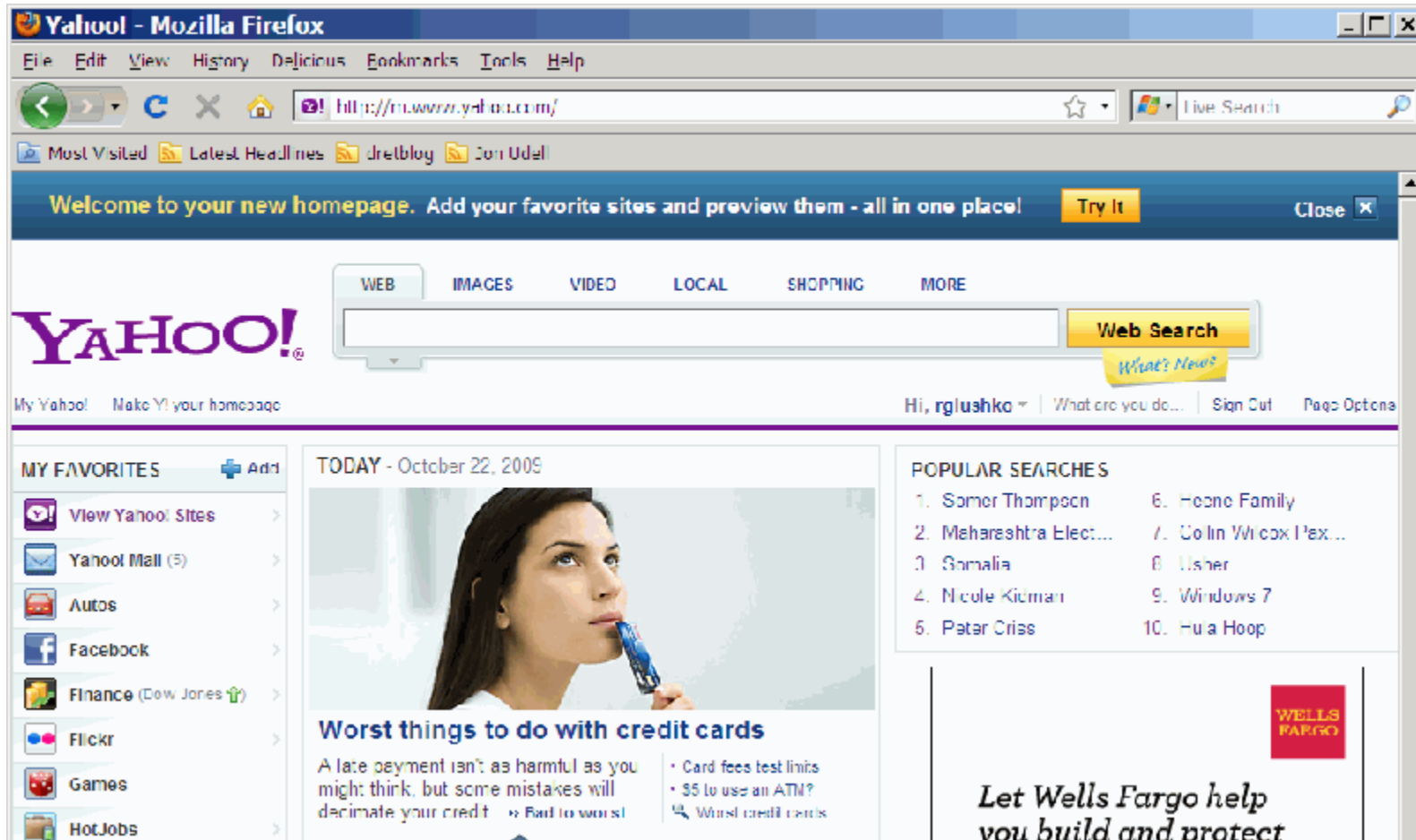
Microsoft Bing Default Search UI



Ask's Default Search UI



Yahoo! Default Search UI



"Ordinary" People Just Don't Get IR

1 in 7 never type URLs in the address bar and others use it wrong

Some use "URL style" (no spaces between words) when entering words into query forms

Only 1 in 6 uses quotes in query forms, and many of these do so incorrectly

Almost no use of any advanced search syntax or functions

They don't appreciate the "vocabulary problem," so if their first query term doesn't work, they just give up rather than trying other terms

"Ordinary" People Don't Understand Boolean Operators

For most people, Boolean semantics are counterintuitive or backwards

- Boolean AND narrows a search, but natural language "and" implies a request for more information
- Likewise, Boolean OR is a union that widens a search, while "or" implies a mutually exclusive choice in everyday language

"Ordinary" People Don't Understand Text Processing in IR Systems

If very frequent words ("stop words") are removed by the search engine, a query like "To be or not to be" won't find anything

A query like "boat fire" is different from "fire boat" but many users don't realize that term order matters

The Search Process and Interface Components

Hearst says "the heart of the search process is an iterative cycle of query specification, inspection and interpretation of query results, and query reformulation"

- Query Specification: Selecting and structuring search terms
- Query Results: Ordered list of documents or other objects matching the query
- Query Reformulation: If nothing in the results satisfies the query, users modify their initial queries and submit new ones

Small Details Matter

UIs for IR require great care in small details because of the text-intensive nature of search

- Tension between more information and clutter

How and where to arrange components of the interface and results matters a lot

- People don't read instructions or help text
- People scan / skim rather than read

Best Practices in Query Specification

Provide advanced capabilities for defining queries and constraining results, but progressively disclose them to hide complexity

Query suggestions

Query expansion and contraction (also used in query refinement stage) that is TRANSPARENT to the user

DWIM / spelling correction

Google Advanced Search

The screenshot shows the Google Advanced Search interface. At the top, the browser window title is "Google Advanced Search - Mozilla Firefox" and the address bar shows "http://www.google.com/advanced_search/". The page features the Google logo and the text "Advanced Search" with links for "Advanced Search Tips" and "About Google".

The main search area is titled "Use the form below and your advanced search will appear here". It contains several sections:

- Find web pages that have...**
 - all these words:
 - this exact wording or phrase: [tip](#)
 - one or more of these words: OR OR [tip](#)
- But don't show pages that have...**
 - any of these unwanted words: [tip](#)
- Need more tools?**
 - Results per page:
 - Language:
 - File type:
 - Search within a site or domain: (e.g. youtube.com, .edu)

At the bottom left, there is a link: [Date, usage rights, numeric range, and more](#). At the bottom right, there is a button labeled "Advanced Search".

