WEB SEARCH BASICS

CS 128

Information Retrieval & Big Data

Prof. Donald J. Patterson

Content adapted from Hinrich Schütze http://www.informationretrieval.org

OVERVIEW

OVERVIEW

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



WHERE DID BIG DATA COME FROM?

- Classic IR
- Web IR
- Generalized data processing frameworks
- Cheap cloud storage
- Lots of data
 - Apply the frameworks to them



Search behavior has changed a lot over time

• How often do you use search?



Search behavior has changed a lot over time

• How often do you use search?

• How often do you use search engines?



GOOGLE HANDLES 63,000 SEARCHES / SEC

- 2 trillion searches per year
- 5.6 billion searches per day
- 228 million per hour
- 3.8 million per minute





15% OF ALL SEARCHES HAVE NEVER BEEN SEARCHED BEFORE

- 9500 per second
- Where are these coming from?





MORE THAN 200 FACTORS ARE USED TO RANK WEB PAGE RELEVANCE

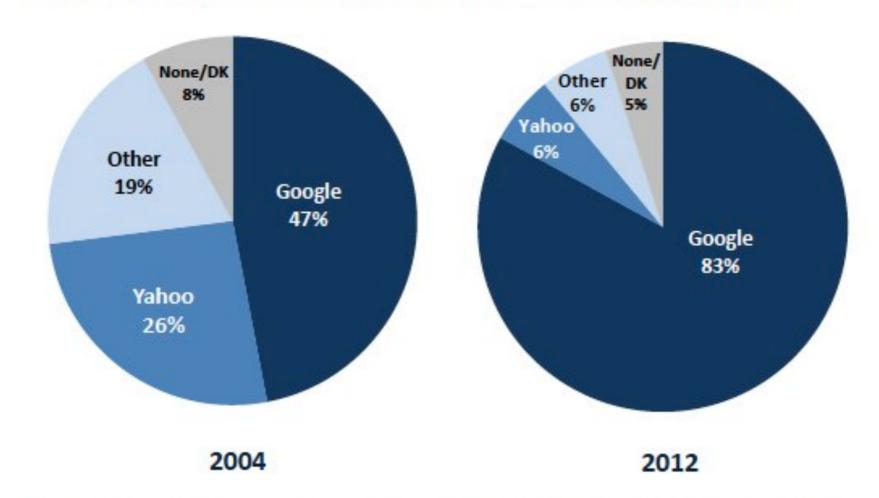
- Domain Factors
- Page-level factors
- Site-level factors
- back link factors
- user interaction
- special Google rules
- brand signals
- spam factors





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?

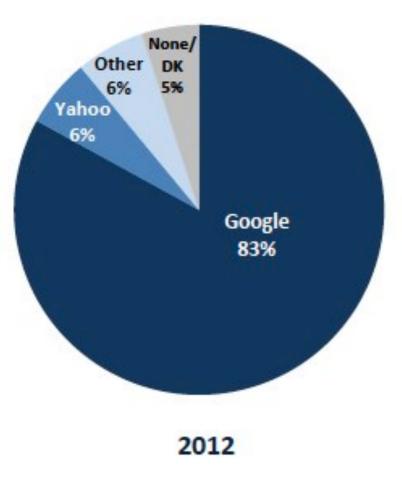


Source: The Pew Research Center's Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults, age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish.

Search Engine Watch 2016
Pew Internet 2012

ch engine of choice,

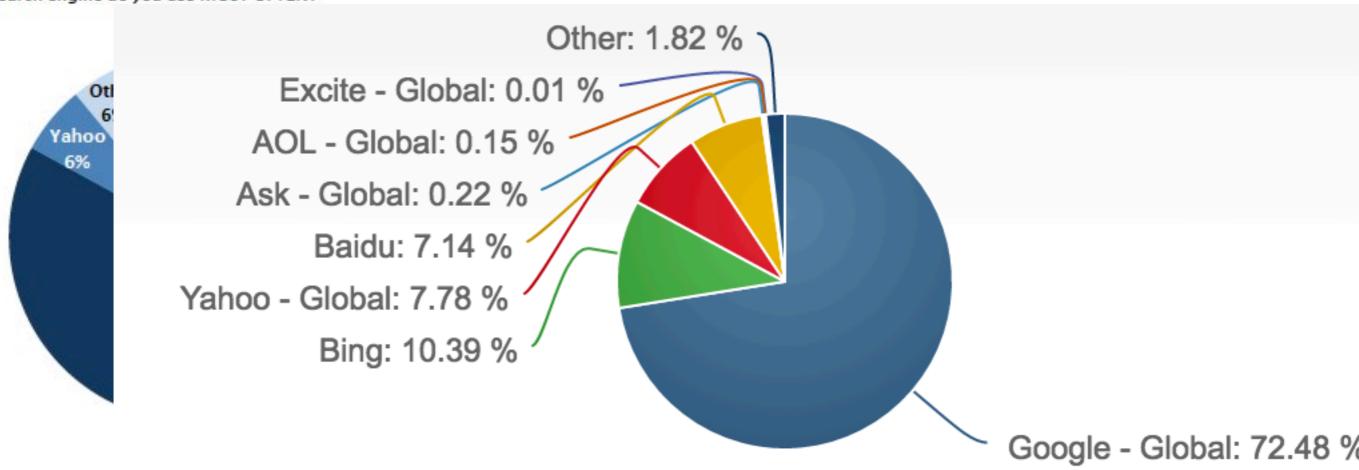
search engine do you use MOST OFTEN?



Project Winter 2012 Tracking Survey, January 20ing 901 cell phone interviews. Interviews conducted in

ch engine of choice, s

search engine do you use MOST OFTEN?

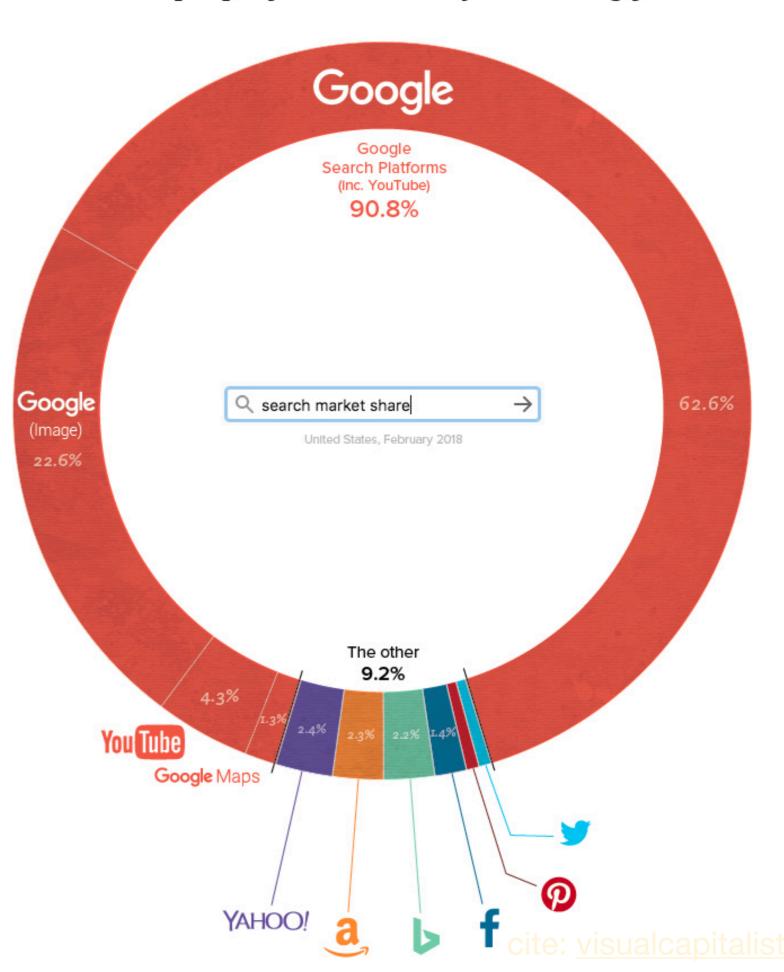


Project Winter ng 901 cell phone interviews. Interviews conducted in

Search Engine Watch 2016
Pew Internet 2012

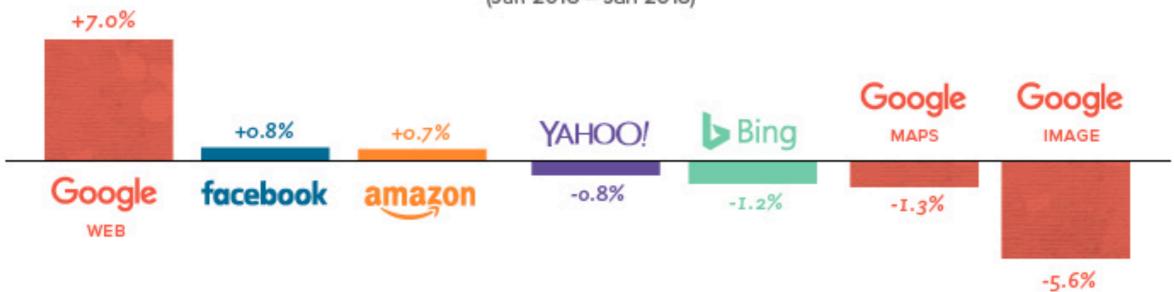
GOOGLE HAS
90.46% OF THE
SEARCH ENGINE
MARKET GLOBALLY
(2018)

How people find what they're looking for

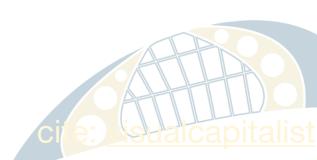


Change in search market share

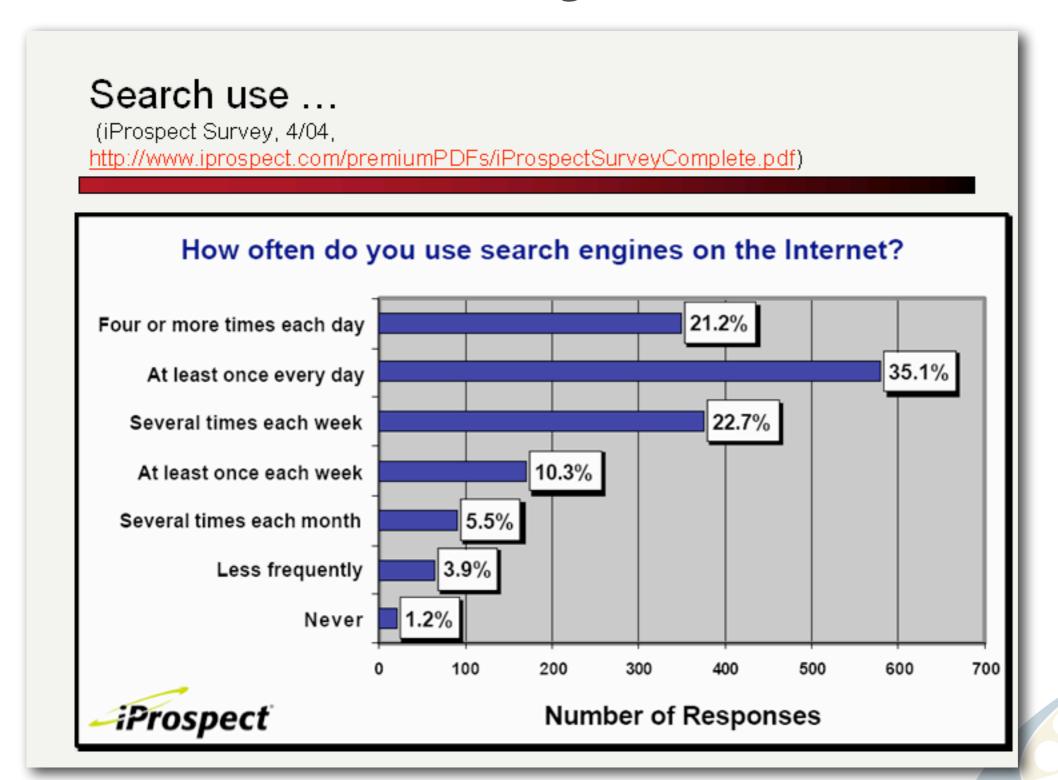
(Jan 2016 - Jan 2018)



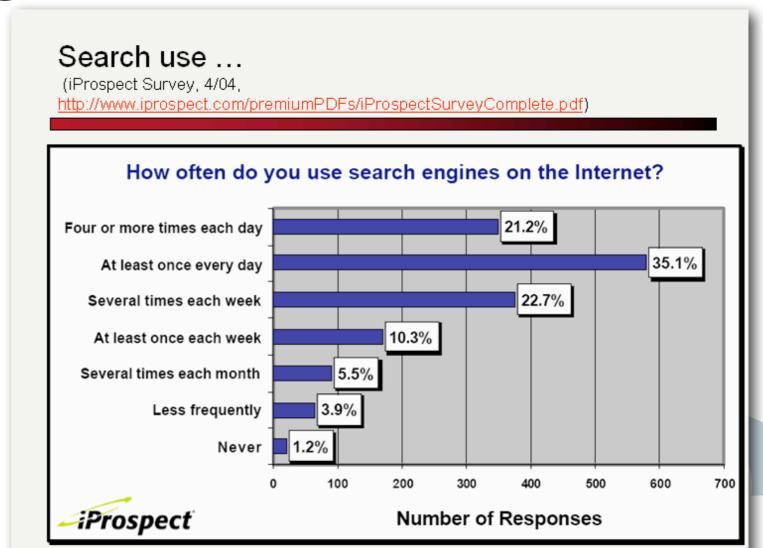
Excludes platforms that experienced a change of less than 0.5%



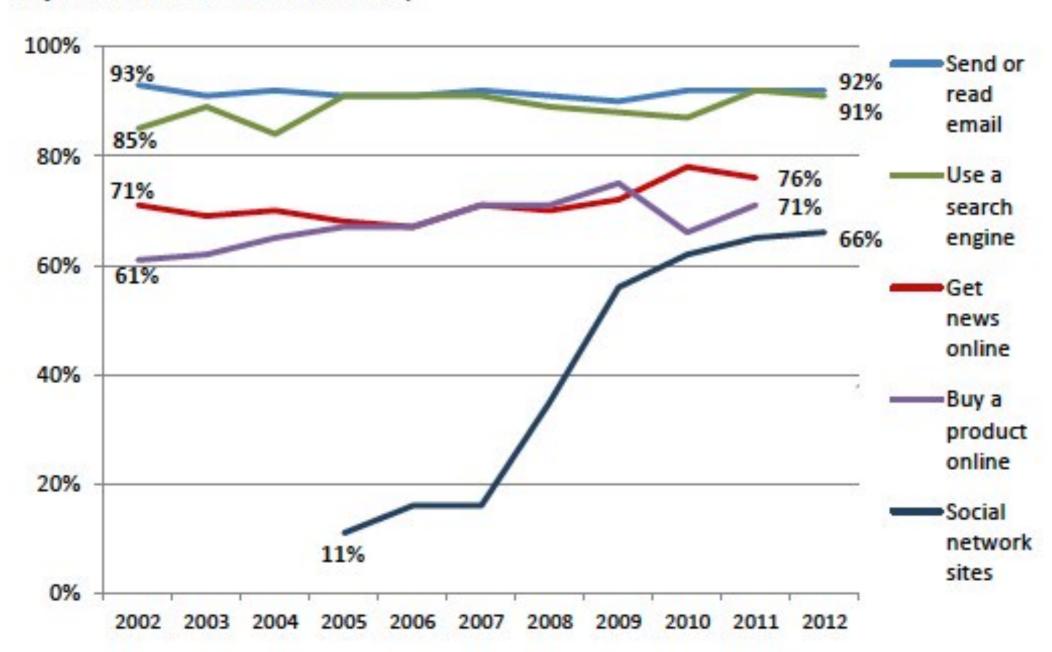
Search behavior has changed a lot over time



- Search behavior has changed a lot over time
- 2004
- 2018 Estimated 3-4 searches per day
 - Why is this hard to figure out?



% of internet users who do each activity



Who uses search?

% of online adults in each group who use search engines

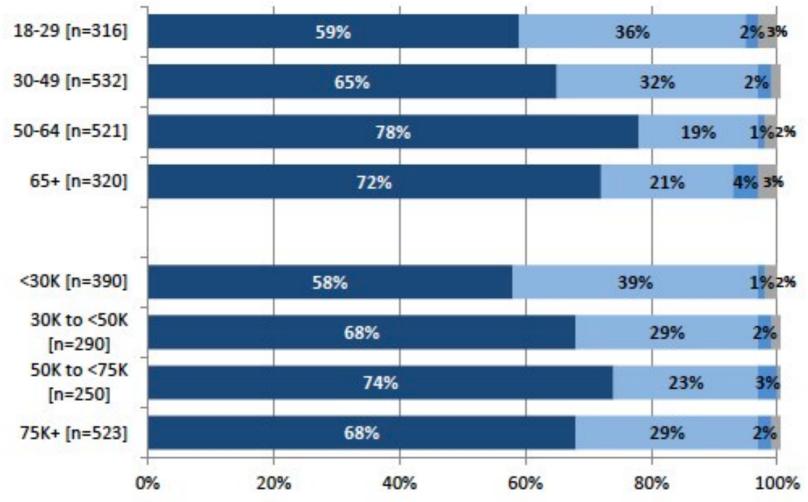
	% of each group who ever use search engines	% of each group who used a search engine yesterday
All online adults	91%	59%
Gender		
Male	90	59
Female	92	60
Race/Ethnicity	AV 2 - 22	
White	93*	63*
African American	89*	44
Hispanic	79	44
Age	·	
18-29	96	66*
30-49	91	65*
50-64	92	52*
65+	80	38
Education		
Some high school	78	34
High school	88*	45*
Some college	94*	65*
College graduate	95*	74*
Household income		
< \$30,000	84	45
\$30,000 - \$49,999	93*	54*
\$50,000 - \$74,999	97*	66*
\$75,000+	95*	76*



Views of targeted advertising vary by age and income

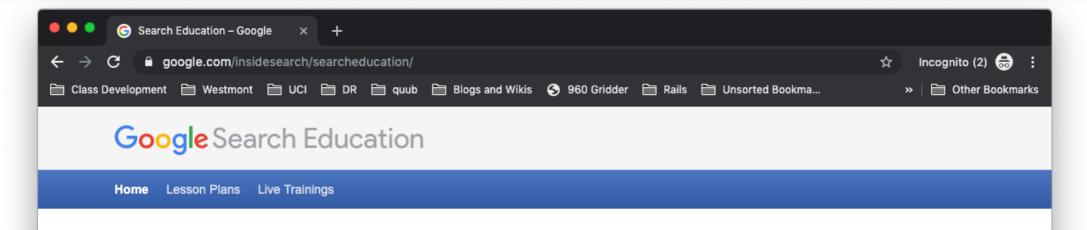
Which of the following statements comes closest to how you, personally, feel about TARGETED ADVERTISING being used online – even if neither is exactly right?

Asked of adult internet users [n=1,729]



- I'm NOT OKAY with targeted advertising because I don't like having my online behavior tracked and analyzed
- I'm OKAY with it because it means I see ads and get information about things I'm really interested in
- Neither (VOL)
- DK/Ref

Pew Internet 2012

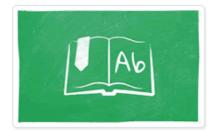


Help your students become better searchers

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.

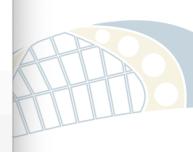
Browse challenges

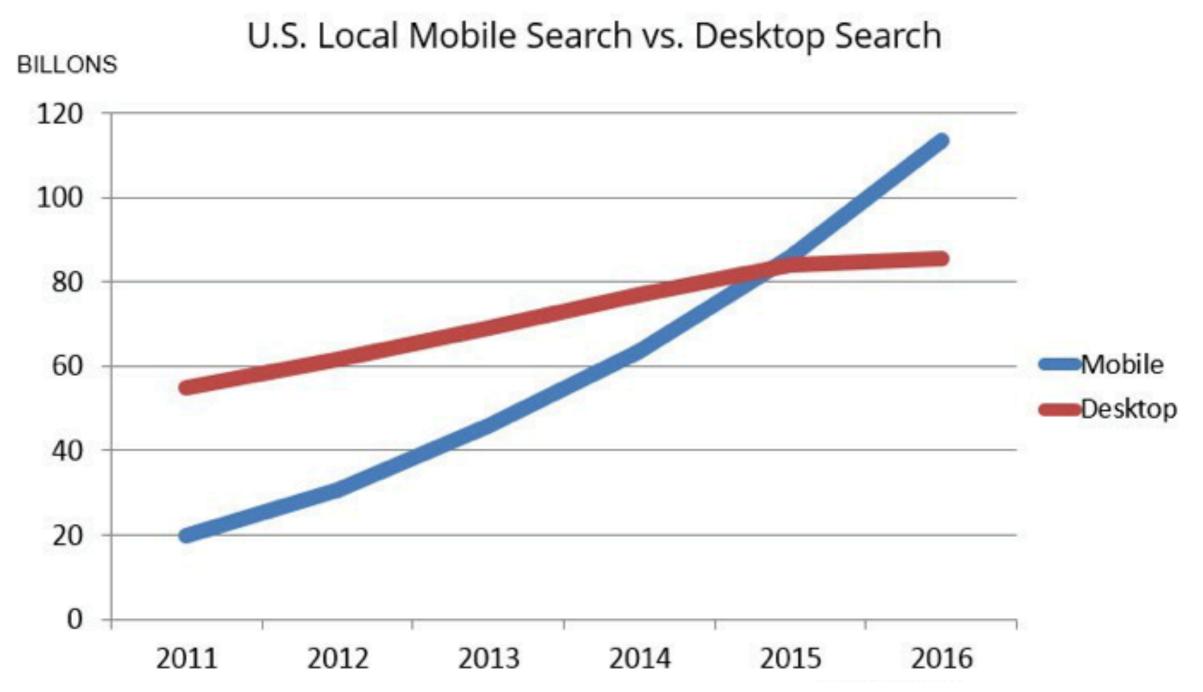


Live Trainings

Join us for live search trainings or watch past trainings from search experts here at Google.