Below the search form, there are links for "Topic-specific search engines from Google":

- [Google Book Search](#)
- [Google Code Search](#) **New!**
- [Google Scholar](#)
- [Google News archive search](#)
- [Apple Macintosh](#)
- [BSD Unix](#)
- [Linux](#)
- [Microsoft](#)
- [U.S. Government Universities](#)

Windows Bing Advanced Search

xml business - Bing - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://www.bing.com/search?q=xml+business&go=&qsn=&qbn=1&FORM=AXRE

Most Visited Latest Headlines dretblog Jon Udell

Web Images Videos Shopping News Maps More MSN Hotmail

bing xml business

ALL RESULTS

SEARCH HISTORY
xml business
[See all](#)
[Clear all](#) | [Turn off](#)

Advanced search x

Search terms | [Site/Domain](#) | [Country/Region](#) | [Language](#)

Look for results that meet the following criteria:

You can add words or phrases one at a time. [Get more help](#)

All of these terms
All of these terms
Any of these terms
This exact phrase
None of these terms

Add to search

ALL RESULTS 1-10 of 59,500,000 results · [Advanced](#)

[Thinking XML: Universal Business Language \(UBL\)](#)
Universal Business Language (UBL) is an ambitious effort to unify the chaotic world of XML formats for business. Recently, the group behind UBL released the first work products for ...
www.ibm.com/developerworks/xml/library/x-think16.html · [Cached page](#)

[Business process integration through XML](#)
Business to Business (B2B) e-commerce is driving a new generation of Internet applications that will use XML as a model for sharing documents and information between organizations
www.gca.org/papers/xml/europe2000/papers/s10-03.html · [Cached page](#)

Query Refinement (They Assumed You Wanted)

weather berkeley - Bing - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://www.bing.com/search?q=weather+berkeley&go=&for

Most Visited Latest Headlines dretblog Jon Udell

Web Images Videos Shopping News Maps More MSN Hotmail Sign In

bing weather berkeley

ALL RESULTS 1-10 of 6,150,000 results · [Advanced](#)

News

RELATED SEARCHES

- San Francisco Weather
- Berkeley College
- Sacramento Weather
- San Diego Weather
- Washington Weather
- California Weather
- Los Angeles Weather
- Berkeley County

ALL RESULTS

[Weather in Berkeley, CA](#)

Today **63° F** · (°C)
haze Wind: 7 mph W
Humidity: 77%
66° / 57°

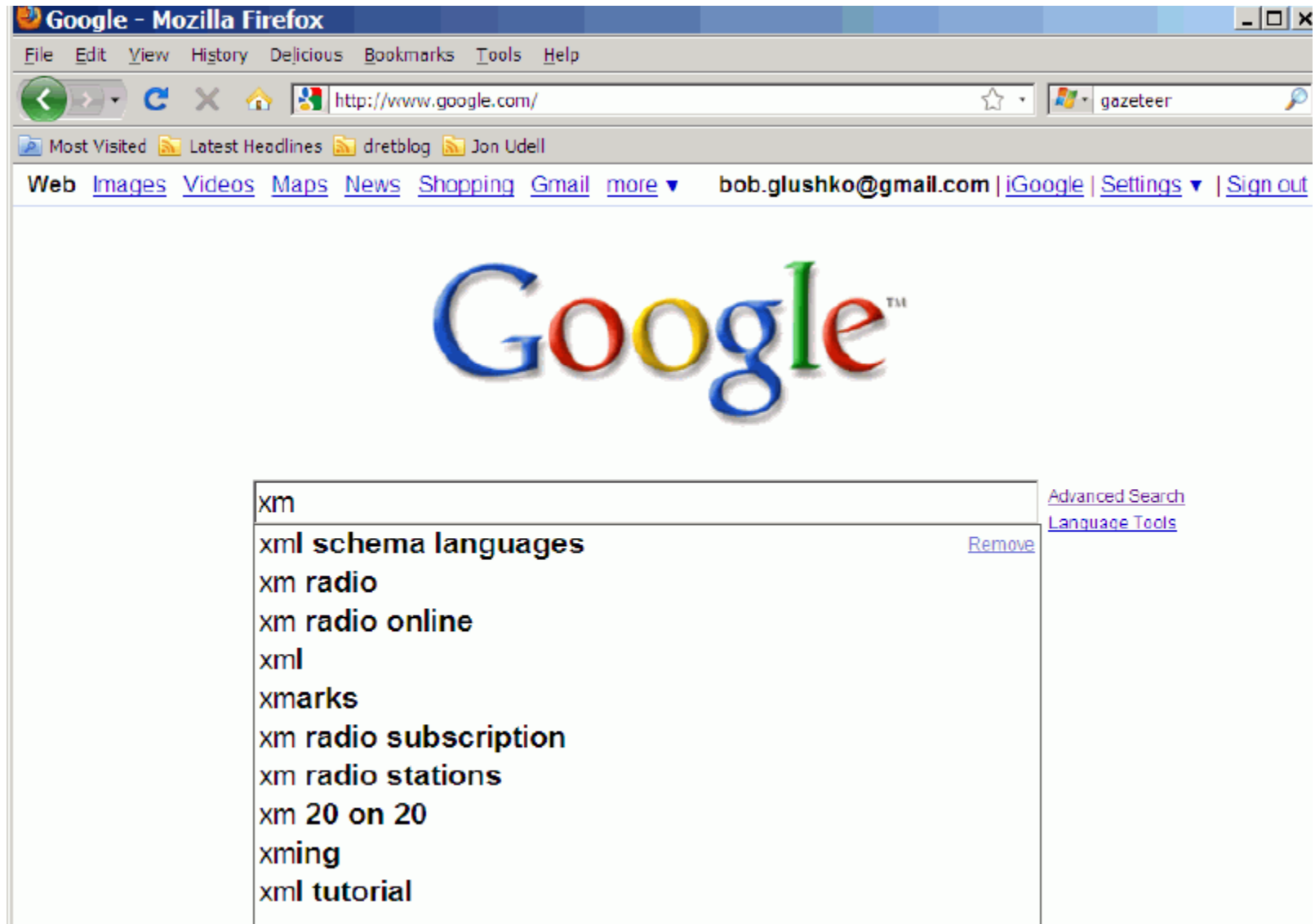
Fri	Sat	Sun	Mon
68° / 57°	69° / 57°	76° / 58°	69° / 52°

[10 Day Forecast](#) · [Hourly Forecast](#) · [Weather Maps](#) · Data provided by WDT

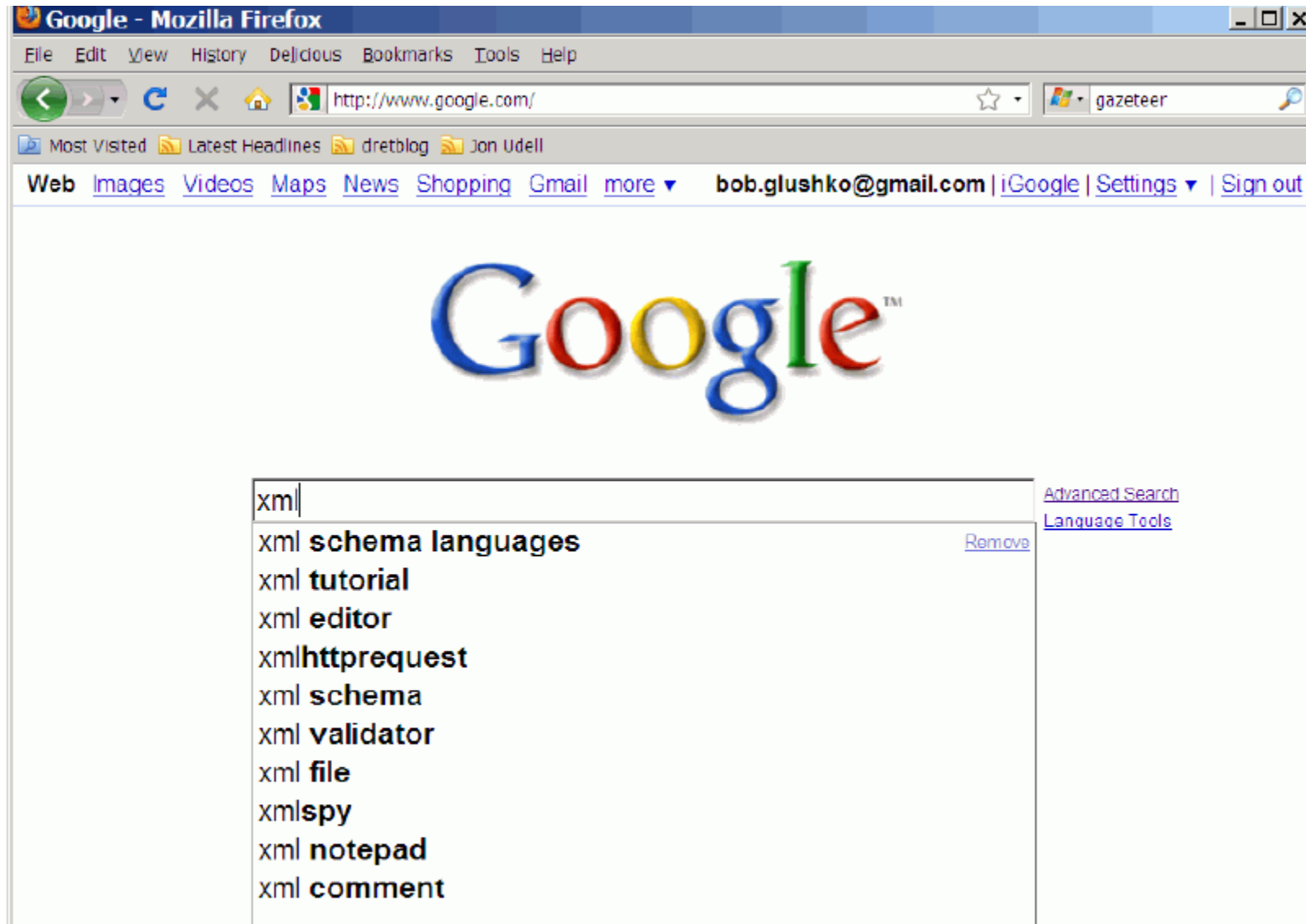
[Berkeley, California \(94701\) Conditions & Forecast : Weather ...](#)
Weather Underground provides weather information for worldwide locations, including current conditions, hourly forecasts, radar and satellite maps. Specialized weather products ...
www.wunderground.com/US/CA/Berkeley.html · [Cached page](#)

[Berkeley Weather Forecast and Conditions California \(94720\)](#)
Berkeley weather forecast and weather conditions for 94720. Today's Berkeley

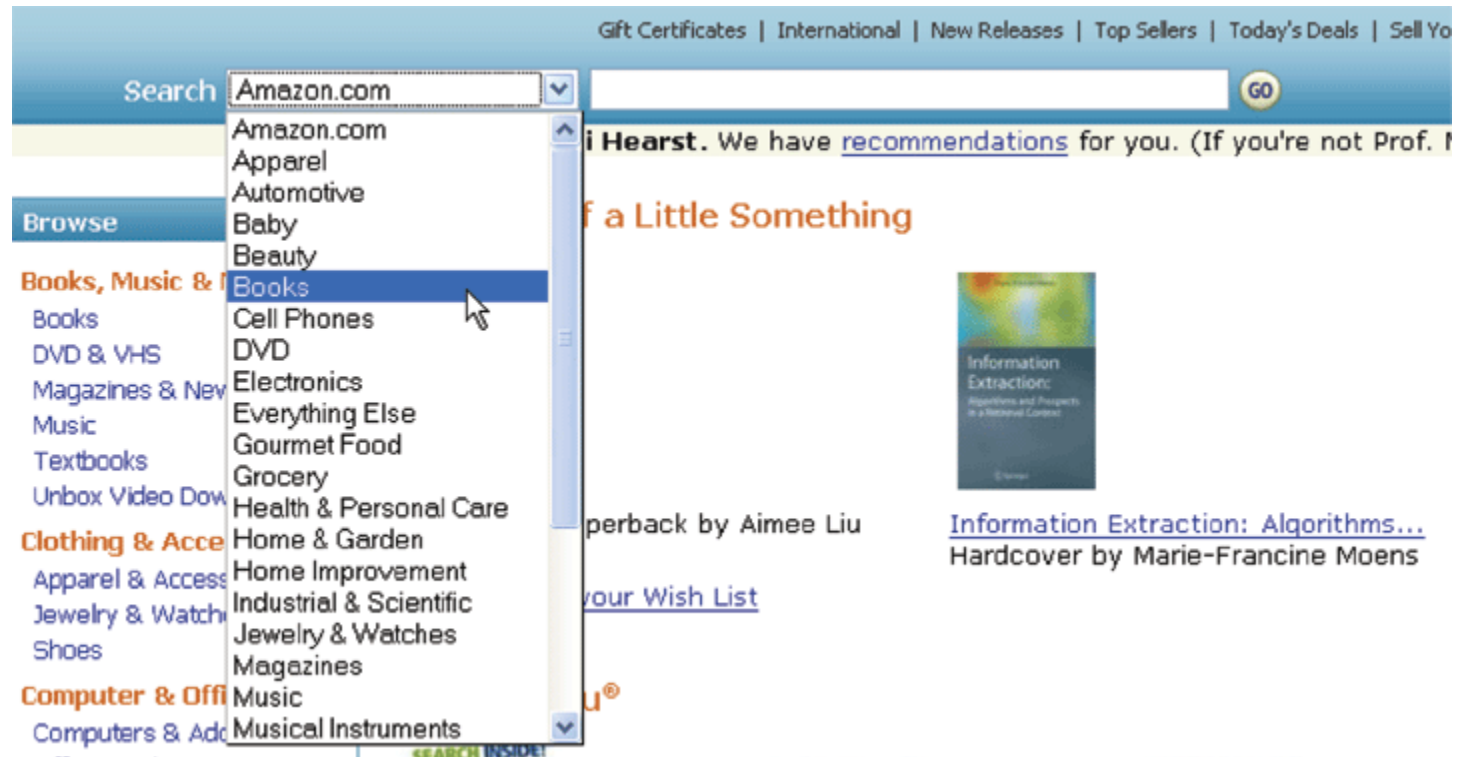
Query Suggestion in Google [xm]



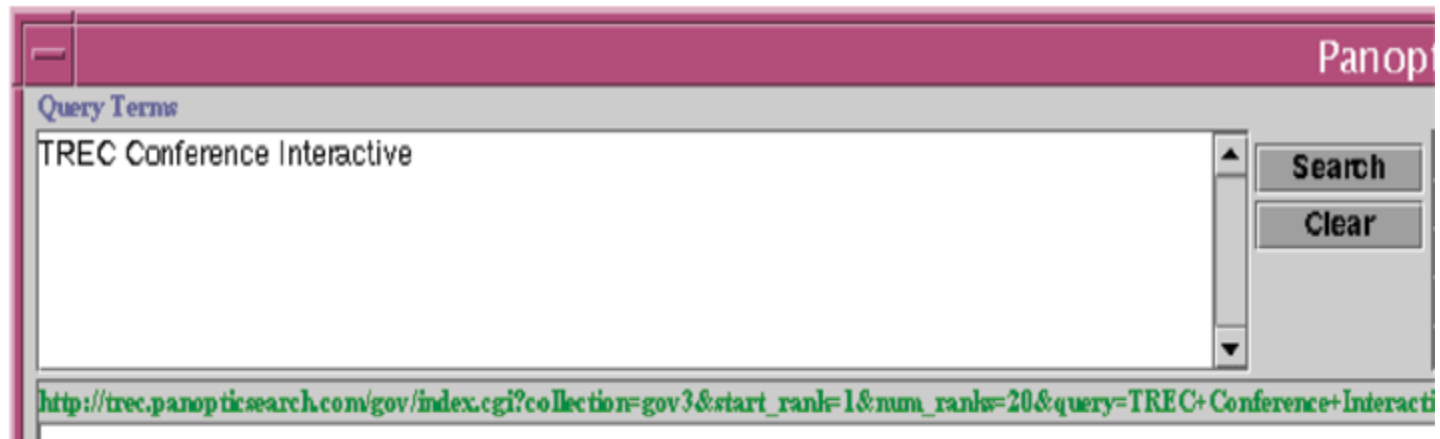
Query Suggestion in Google [xml]