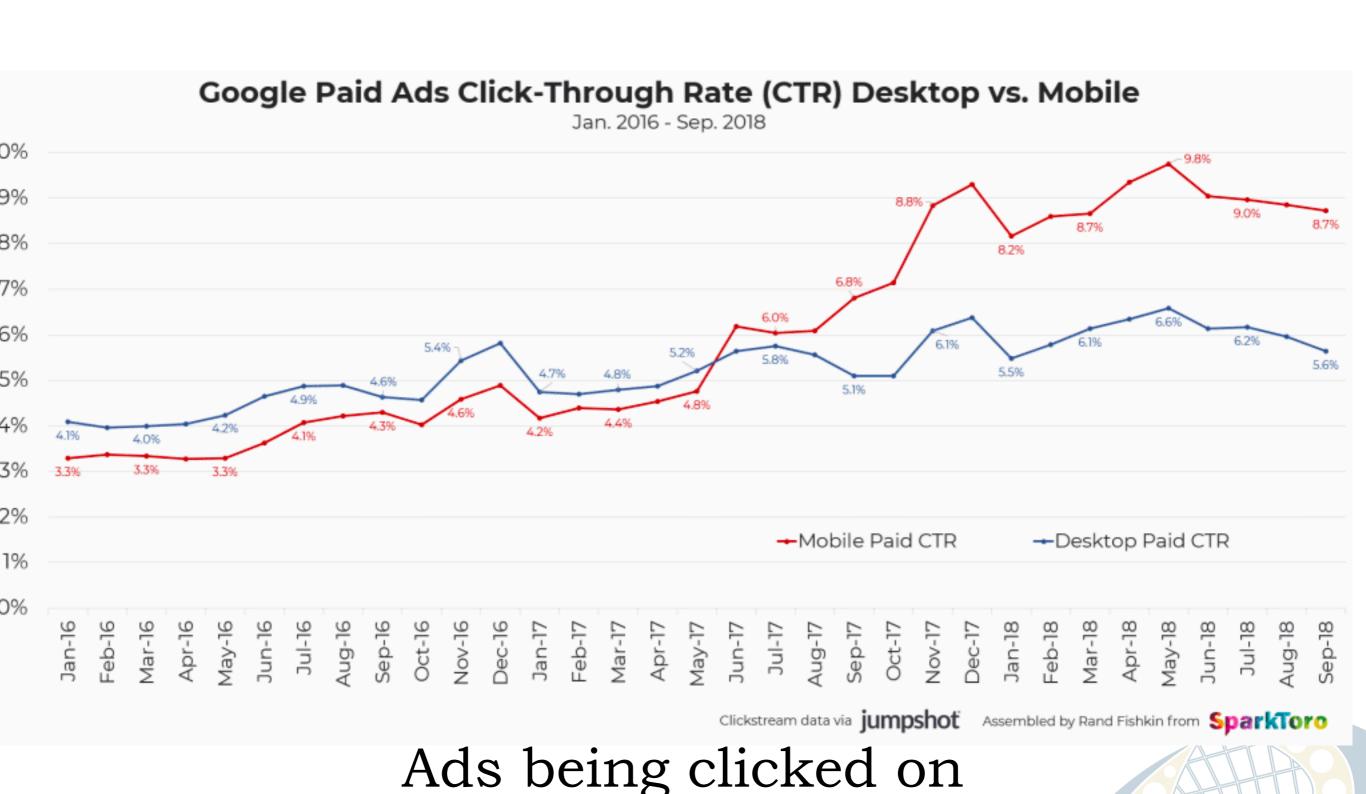
Start training





Users searching for local businesses





sparktoro.com

65%

60%

55%

50%

45%

25%

Google Searches Resulting in ZERO Clicks



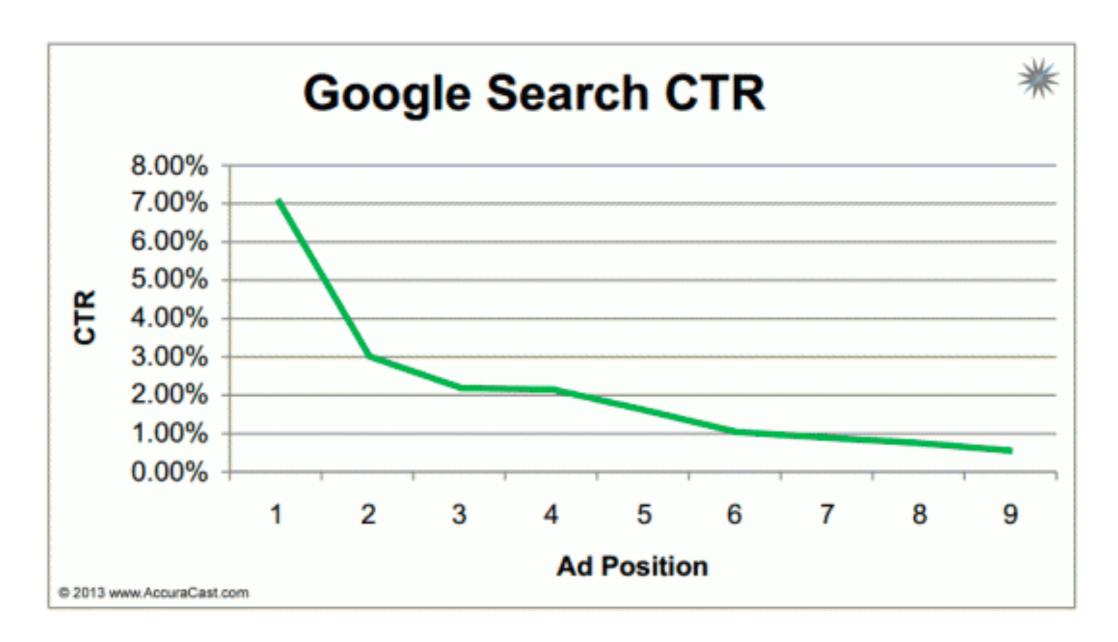


-ep-16 Mar-16 Apr-16 May-16 Jun-16 Aug-16 Sep-16 Oct-16 Nov-16 Dec-16 Nov-17 Dec-17 Jan-18 Feb-18 Mar-18 Jul-16 Feb-17 Mar-17 Apr-17 May-17 Oct-17 Apr-18 Jan-17 Jun-17 Jul-17 Aug-17 Sep-17 1ay-18

Clickstream data via jumpshot Assembled by Rand Fishkin from SparkToro



What's going on?



CTR on Google Sidebar Ads





paid vs organic?

Stats for organic results



WITHOUT SEARCH ENGINES THE WEB WOULDN'T SCALE

- No incentive in creating content unless it can be found.
 - Taxonomies, bookmarks can't keep up
 - Or can they? (del.icio.us)
- The web is both a technology artifact and a social environment
 - "The Web has become the 'new normal' in the America way of life; those who don't go online constitute an every-shrinking minority" [Pew Foundation report, January 2005]

WITHOUT SEARCH ENGINES THE WEB WOULDN'T SCALE

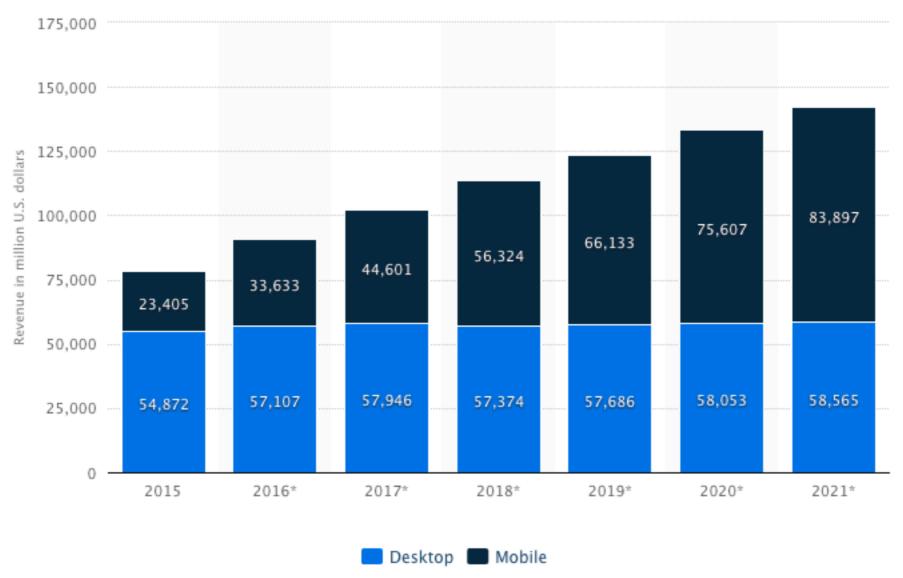
- Search engines make aggregation of interest possible:
 - Create incentives for very niche players
 - Economical specialized stores, providers, etc.
 - Social narrow interests, specialized communities
- The acceptance of search interaction makes "unlimited selection" stores possible
 - Amazon, Netflix, etc.

WITHOUT SEARCH ENGINES THE WEB WOULDN'T SCALE

- Search turned out to be the best mechanism for advertising on the web,.
 - Growing very fast (entire US advertising industry is \$250 billion though)
 - \$15 billion plus industry in 2009
 - \$36 billion in 2012
 - \$95.4 billion in 2017 (cite)



WITHOUT SEARCH ENGINES THE WEB WOULDN'T SCALE



Digital search advertising revenue worldwide from 2015 to 2021, by device (in million U.S. dollars)

Statista 2017

OVERVIEW

OVERVIEW

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



CLASSIC IR ASSUMPTIONS

- Corpus: Fixed document collection
- Goal: Retrieve information content relevant to information need





CLASSIC IR GOAL

• Classic "Relevance"



- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)



- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)
 - R(Q,D) is averaged over users, U, and contexts, C



- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)
 - R(Q,D) is averaged over users, U, and contexts, C
 - Maximize R(Q,D) instead of R(Q,D,U,C)



- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)
 - R(Q,D) is averaged over users, U, and contexts, C
 - Maximize R(Q,D) instead of R(Q,D,U,C)
 - Context is ignored



CLASSIC INFORMATION RETRIEVAL

CLASSIC IR GOAL

- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)
 - R(Q,D) is averaged over users, U, and contexts, C
 - Maximize R(Q,D) instead of R(Q,D,U,C)
 - Context is ignored
 - Individuals are ignored



CLASSIC INFORMATION RETRIEVAL

CLASSIC IR GOAL

- Classic "Relevance"
 - For each query, Q, and stored document, D, in a corpus there exists a relevance score: R(Q,D)
 - R(Q,D) is averaged over users, U, and contexts, C
 - Maximize R(Q,D) instead of R(Q,D,U,C)
 - Context is ignored
 - Individuals are ignored
 - Corpus is static



OVERVIEW

OVERVIEW

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



WEB IR: DIFFERENCES FROM TRADITIONAL IR

- On the web, search and ads are intricately connected
- The web is huge
- The web is a rapidly changing collection.
- There is spam on the web
 - Adversarial IR
 - Huge difference from traditional IR
- One interface for hugely divergent needs
 - Queries, Maps, Stocks, Weather, Calculations

- Early keyword-based engines
 - (1995-1997) Altavista, Excite, Infoseek, Inktomi
- Paid placement ranking
 - Goto.com -> Overture.com -> Yahoo!
 - Results based on auction for keyword placement



www.goto.com/d/search/;\$sessionid\$AQ4214AAAFDRSQFIEF3QPUQ7type=home8tm=18Keywords=Wilmington+ Wilmington real estate. Access 75% of all users now! Premium Listings reach 75% of all Internet users, Sign up for Premium Listings today! Wilmington Real Estate - Buddy Blake Wilmington's information and real estate guide. This is your onanything to do with Wilmington. bb of www.buddyblake.com (Cost to advertis : \$0.38) ow! stings of all 2. Coldwell Banker Sea Coast Realty 40.5 Wilmington's number one real estate company. www.cbseacoast.com (Cost to advertiser: \$0.37) stings 3. Wilmington, NC Real Estate Becky Bullard Everything you need to know about buying or selling a home of on my Web site! WWW.IWWC.Net (Cost to advertiser: \$0.35)



- (1998+) Link-based ranking pioneered by Google
 - Links added the idea of "authoritativeness" to "relevance"
 - Blew away all early engines save Inktomi
 - Great user experience looking for a business model
 - Meanwhile Goto/Overture's annual revenues were nearing \$1 billion



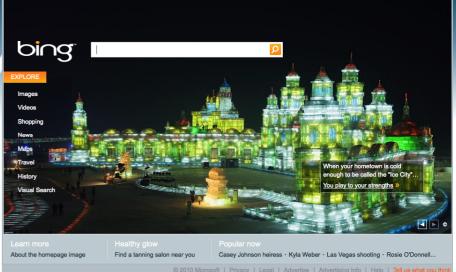
- Result
 - Google:
 - Added paid placement ads on the side
 - Differentiated from search results
 - Yahoo! built a similar architecture
 - Buys Overture for paid placement
 - Buys Inktomi for search