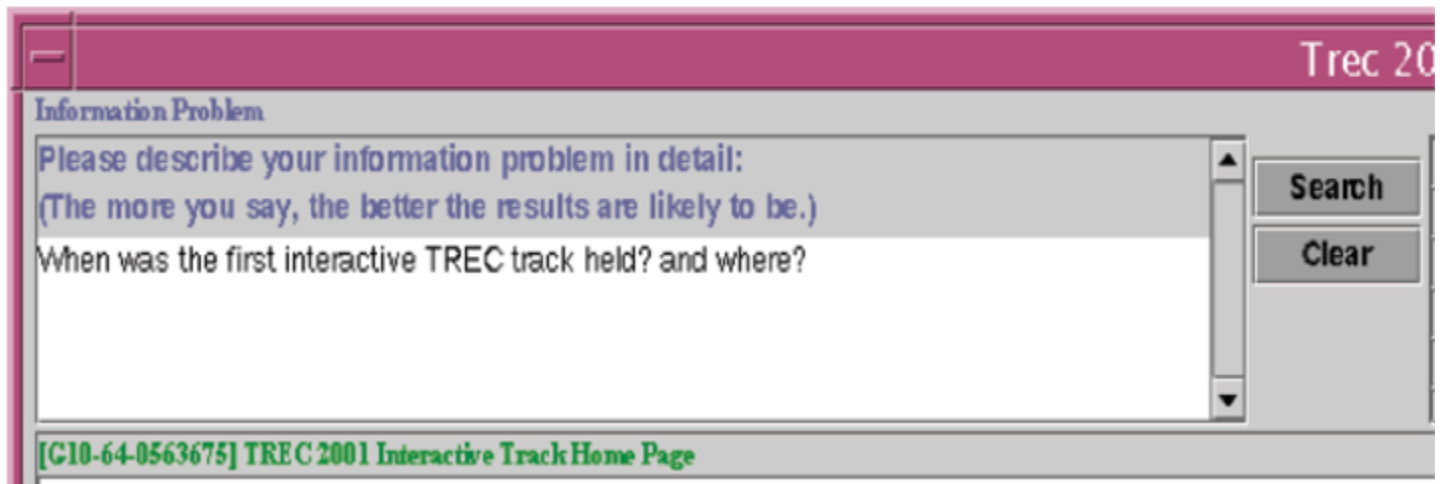
Search Window with Subject Drop Down Menu



Search Window with Terse and Verbose Instruction



(a)



Google Presents a Spelling Correction

Arnold **Schwartzeneqger** Sound Board Limited - ebaumsworld.com. Need more fun movie phrases? Click here for even more phrases! - ARNOLD EXTENDED OPEN. ...
home.rochester.rr.com/ebaumsworld/arnolds.html - 3k - [Cached](#) - [Similar pages](#)

[Stallone, Sylvester : The Stallone-**Schwartzeneqger** Thing](#)
... Stallone, Sylvester - The Stallone-**Schwartzeneqger** Thing. For a while, Arnold **Schwartzeneqger** and Sylvester Stallone seemed to have this running joke where they ...
www.eeggs.com/items/5647.html - 22k - [Cached](#) - [Similar pages](#)

[Send a letter to Governor **Schwartzeneqger** about higher education ...](#)
... Send a letter to Governor **Schwartzeneqger** about higher education fee increases
Posted by on Mar 7, 2004, 16:11, Graduate and Professional ...
www.myhastings.org/artman/publish/article_61.shtml - 27k - [Cached](#) - [Similar pages](#)

[Arnold **Schwartzeneqger**](#)
Arnold **Schwartzeneqger**. Arnold **Schwartzeneqger** did not appear in this film. ...Return to Home Page.
www.thequickandthedead.net/as.html - 2k - [Cached](#) - [Similar pages](#)

Did you mean to search for: [schwarzeneqger](#)

Gooooooog**le** ▶

Result Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Next](#)

[Search within results](#) | [Language Tools](#) | [Search Tips](#) | [Dissatisfied? Help us improve](#)



DWIM

"Do What I Mean" mechanisms try to be "smart" and determine the searcher's unstated intentions or goals

Examples:

- Automatically suggest spelling corrections
- Automatically augment my query with related terms, synonyms, abbreviations, etc.
- Pop up "the paperclip" that tells me what kind of help I need

CRITICAL POINT: Users love DWIM when it works, but **DESPISE** it when it doesn't

Best Practices in Results Presentation

Present ranked results (people won't look past the first page) but don't show ranks

Sort of search results according to important criteria (date, author)

Group results according to well-organized category labels (see Flamenco)

Highlight query terms

Present query terms in context

Counter-intuitive failure to help: visualization

Search Query Term Highlighting

TIP

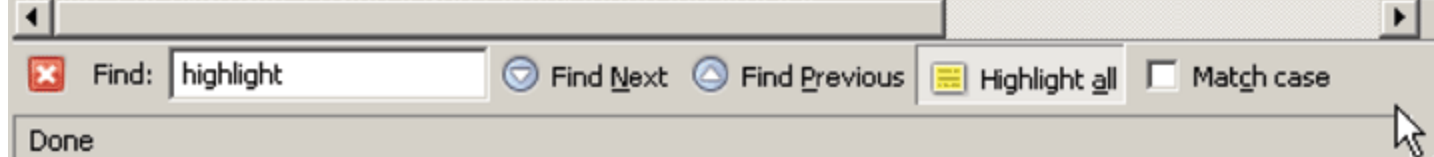
Firefox's Find bar can **highlight** all occurrences of the find text in the document if you click the **Highlight All** button.

NOTE

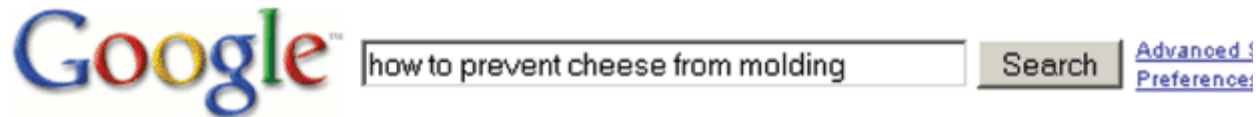
The Firefox Find bar—or, as it's sometimes called, the *Find toolbar*—has a number of controls. From the left they include an area in which to type characters to look for, a Find Next button, a Find Previous button, a **Highlight All** button, a Match Case button, and a message area.