- 2004
 - Microsoft begins in house development of a search engine called Live
- May 28, 2009
 - Microsoft rebrands Live Search to Bing!
 - Search Engine wars intensify
 - New innovations appears at every turn
 - Technology becomes much more closely guarded

WEB INFORMATION RETRIEV bing

- 2004
 - Microsoft begins in house development of a search engine called Live
- May 28, 2009
 - Microsoft rebrands Live Search to Bing!
 - Search Engine wars intensify
 - New innovations appears at every turn
 - Technology becomes much more closely guarded



HISTORY

- Internationally
 - Chinese search engine Baidu "owns" Chinese search
 - Launched around 2000, specializes in Chinese

content



BEFORE (1/7/2010)

- 1. google.com
- 2. facebook.com
- 3. youtube.com
- 4. yahoo.com
- 5. live.com
- 6. wikipedia.org
- 7. blogger.com
- 8. baidu.com
- 9. msn.com
- 10. yahoo.co.jp

- 13. twitter.com
- 15. google.cn
- 22. bing.com
- 29. google.co.jp
- 56. ask.com
- 64. cnn.com

Top sites globally according to <u>alexa.com</u>



(8/29/2015)

- 1. Google.com
- 2. Facebook.com
- 3. Youtube.com
- 4. <u>baidu.com</u>
- 5. Yahoo.com
- 6. Amazon.com
- 7. Wikipedia.org
- 8. Qq.com
- 9. Twitter.com
- 10. Google.co.in

- 11. Live.com
- 12. Taobao.com
- 13. Sina.com.cn
- 14. Linkedin.com
- 15. Yahoo.co.jp
- 16. Weibo.com
- 17. Ebay.com
- 18. Google.co.jp
- 19. Yandex.ru
- 20. blogspot.com

- 21. Vk.com
- 22. Hao123.com
- 23. T.co
- 24. Bing.com
- 25. Google.de
- 26. Instagram.com
- 27. Msn.com
- 28. Aliexpress.com
- 29. Microsoft.com
- 30. google.co.uk



(8/31/2017)

- 1 Google.com
- 2 Youtube.com
- 3 Facebook.com
- 4 Baidu.com
- 5 Wikipedia.org
- 6 Yahoo.com
- 7 Reddit.com
- 8 Google.co.in
- 9 Qq.com
- 10 Amazon.com

- 11 Taobao.com
- 12 Google.co.jp
- 13 Twitter.com
- 14 Tmall.com
- 15 Vk.com
- 16 Live.com
- 17 Instagram.com
- 18 Sohu.com
- 19 Sina.com.cn
- 20 Weibo.com

- 21 Jd.com
- 22 360.cn
- 23 Google.de
- 24 Google.co.uk
- 25 Google.ru
- 26 Google.fr
- 27 Google.com.br
- 28 Linkedin.com
- 29 List.tmall.com
- 30 Google.com.hk



TODAYISH (9/3/2019)

4					
	Google.com	11	Yahoo.com	21	netflix.com
	Youtube.com		360.cn	22	blogspot.com
	Tmall.com	13	Amazon.com		alipay.com
4	Baidu.com	14	jd.com		csdn.net
	qq.com				Yahoo.co.jp
	Sohu.com		Weibo.com		Okezone.com
7	Facebook.com		Pages.tmall.com		
	Taobao.com		live.com		instagram.com
	login.tmall.com		reddit.com		Google.com.hk
10	Wikipedia		vk.com		office.com



OVERVIEW

OVERVIEW

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam





search engine optimization

Search

Personalized Results 1 - 10 of

Ads search engine optimization.

onsored Links

Results

Ads

Search Engine Optimize

Blogs News

SEOP.com Guaranteed Top Ranking w/ Warranty. Free Site Analysis! 877-231-1557

Guaranteed Page 1 Ranking

www.berankednumber1.com Guaranteed Page 1 Rankings \$49.95 No Charge Until You are on Page 1

Search engine optimization - Wikipedia, the free encyclopedia

Search engine optimization (SEO) is the process of improving the volume and quality of traffic to a web site from search engines via "natural" ("organic" or ... en.wikipedia.org/wiki/Search_engine_optimization - 87k - Cached - Similar pages - Note to

Search Engine Optimization, Google Optimization - SEO Chat

Search Engine Optimization, Google Optimization - SEO Chat.

www.seochat.com/ - 111k - Cached - Similar pages - Note this

Search Engine Optimization (SEO) Marketing Firm & Placement Company

Offers search engine optimization (SEO) marketing services & placement since 1998. Submit your website URL to 40 major search engines for FREE! www.submitexpress.com/ - 42k - Cached - Similar pages - Note this

News results for search engine optimization



<u>CIBER Selected as E-Commerce Vendor by Elite Island Resorts</u> - Jan 3, 2008
Their **search engine** marketing program will help us lower acquisition costs ... CIBER's advanced **search engine** marketing services will help Elite direct more ...
FOX News - 10 related articles »

bruceclay.com - Search Engine Optimization - SEO Training, Tools ...

Search Engine Optimization, ranking, placement, and submission tutorial. Free step-by-step SEO tools and advice. SEO training and services offered. ... www.bruceclay.com/web_rank.htm - 87k - Cached - Similar pages - Note this

Inteliture M Search Engine Optimization, Internet Marketing, and ...

Inteliture™ a professional **search engine optimization** and internet marketing company. Offers internet marketing solutions, **search engine optimization** ...

www.inteliture.com/ - 12k - Cached - Similar pages - Note this

Search engine optimizer

Use Network Solutions online tools to drive business to your web site. marketing.networksolutions.com

Sponsored Links

Search Optimization Firm

Agorithmic Receive a free analysis.

Company

since 1998 with proven results.
www.iClimber.com

Get Optimization Help Now

Top SEO Firms Want Your Business. Fast, Free Competitive Quotes! www.TopSeos.com/SEO

Check your SEO for Free

PPC vs Natural search Keyword ranks costs & robot stats: 15 days free www.ClickTracks.com/15_Days_Free

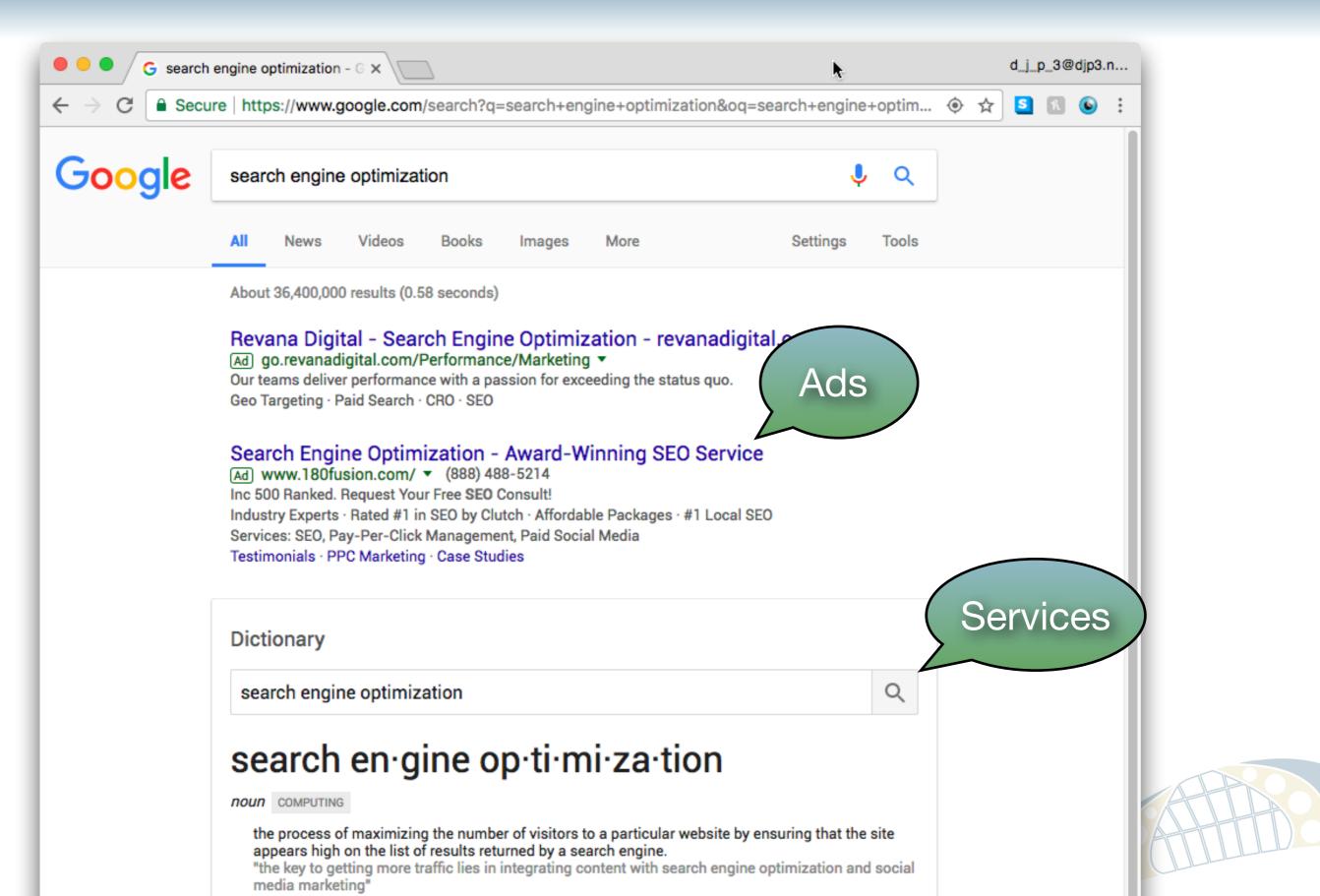
Search Engine Marketing

Boost Online Traffic and Sales!
Free Site Optimization Analysis.
www.corporatesearchoptimization.com

Free Website Visitors

Free Visitors Plus Top 10 Positions In 8 Hours! FREE Trial Offer. www.EngineSeeker.com





Testimonials · PPC Marketing · Case Studies

Services

Dictionary

search engine optimization

Q

Feedback

search en-gine op-ti-mi-za-tion

noun COMPUTING

the process of maximizing the number of visitors to a particular website by ensuring that the site appears high on the list of results returned by a search engine.

"the key to getting more traffic lies in integrating content with search engine optimization and social media marketing"

Translations, word origin, and more definitions

What Is SEO / Search Engine Optimization? - Search Engine Land searchengineland.com/guide/what-is-seo ▼

Get started learning all about SEO from the industry's most trusted source, Search Engine basics of search engine optimization, ranking factors ...

Algorithmic Results

Beginner's Guide to SEO (Search Engine Optimization) - Moz

https://moz.com/beginners-guide-to-seo ▼

Dec 18, 2015 - New to SEO? Need to polish up your knowledge? The Beginner's Guide to SEO has been read over 3 million times and provides the ...

Keyword Research · How Usability, Experience and ... · Measuring & Tracking

Search engine optimization - Wikipedia

https://en.wikipedia.org/wiki/Search_engine_optimization •

Search engine optimization (SEO) is the process of affecting the visibility of a website or a web page in a web search engine's unpaid results—often referred to ...

History · Methods · As a Marketing Strategy · International markets

SEO basics: 22 essentials you need for optimizing your site | Search ...

https://searchenginewatch.com > SEO ▼

Jan 21, 2016 - Quite simply, **SEO** is the umbrella term for all the methods you can use to ensure the visibility of your website and its content on search engine ...