Firefox always **highlights** the first occurrence of the term typed. Clicking the **Highlight** button tells Firefox to **highlight** all occurrences of the search term found on the page. Pressing F3 or clicking Find Next takes you to the next occurrence of the find text.

Using the Find bar's Find area, you can type a word(s). As you type, Firefox searches for what you type, finding and **highlighting** the first occurrence. As you type more letters, the search moves to the next match. For example, I typed **h** and Firefox found the first **h** it



Search Query Term in Context



Web [Books](#) Results 1 - 10 of about 175,000 for [how to prevent](#)

Book results for [how to prevent cheese from molding](#)



[Food Science](#) - by Norman N. Potter, Joseph H. Hotchkiss - 608 pages

[Extraordinary Uses for Ordinary Things](#) - by Reader's Digest Association - 400 pages

[Sargento.com | FAQs](#)

Mold can develop on **cheese** once the product is exposed to air. ... packaging is not 100 percent airtight and does not always **prevent mold** development. ...

[www.sargentocheese.com/aboutus/faq.jsp](#) - 26k - [Cached](#) - [Similar pages](#)

[A Forest Ranger's Cookbook - Household Suggestions](#)

To **prevent cheese from molding**, wrap in a cloth wrung out of vinegar. Then roll in paper. In cooking vegetables: Cover those that grow under the ground. ...

[forestry.about.com/cs/foresthstory1/a/camp_reci_tips.htm](#) - 26k - [Cached](#) - [Similar pages](#)

[Preventing Moldy Cheese](#)

It's the air that makes the **mold** on **cheese** so keeping air from getting at ... In the bowl I add a little corn starch which will **prevent** it from sticking. ...

[www.stretcher.com/stories/981123b.cfm](#) - 42k - [Cached](#) - [Similar pages](#)

[How to Prevent Cheese Mold | eHow.com](#)

How to Prevent Cheese Mold. It is frustrating to go into the refrigerator to get **cheese** for a recipe or sandwich and find that it has gone green with **mold**.

[www.ehow.com/how_2043886_prevent-cheese-mold.html](#) - 49k - [Cached](#) - [Similar pages](#)

Grouping of Search Results

Interviews with lay users often reveal a desire for better organization of retrieval results

Useful for suggesting where to look next

Variety of techniques - categories vs clusters, single vs complex (faceted) category structure

Comparing the Techniques -- Clustering

Clustering is an automated technique for assigning results to groups (exclusively)

More flexible than pre-defined categories

Disambiguates ambiguous terms

Automatically generated labels can be unintuitive and occur at different levels of description

Potentially useful if the user wants a summary of the main themes in the subcollection

Potentially harmful if the user is interested in less dominant themes

Results Clustering by Carrot2

The screenshot shows the Carrot2 Clustering Engine interface in a Mozilla Firefox browser. The search term is "berkeley". The results are clustered into a tree structure on the left and a list of top 100 results on the right.

Tree Visualization:

- All Topics (100)
 - University of California (20)
 - Berkeley School (9)
 - Berkeley County (7)
 - International Society for Heart and Lung Transplantation (5)
 - National (5)
 - Public (5)
 - George Berkeley (4)
 - College (3)
 - Computer Sciences (3)
 - Philosophy (3)
 - more | show all

Top 100 results of about 3320000 for berkeley:

- University of California, Berkeley**
University of California, **Berkeley** home page. Gateway to information on studying , teaching, research and public service at UC **Berkeley**, flagship campus of ...
<http://berkeley.edu/> [Ask, Google, Yahoo]
- Official web site of the City of Berkeley, California**
Information on city services and departments as well as calendar and links.
<http://www.ci.berkeley.ca.us/> [Ask, Bing, Entireweb, Exalead, Google, Yahoo]
- Berkeley Convention & Visitors Bureau**
Visit **Berkeley.com** - **Berkeley** Convention & Visitors Bureau.
<http://www.visitberkeley.com/> [Ask, Google]
- Berkeley, California - Wikipedia, the free encyclopedia**
Berkeley (pronounced /ˈbɜrklɪ/) is a city on the east shore of the San Francisco Bay in Northern California, United States. Its neighbors to the south are ...
http://en.wikipedia.org/wiki/Berkeley,_California [Cui, Google, Wikipedia]

Comparing the Techniques -- Categories

Human-created categories, but items can often be automatically assigned (to multiple categories)

Usually restricted to a fixed set

Intended to be readily understandable to those who know the underlying domain

Provide a novice with a conceptual structure

Results Categorization by Google

Vitamin A - Google Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help del.icio.us

http://www.google.com/search?q=Vitamin+A&ie=utf-8&oe=utf-8&

Latest Headlines

Web Images Video News Maps Gmail more ▾ bob.glushko@gmail.com | My Not

Google Vitamin A Search [Advanced Search](#) [Preferences](#)

Web News Personalized Results 1 - 10 of about 56,500,000

Refine results for Vitamin A:

[Drug uses](#) [Interactions](#) [For patients](#) [From medical authorities](#)
[Side effects](#) [Warnings/recalls](#) [For health professionals](#)

[Vitamin A - Wikipedia, the free encyclopedia](#)
Vitamin A is an essential human nutrient. It exists not as a single compound, but in several forms. In foods of animal origin, the major form of **vitamin A** ...
en.wikipedia.org/wiki/Vitamin_A - 41k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Dietary Supplement Fact Sheet: Vitamin A and Carotenoids](#)
Extensive information about the fat soluble **vitamin** including suggested RDA intake, role in human body functions, food sources, and deficiency signs.
ods.od.nih.gov/factsheets/vitamina.asp - 82k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Vitamin A](#)
Discussion of function, deficiency, disease prevention, disease treatment, food sources, supplements, new research, and recommendations from the Linus ...
lpi.oregonstate.edu/infocenter/vitamins/vitaminA/ - 60k - [Cached](#) - [Similar pages](#) - [Note this](#)

Results Categorization by Google [2]

Vitamin A more:drug_uses - Google Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help del.icio.us

http://www.google.com/search?hl=en&client=firefox-a&rls=org.mozilla:en-US:official&hs=V5g&pwst=1&q=Vitamin+

Latest Headlines

Web Images Video News Maps Gmail more

Google Vitamin A more:drug_uses Search Advanced Search Preferences

Web Personalized Results 1

Refine results for Vitamin A:

Drug uses Interactions For patients From medical authorities
Side effects Warnings/recalls For health professionals

[Aquasol A \(Vitamin A\) drug indications and dosage - prescription ...](#)
Learn the indications and **dosage** for Aquasol A (Vitamin A) and other ... Vitamin A injection is effective for the treatment of **vitamin A** deficiency. ...
www.rxlist.com/cgi/generic/aquasola_ids.htm - 34k - [Cached](#) - [Similar pages](#) - [Note this](#)
Labeled by [WebMD](#) [MLA](#) [HON](#)

[MedlinePlus Herbs and Supplements: Vitamin A \(Retinol\)](#)
The World Health Organization (WHO) has established **dosage** guidelines for children between 6-11 months-old to receive 100000 IU of **vitamin A**. This increases ...
www.nlm.nih.gov/medlineplus/druginfo/natural/patient-vitamina.html - 45k - [Cached](#) - [Similar pages](#) - [Note this](#)
Labeled by [HON](#) [CDC](#)

[VITAMIN A - ORAL \(Aquasol A, Palmitate-A\) side effects, medical ...](#)
Avoid taking cholestyramine or mineral oil at the same time as **vitamin A** because they can prevent absorption of **vitamin A**. Separate the dosing by at least 2 ...
www.medicinenet.com/vitamin_a-oral/article.htm - 41k - [Cached](#) - [Similar pages](#) - [Note this](#)
Labeled by [HON](#)

[Drugs and Treatments - Vitamin D Oral - Patient Handout from WebMD](#)
Vitamin D is used to treat and prevent bone disorders caused by low levels of ... If your doctor has prescribed this medication, the **dosage** is based on your ...
www.webmd.com/drugs/drug-6414-Vitamin+D+Oral.aspx?drugid=6414&drugname=Vitamin+D+Oral - 87k - [Cached](#) - [Similar pages](#) - [Note this](#)
Labeled by [HON](#)

Results Organized Using Faceted Categories (Flamenco)

Flamenco

Refine your search further within these categories:

These terms define your current search. Click the **X** to remove a term.