[PDF] Search Engine Optimization Starter Guide - Google

www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf •



ADS VS. SEARCH RESULTS

• Google maintains that ads (based on vendors bidding for search queries) do not affect vendors ranking in

search results

Search engine optimization - Wikipedia, the free encyclopedia

Search engine optimization (SEO) is the process of improving the volume and quality of traffic to a web site from search engines via "natural" ("organic" or ... en.wikipedia.org/wiki/Search_engine_optimization - 87k - Cached - Similar pages - Note this

Search Engine Optimization, Google Optimization - SEO Chat

Search Engine Optimization, Google Optimization - SEO Chat. www.seochat.com/ - 111k - Cached - Similar pages - Note this

Search Engine Optimization (SEO) Marketing Firm & Placement Company

Offers search engine optimization (SEO) marketing services & placement since 1998. Submit your website URL to 40 major search engines for FREE! www.submitexpress.com/ - 42k - Cached - Similar pages - Note this

News results for search engine optimization



<u>CIBER Selected as E-Commerce Vendor by Elite Island Resorts</u> - Jan 3, 2008
Their **search engine** marketing program will help us lower acquisition costs ... CIBER's advanced **search engine** marketing services will help Elite direct more ...
FOX News - 10 related articles »

Sponsored Links

Search engine optimizer

Use Network Solutions online tools to drive business to your web site. marketing.networksolutions.com

Search Optimization Firm

Looking for top rankings? Get real results. Receive a free analysis. www.customermagnetism.com

SEO Company

Search Engine Optimization services since 1998 with proven results. www.iClimber.com



RANKING OF ADS

- Other search engines (Yahoo!, MSN) have made similar statements on occasion
 - Any of them can change at any time
 - Facebook had a class action lawsuit over testing the waters with "Beacon" social ads
- We will ignore the possibility of paid placement ads being interspersed in search results.





RANKING OF ADS

- Goto model:
 - Rank according to how much advertiser pays
- Current model:
 - Balance auction price and relevance
 - Irrelevant ads (few click-throughs)
 - Decrease opportunities for relevant ads
 - Harm the user experience
 - Idea: Well-targeted advertising is good for everyone

PAYING FOR ADVERTISEMENTS - TERMS

- CPM
 - "Cost Per Mil"
 - Pay for 1000 eyeballs
 - Important for branding campaigns
- CPC
 - "Cost per Click"
 - Pay for clicking on ads
 - Important for sales campaigns
 - Related "CTR" = Click Through Rate



WHAT ARE THE STAKES HERE?

• What role is Google playing?





OVERVIEW

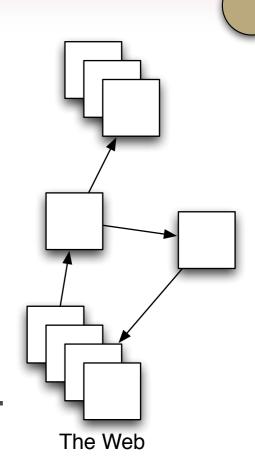
OVERVIEW

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



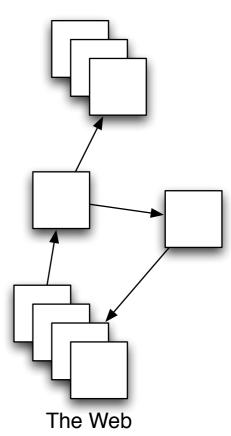
The Web Corpus

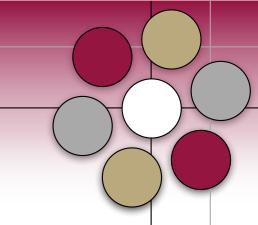
- No design/coordination
- Distributed content creation, linking
- "Democratization of publishing"
- Content includes truth, lies, contradictions, etc.
- Unstructured Data (text, html)
- Semi-Structured (XML, annotated photos)
- Structured (Databases)
- Scale is much larger than previous text corpora



The Web Corpus

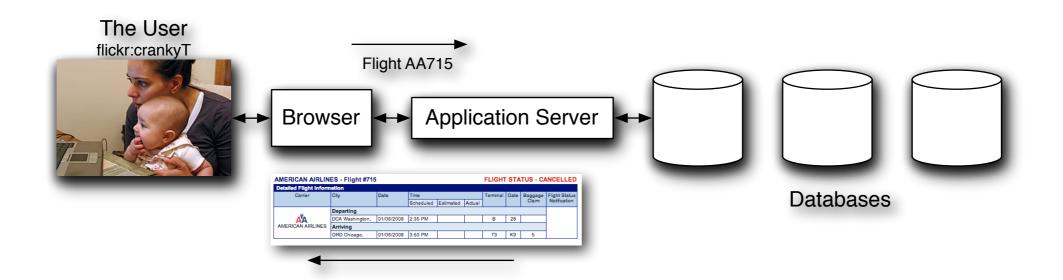
Growth - slowing from "doubling every few months", but still expanding





Dynamic Content

- Content can by dynamically generated
 - There is no static html version
 - Flight status information, evite responses
 - Assembled on request ("?" in URL is a clue)

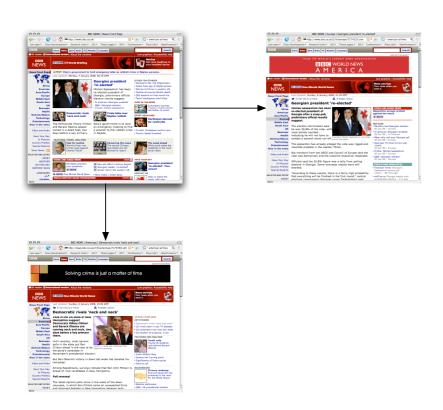


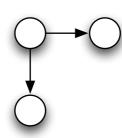
Dynamic Content

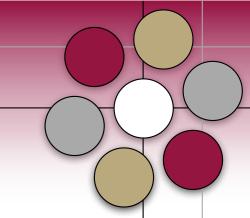
- Most (truly) dynamic content is ignored by search engines
 - Too much to index
 - Static information is more important for search
 - Spider Traps look dynamic
- Actually a lot of "static" content is assembled on the fly also
 - ASP, PHP, JSP, ads, etc....

The Web as a graph

- Web pages are nodes
- Hyperlinks are directed edges

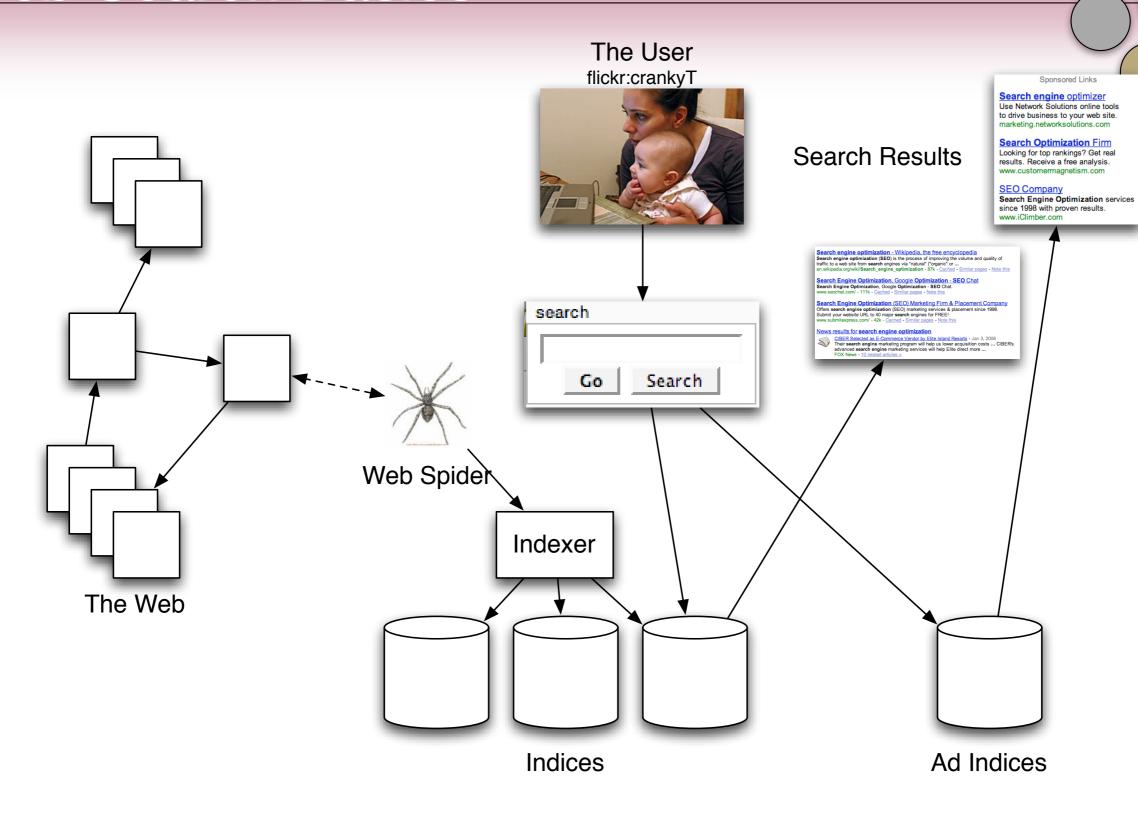




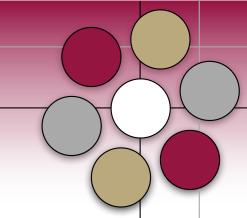


Characteristics of the web

- Significant Duplication
 - 30%-40% is some studies [Brod97, Shiv99]
 - www.copyscape.com
- High linkage
 - more than 8 links per page on average
- Spam
 - Billions of pages of it.



Web Search Basics: Size of the web



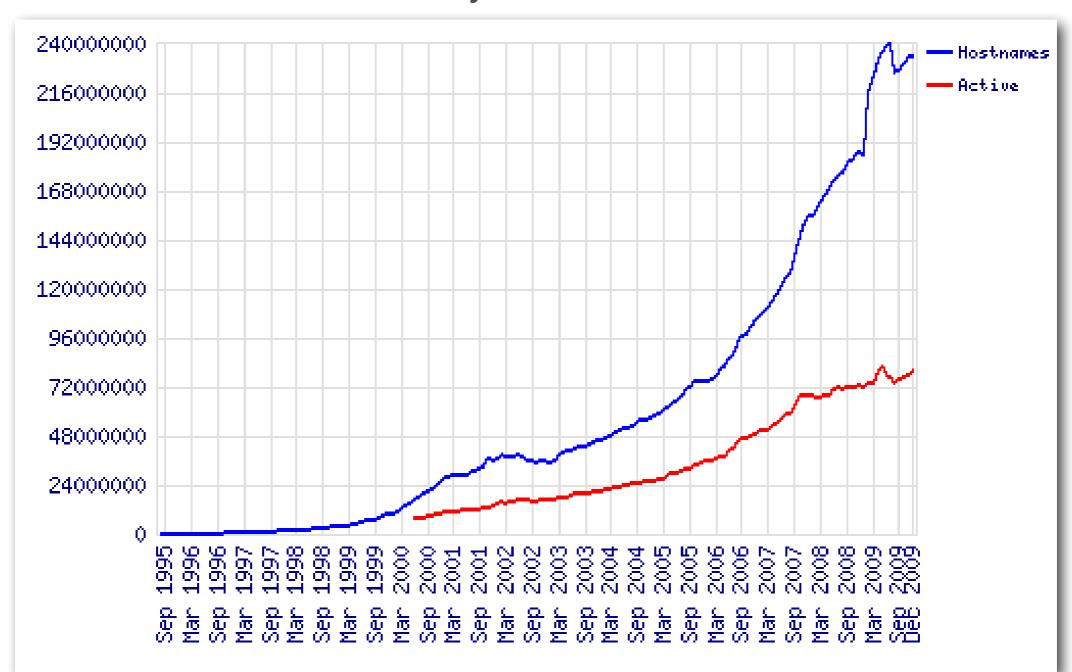
How big is the web?