Media (group results)
costume (3), drawing (2), lithograph (1), woodcut (6), woven object (2)

Location: all > Asia
Afghanistan (1), China (4), China or Tibet? (3), India (2), Japan (13), Russia (1), Turkey (3), Turkmenistan (1)

Date (group results)
17th century (3), 18th century (3), 19th century (10), 20th century (3), date ranges spanning multiple centuries (7), date unknown (2)

Themes (group results)
music, writing, and sport (5), nautical (1), religion (2)

Objects (group results)
clothing (5), food (1), furnishings (4), timepieces (1)

Nature (group results)
bodies of water (3), fish (1), flowers (2), geological formations (1), heavens (3), invertebrates and arthropods (1), mammals (2), plant material (3), ...


Location: Asia **X**

Shapes, Colors, and Materials: fabrics **X**

all items within current results




28 items (grouped by location) [view ungrouped items](#)

Afghanistan 1



Girl's Ceremonia...
no artist
20th century

China 4



Best Practices in Query Reformulation

(Do what works in query specification within the context of current results)

Make it easy to conduct iterative search by modifying queries to search within the current results

Get relevance feedback from searcher ("more like this")

Query Refinement Based on Relevance Feedback

Many studies show that if users engage in relevance feedback, the results are much better.

But the explicit effort required to rate relevance is usually a roadblock

This is one motivation for using "social" or indirect methods of assessing relevance, (+1, LIKE, etc)

Four Big Ideas

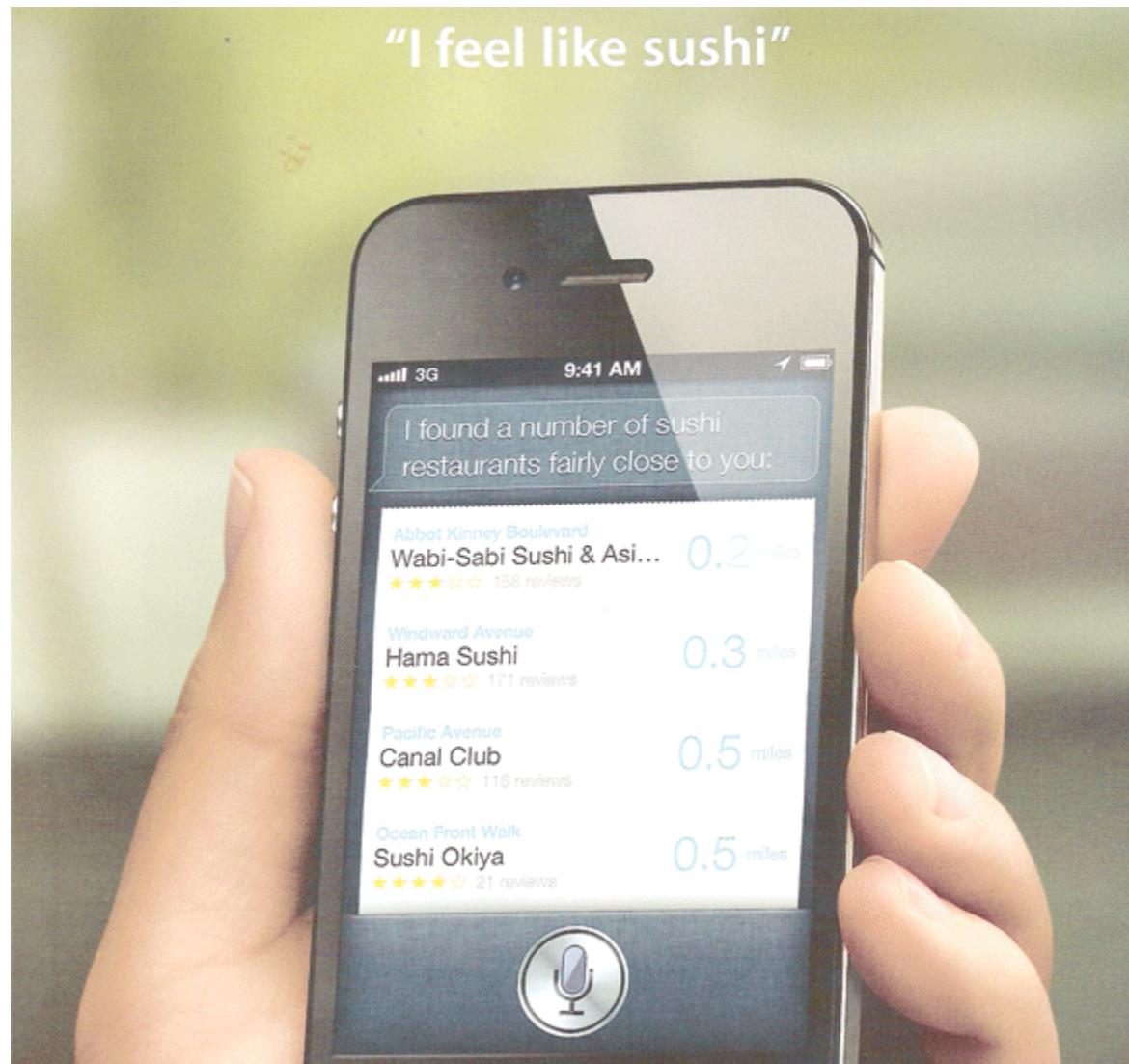
Personalization

Timeliness (<http://bits.blogs.nytimes.com/2011/11/03/google-cha>)

"Smart" voice recognition - SIRI

"Social" search - exploit content created by your
"friends" <http://www.bing.com/explore/social/>

SIRI in the iPhone 4



Watson Plays Jeopardy

[Watson Beats the Human Champs](http://www.youtube.com/watch?v=xm8iUjzgPTg&feature=related) (<http://www.youtube.com/watch?v=xm8iUjzgPTg&feature=related>)

Jeopardy uses a broad and open knowledge domain, uses complexly (with puns and abbreviations) worded clues, demands precise answers, and you have to be quick!

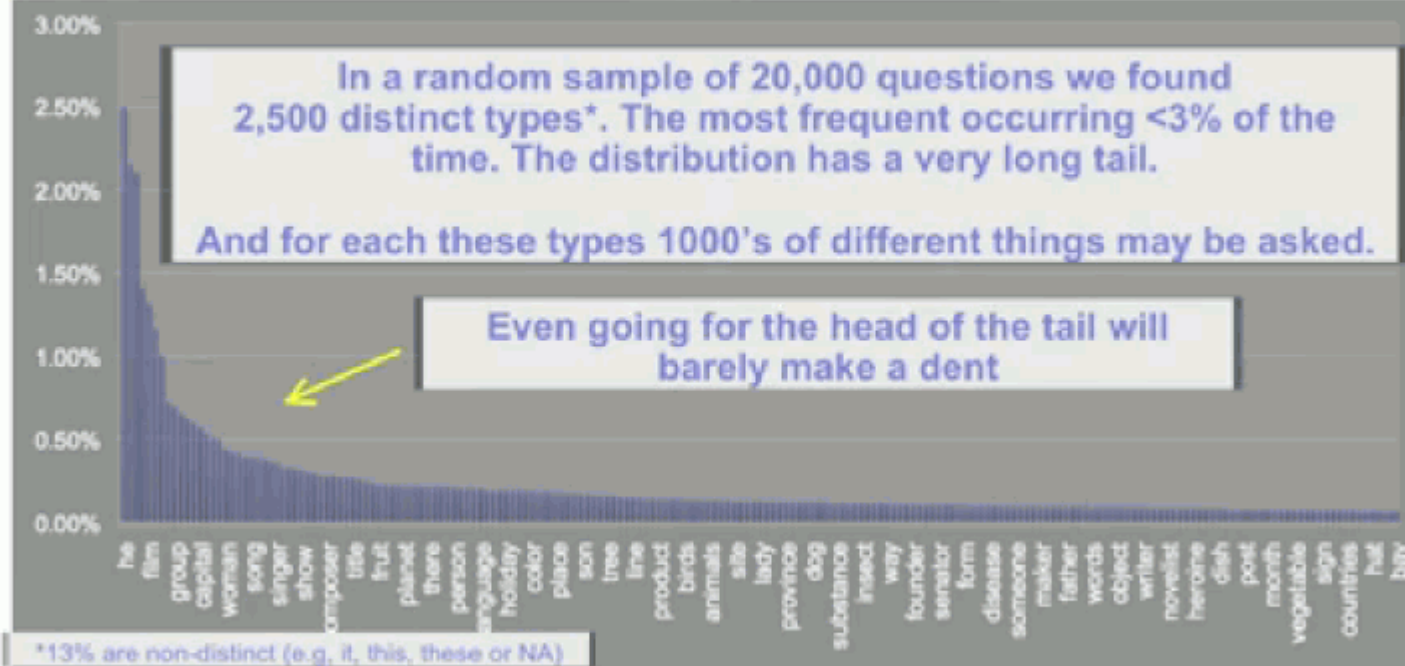
A compelling technology (and marketing) demonstration for IBM

Watson's Approach to QA

Broad Domain

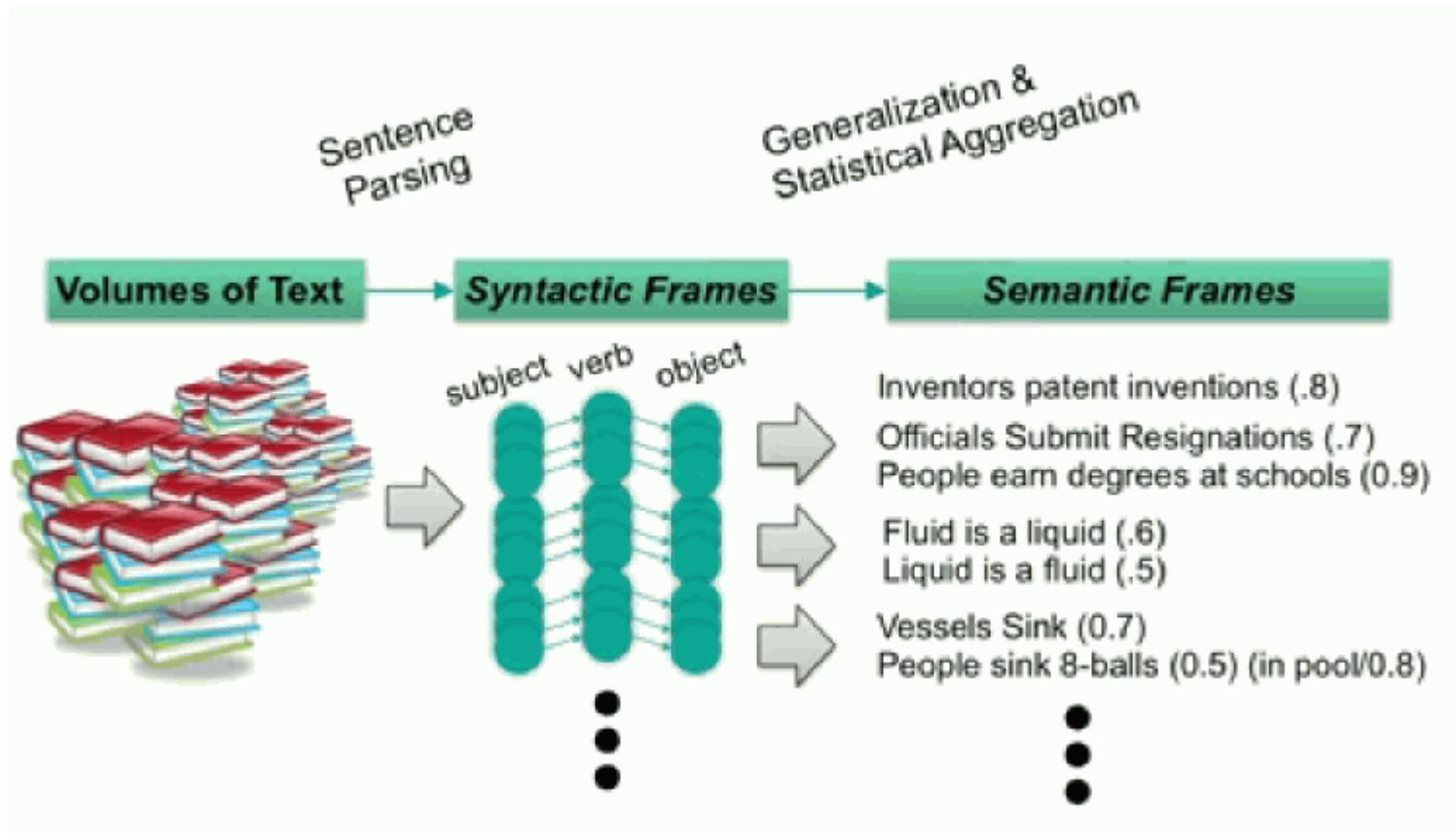
We do NOT attempt to anticipate all questions and build databases.