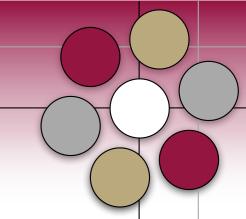
- What is measured?
 - Number of hosts
 - Number of "static" html pages
- Number of hosts netcraft survey
 - <u>http://news.netcraf</u>t.com/archives/web_server_survey.html
 - Monthly report on hosts and servers
- Number of pages
 - Lots of estimates which warrant further discussion

Web Search Basics: Size of the web

How big is the web?

Netcraft Web Server Survey

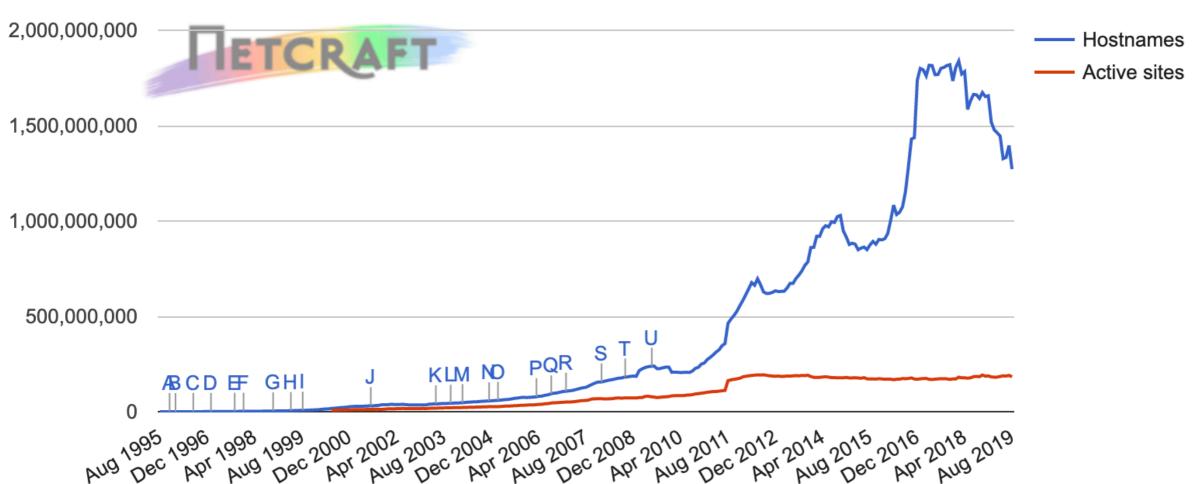


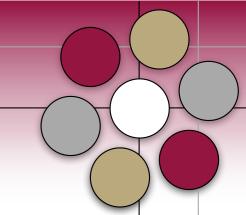


How big is the web?

Netcraft Web Server Survey

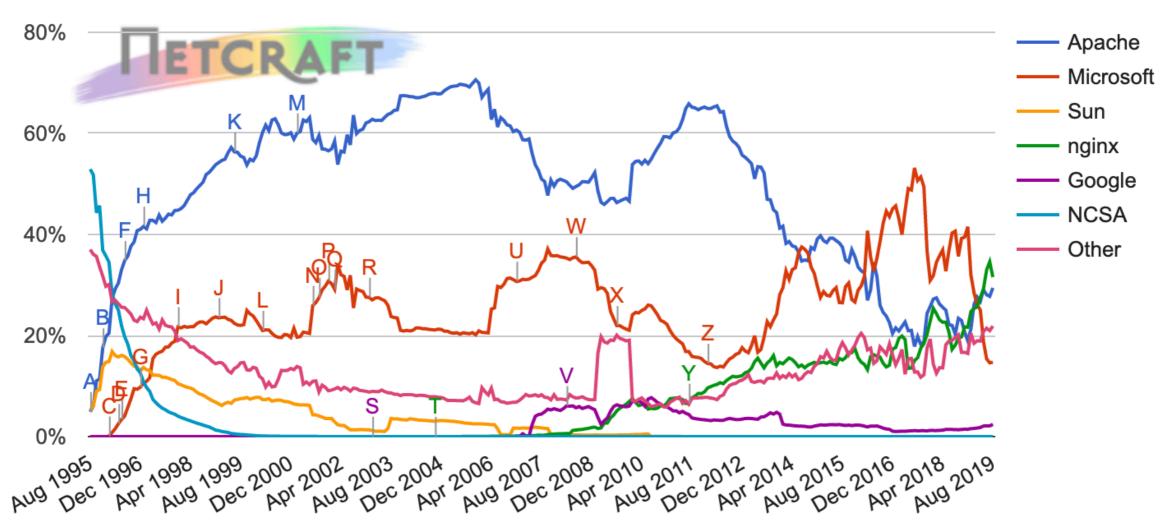






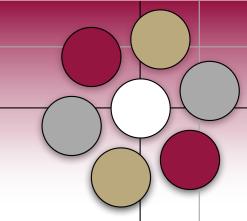
Netcraft Web Server Survey

Web server developers: Market share of all sites



Rate of change

- [Cho00] 720k pages from 270 popular sites sample daily for 5 months in 1999
 - 40% changed weekly, 23% daily
- [Fett02] Massive study: 151M pages checked over a few months
 - Significant changes 7% weekly
 - Any change 25% weekly

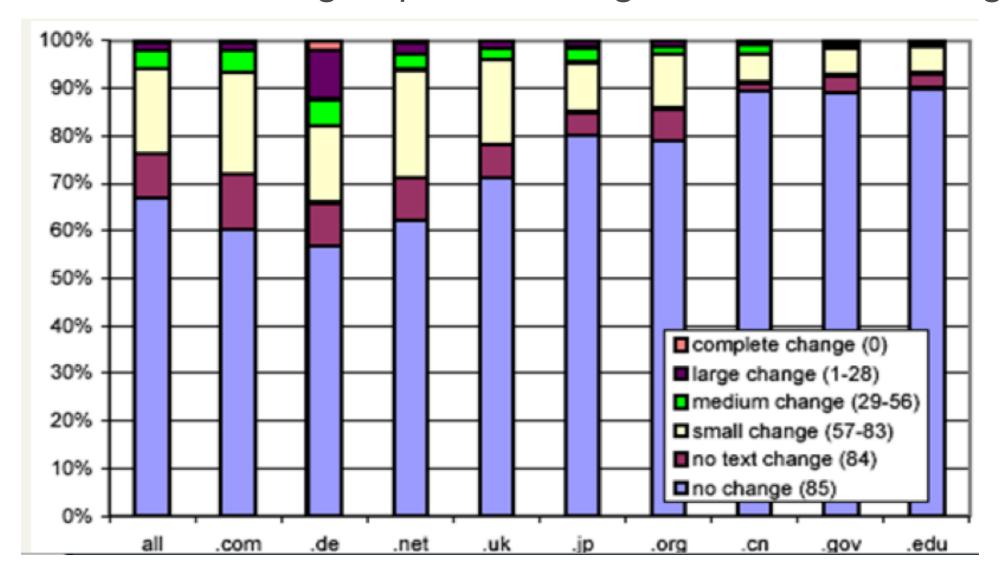


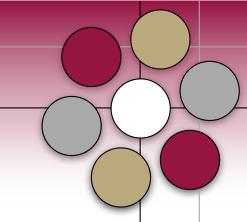
Rate of change

- [Ntul04] 154 large sites recrawled from scratch weekly
 - 8% had new pages ever week
 - 8% die
 - 5% new content
 - 25% new links per week

Rate of change

- Fetterly et al. study in 2002
 - 150 million pages over 11 weekly crawls
 - Bucketed into 85 groups according to amount of change





Web Evolution

- The nature of the web is change
- Not much work on studying web evolution
 - Exception is Fetterly et. al, 2003
- Some effort has been made to extrapolate from small samples using fractal models [Dill et. al. 2001]

The very nature of the web is changing as well

Transforming from a source of information

- Transforming from a source of information
- to what?

- Transforming from a source of information
- to what?
 - a communication platform?

- Transforming from a source of information
- to what?
 - a communication platform?
 - a source of computation?

- Transforming from a source of information
- to what?
 - a communication platform?
 - a source of computation?
 - an application-space?

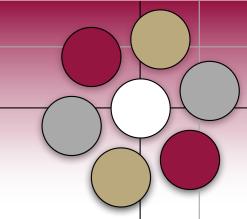
- Transforming from a source of information
- to what?
 - a communication platform?
 - a source of computation?
 - an application-space?
 - a mirror-world?

- Transforming from a source of information
- to what?
 - a communication platform?
 - a source of computation?
 - an application-space?
 - a mirror-world?
 - an augmentation of reality?

- Transforming from a source of information
- to what?
 - a communication platform?
 - a source of computation?
 - an application-space?
 - a mirror-world?
 - an augmentation of reality?
 - a cognitive orthotic?

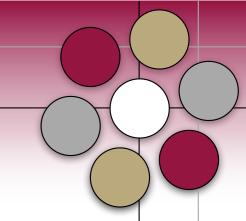


Overview



- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam

Overview

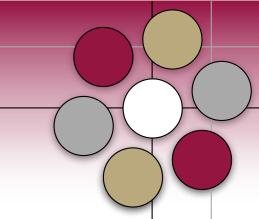


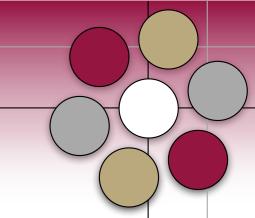
- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web



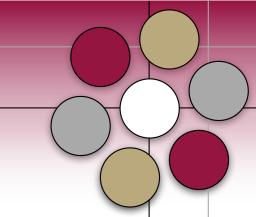
Web Users

Spam

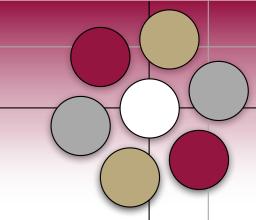




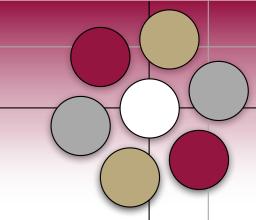
- Informational
 - Want to learn about something (~40%/65%)



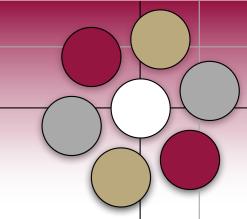
- Informational
 - Want to learn about something (~40%/65%)
- Navigational
 - Want to go to that page (~25%/15%)



- Informational
 - Want to learn about something (~40%/65%)
- Navigational
 - Want to go to that page (~25%/15%)
- Transactional
 - Want to do something (~35%/20%)
 - Access a service, download, shop

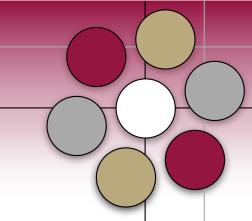


- Informational
 - Want to learn about something (~40%/65%)
- Navigational
 - Want to go to that page (~25%/15%)
- Transactional
 - Want to do something (~35%/20%)
 - Access a service, download, shop
- Others?
 - Exploration, social, etc...