We do NOT try to build a formal model of the world

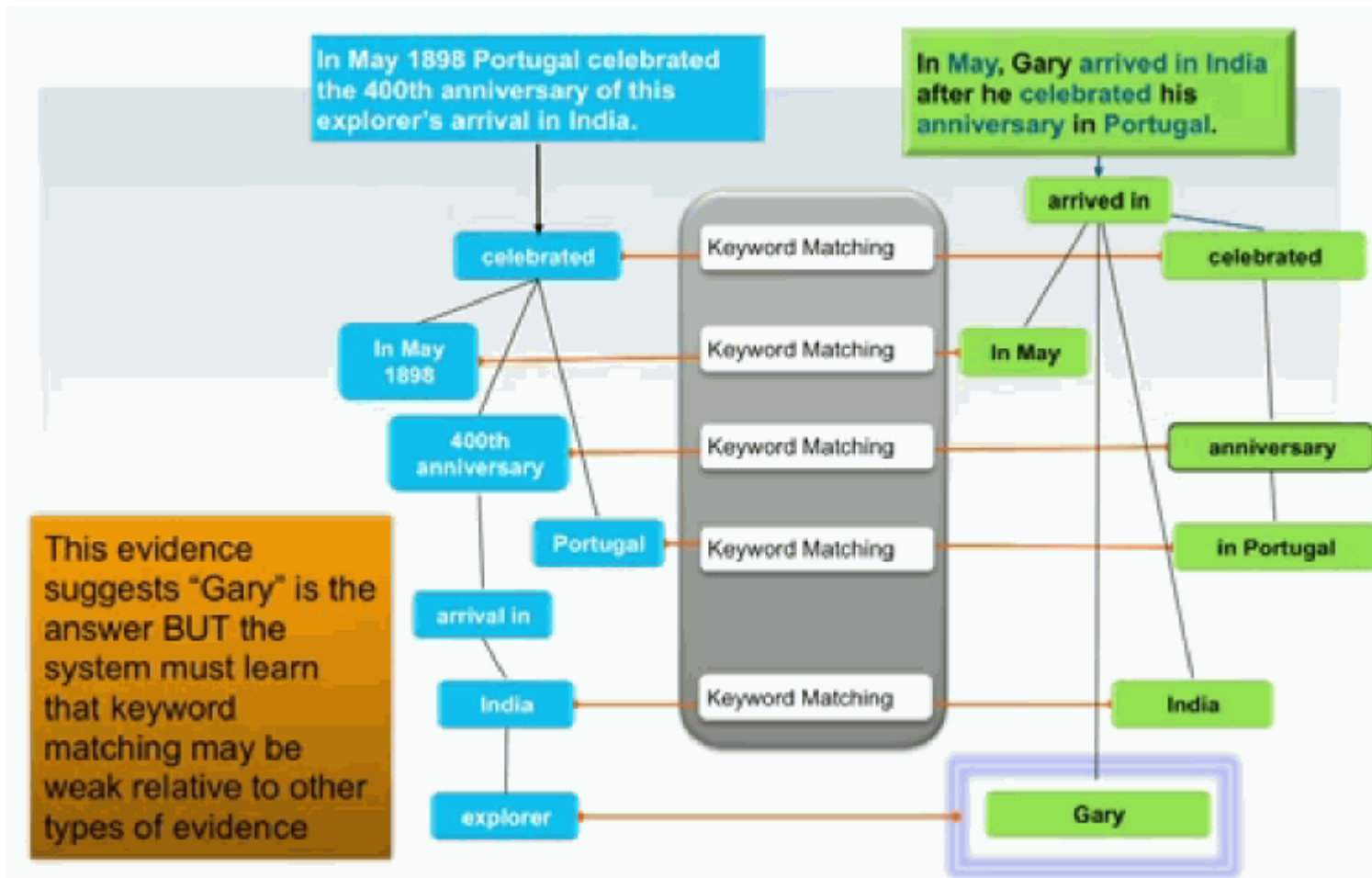


Focus is on reusable NLP technology for analyzing vast volumes of *as-is* text. Structured sources (DBs and KBs) provide background knowledge for interpreting the text.

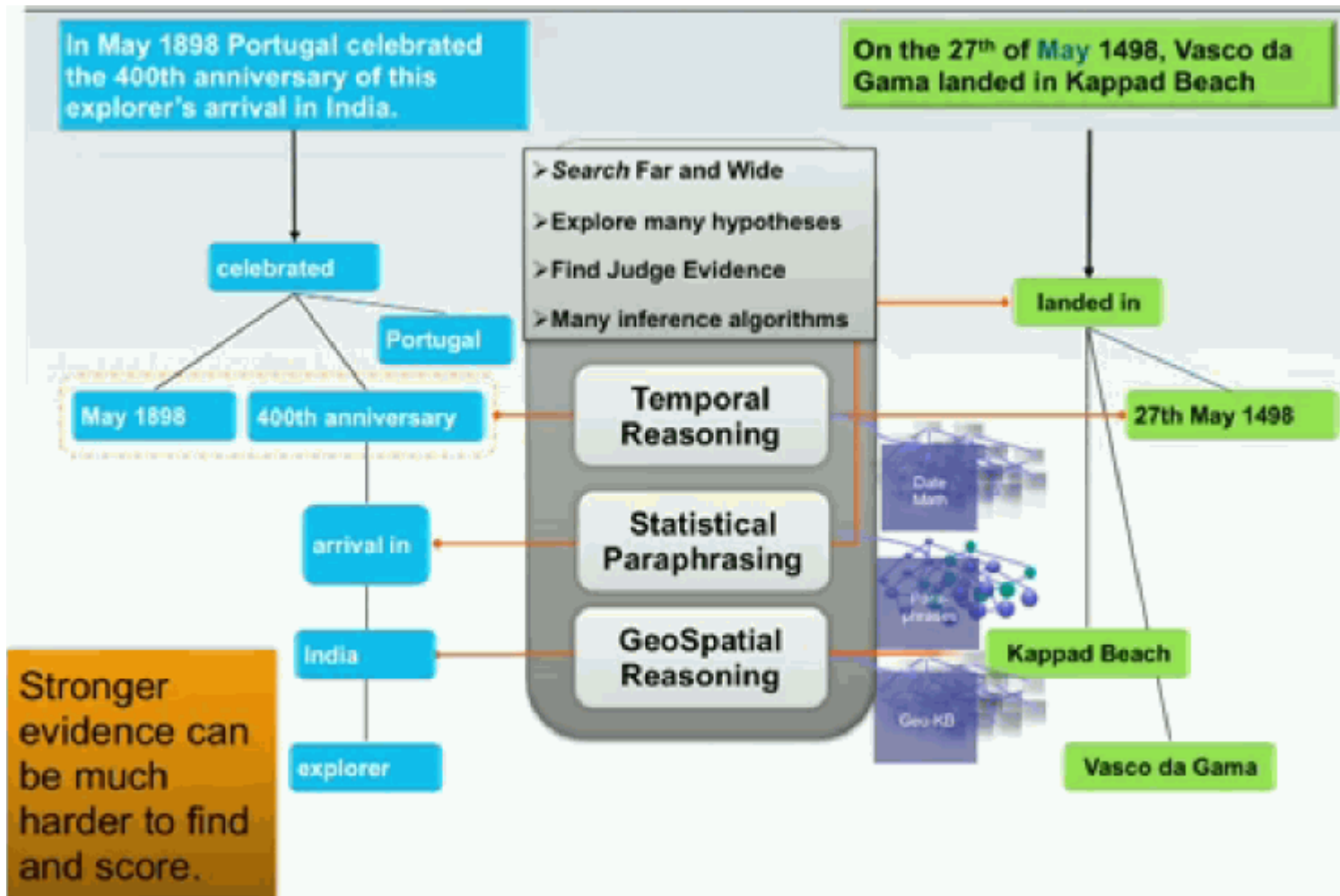
Watson - Learning from Reading



Watson - Why Keywords Won't Work



Watson - Using Deeper Evidence



Readings for Next Lecture (11/15)

Manning: Chapter 6 (focus on sections 6.2 and 6.3)

Yu, Clara, et al. – "Patterns in Unstructured Data" (from "Latent Semantic Indexing" through "Applications")