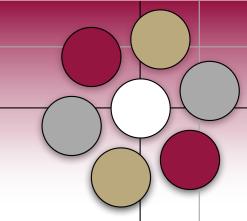
Web Users

- Make ill defined queries
 - Short
 - Reported in 2012: ~3.5 (Dan Russell personal conv.)
 - Allegedly in 2009: 3-4 terms (lots of nuance)
 - Average in 2001: 2.54 terms (80% < 3 words)
 - Average in 1998: 2.35 terms (88% < 3 words) [Silv98]
 - Imprecise terms
 - Suboptimal syntax (no operators)
 - Low effort (spelling mistakes)



Web Users

- Wide Variance in
 - Needs
 - Expectations
 - Knowledge
 - Bandwidth

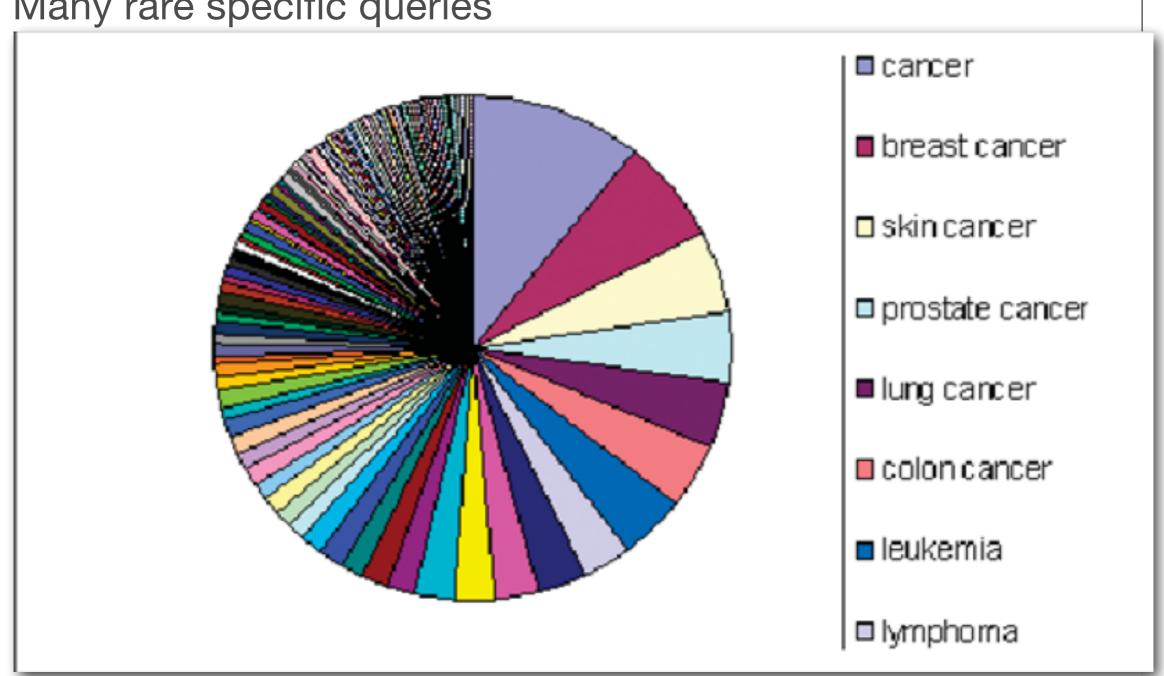


Web Users

- Behavior
 - 85% look over one result screen only
 - 78% of queries are not modified
 - Follow links ("the scent of information")

Power law

- Few popular broad queries
- Many rare specific queries



Top queries

- Many are related to sex
- 2008 Who What How (Google)

Who is...

- 1. who is obama
- who is mccain
- who is palin
- 4. who is lil wayne
- 5. who is miley cyrus
- 6. who is dolla
- 7. who is jonas brothers
- 8. who is chris brown
- who is biden
- 10. who is martin luther

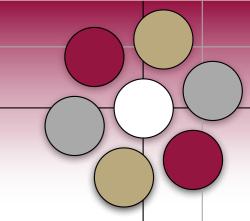
What is...

- what is love
- what is life
- what is java
- 4. what is sap
- 5. what is rss
- 6. what is scientology
- what is autism.
- 8. what is lupus
- 9, what is 3g
- 10. what is art

How to...

- how to draw
- 2 how to kiss
- how to write
- 4 how to cook
- how to tie
- how to hack
- how to run
- 8. how to cite
- 9. how to paint
- how to spell

http://www.google.com/intl/en/press/zeitgeist2008/mind.html



Top queries

- Differences today
- 01/2014 Who What How (Google)

who is

who is

who is red john

who is banksy

who is gossip girl

Press Enter to search.

what is

what is my ip

what is obamacare

what is bitcoin

Press Enter to search.

what is my ip address

how to

how to tie a tie

how to take a screenshot on a mac

how to make french toast

how to get rid of acne

Press Enter to search.

• 09/2015

who is

who is a

who is job

who is the next bachelor

who is the arkham knight

Press Enter to search.

what is

what is my ip

what is flakka

what is my ip address

what is **gluten**

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to write a cover letter

how to screenshot on pc

Press Enter to search.

who is

who is

who is red john

who is banksy

who is gossip girl

Press Enter to search.

who is

who is a

who is job

who is the next bachelor

who is the arkham knight

Press Enter to search.

what is

what is my ip

what is obamacare

what is bitcoin

what is my ip address

Press Enter to search.

what is

what is my ip

what is flakka

what is my ip address

what is gluten

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to make french toast

how to get rid of acne

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to write a cover letter

how to screenshot on pc

Press Enter to search.

9/2019

who is

who is

who is red john

who is banksy

who is gossip girl

Press Enter to search.

who is

who is a

who is job

who is the next bachelor

who is the arkham knight

Press Enter to search.

what is

what is my ip

what is obamacare

what is bitcoin

what is my ip address

Press Enter to search.

what is

what is my ip

what is flakka

what is my ip address

what is gluten

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to make french toast

how to get rid of acne

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to write a cover letter

how to screenshot on pc

Press Enter to search.

9/2019

who is

who is

who is brock turner

whois domain

who is allah

who is in custody

who is america

who is **banksy**

who is the next bachelor

who is cardi b

who is mac miller

who is

who is

who is red john

who is **banksy**

who is gossip girl

Press Enter to search.

who is

who is a

who is job

who is the next bachelor

who is the arkham knight

Press Enter to search.

what is

what is my ip

what is obamacare

what is bitcoin

what is my ip address

Press Enter to search.

what is

what is my ip

what is flakka

what is my ip address

what is gluten

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to make french toast

how to get rid of acne

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to write a cover letter

how to screenshot on pc

Press Enter to search.

9/2019

who is

who is

who is brock turner

whois domain

who is allah

who is **in custody**

who is america

who is **banksy**

who is the next bachelor

who is cardi b

who is mac miller

what is

what is

what is the name

what is the capital of

what is the meaning of

what is my ip

what is your name

what is google

what is my name

what is science

what is the definition of

who is

who is

who is red john

who is **banksy**

who is gossip girl

Press Enter to search.

who is

who is a

who is job

who is the next bachelor

who is the arkham knight

Press Enter to search.

what is

what is my ip what is obamacare

what is bitcoin

what is my ip address

Press Enter to search.

what is

what is my ip what is flakka what is my ip address what is gluten

Press Enter to search.

how to

how to tie a tie

how to take a screenshot on a mac

how to make french toast

how to get rid of acne

Press Enter to search.

how to

how to tie a tie how to take a screenshot on a mac how to write a cover letter

Press Enter to search.

how to screenshot on pc

9/2019

who is

who is

who is brock turner

whois domain

who is allah

who is in custody

who is america

who is **banksy**

who is the next bachelor

who is cardi b

who is mac miller

what is

what is

what is the name

what is the capital of

what is the meaning of

what is my ip

what is your name

what is google

what is my name

what is **science**

what is the definition of

how to

how to screenshot on mac

how to tie a tie

how to delete your instagram

how to write a check

how to make slime

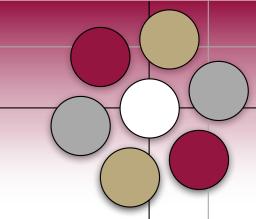
how to train your dragon

how to screenshot on pc

how to hard boil eggs

how to write a cover letter

how to lose weight

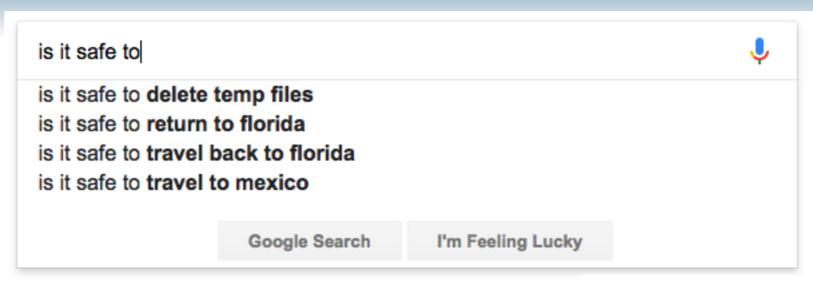


Top queries

- Live demo WARNING this is not very safe....
 - "Is it safe to"
 - "Is it legal to"
 - "why does"
 - "why doesn't
 - "why is there"
 - "why isn't"
 - "americans are"

EXERCISE

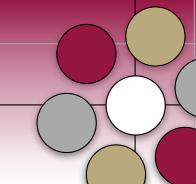
WORK IN TEAMS



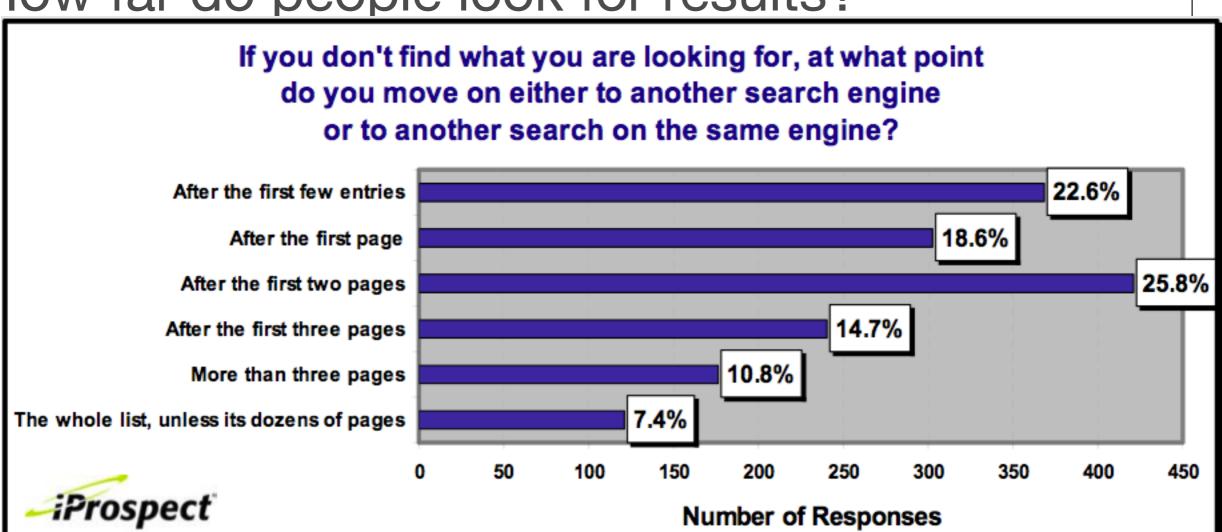
Report inappropriate predictions

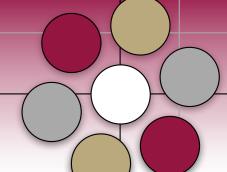
Imagine you have the resources of Google. Propose a way to generate query completions queries like the above example. Present your proposal.



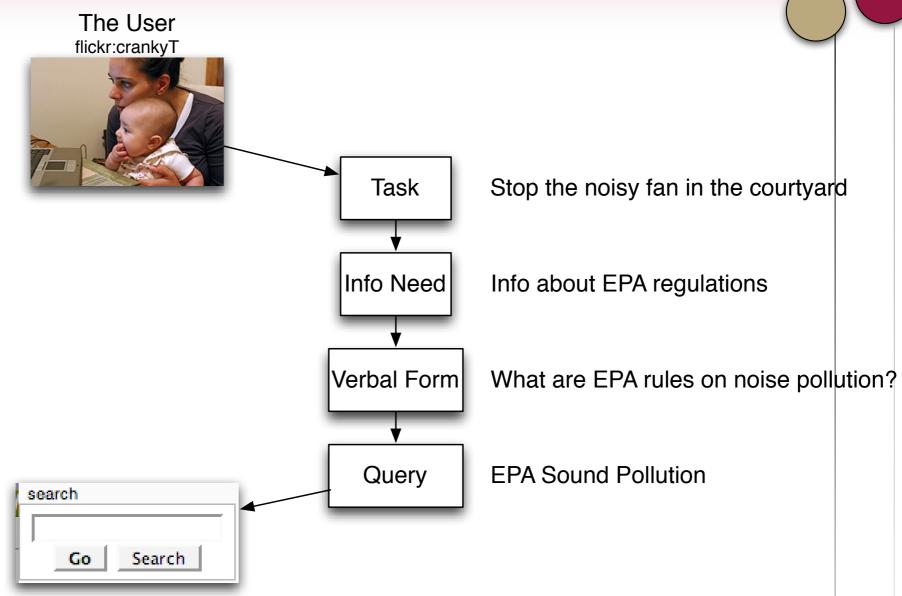


How far do people look for results?





True Example *



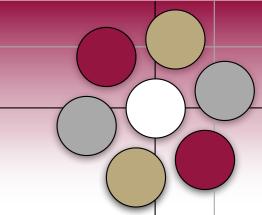
"To Google or to GoTo" Business Week Online 9/28/2001

EXERCISE

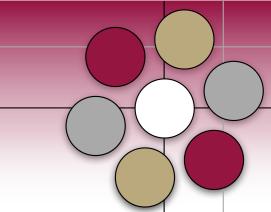
WORK IN TEAMS

What are 4 things that would have to be true about a page of search engine results for you to consider them high-quality?

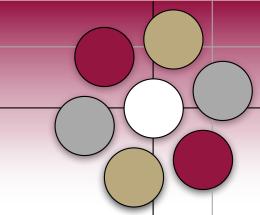




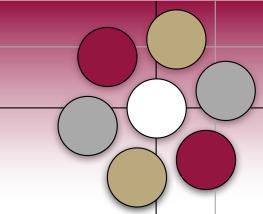
- Quality of pages
 - Classic IR relevance
 - Also important:
 - Trust
 - Duplicate elimination
 - Readability
 - Fast Access
 - No pop-ups



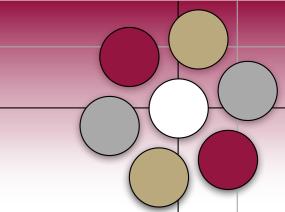
- Precision is more important than recall
 - Precision:
 - How precise is a portal in locating relevant results?
 - Recall
 - How thorough is the coverage of available relevant results?
- Precision with 1 result, 10 results, 2-3 pages of results.
- When is recall important?



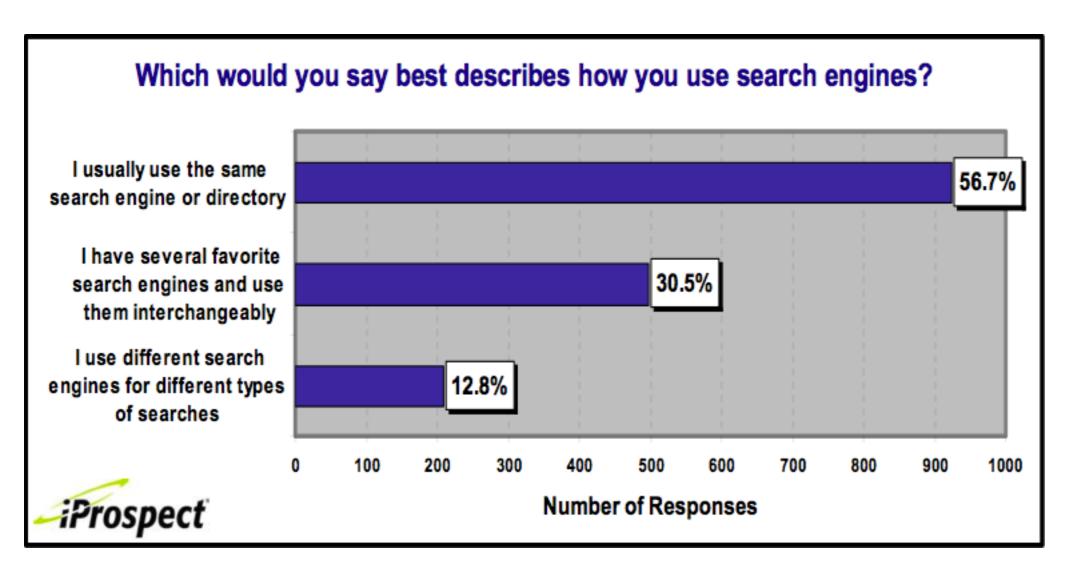
- Recall is sometimes important:
 - Googling for a new doctor
 - Googling a prospective employee
 - Googling your date

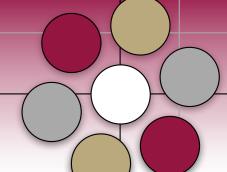


- Good U/I
 - Simple
 - No Clutter
- Pre and post processing tools
 - Spell check ("Did you mean?")
 - Suggested alternative searches
 - Links to resources (maps, images, stock quotes)
- Able to deal with typical behavior
 - e.g., a URL typed into a search box

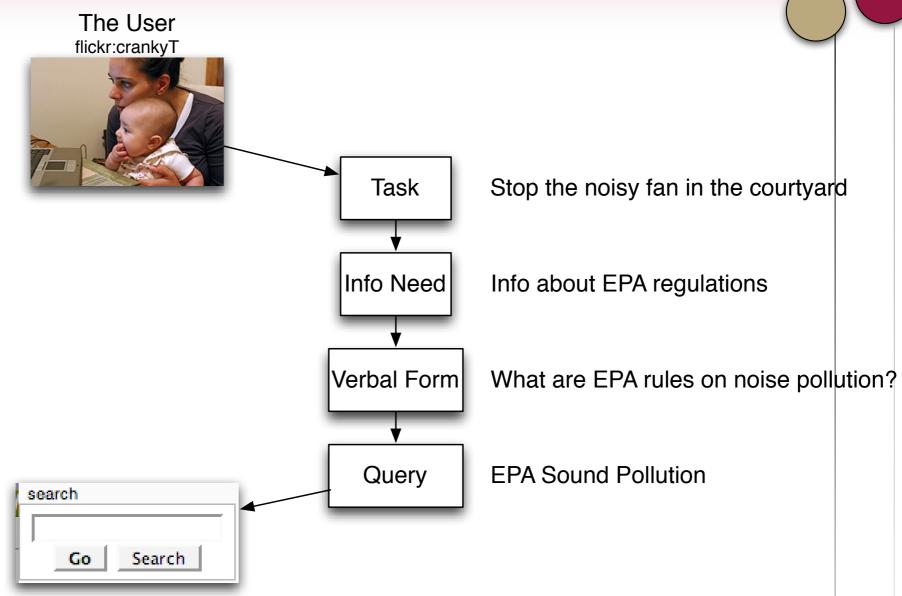


Loyalty to a given search engine





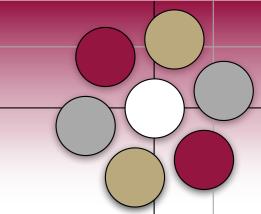
True Example *



"To Google or to GoTo" Business Week Online 9/28/2001

Answering "the need behind the query"

- The query is often an imprecise indicator of what the user really wants
- What can we do to get a better handle on the underlying information need?
- Query language
 - Adjust rank of English results for a Japanese query
- Use user context
 - In particular geographic context



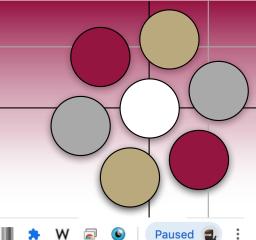
Answering "the need behind the query"

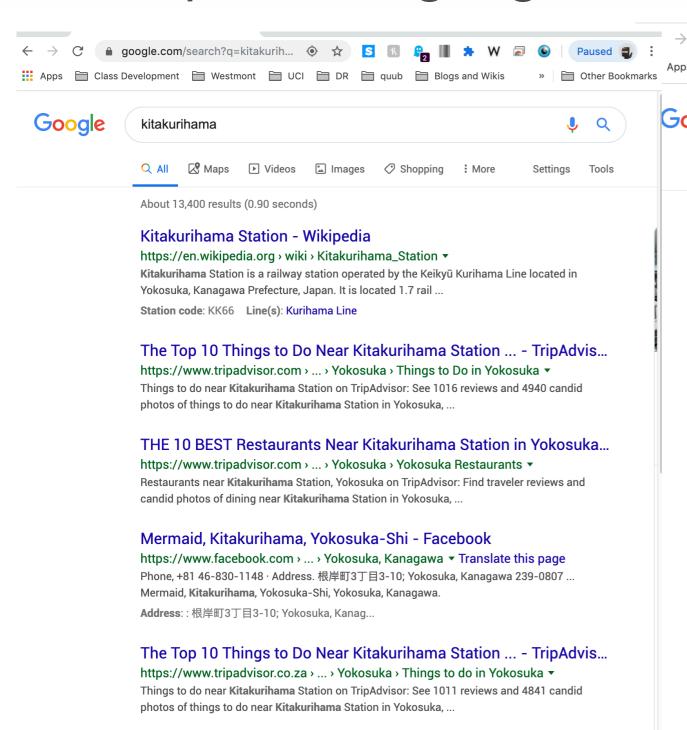
- Guess what type of information the user wants
 - a web page?
 - a map?
 - a stock price?
 - what else?
- Correct queries
 - Suggest correct spellings
 - Suggest related searches (google-fu)

Vokocuka/Hayama farm acaparatiya/Kitakurihama Branch)



Examples - language







北久里浜駅 - Wikipedia

google.co.jp/search?source=hp...

https://ja.wikipedia.org > wiki > 北久里浜駅 ▼

北久里浜駅(きたくりはまえき)は、神奈川県横須賀市根岸町二丁目にある京浜急行電鉄久里浜線の駅である。駅番号はKK66。目次. 1 年表; 2 駅構造. 2.1 のりば. 3 利用状況; 4 駅周辺; 5 バス路線: 6 隣の駅; 7 脚注; 8 出典; 9 関連項目; 10 外部リンク ...

所属事業者: 京浜急行電鉄 駅番号: ○KK66 所属路線: ■久里浜線 駅構造: 地上駅

年表・利用状況・駅周辺・バス路線

Kitakurihama Station - Wikipedia

https://en.wikipedia.org > wiki > Kitakurihama_Station ▼ このページを訳す

Kitakurihama Station is a railway station operated by the Keikyū Kurihama Line located in Yokosuka, Kanagawa Prefecture, Japan. It is located 1.7 rail kilometers from the Horinouchi Station, and 54.0 rail kilometers from the northern terminus ...

Station code: KK66 Line(s): Kurihama Line

The Top 10 Things to Do Near Kitakurihama Station ... - TripAdvisor

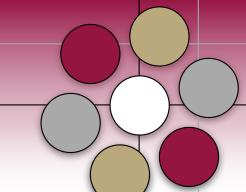
https://www.tripadvisor.com → AttractionsNear-g298174-... ▼ このページを訳す

Things to do near **Kitakurihama** Station on TripAdvisor: See 1016 reviews and 4940 candid photos of things to do near **Kitakurihama** Station in Yokosuka, Kanagawa Prefecture.

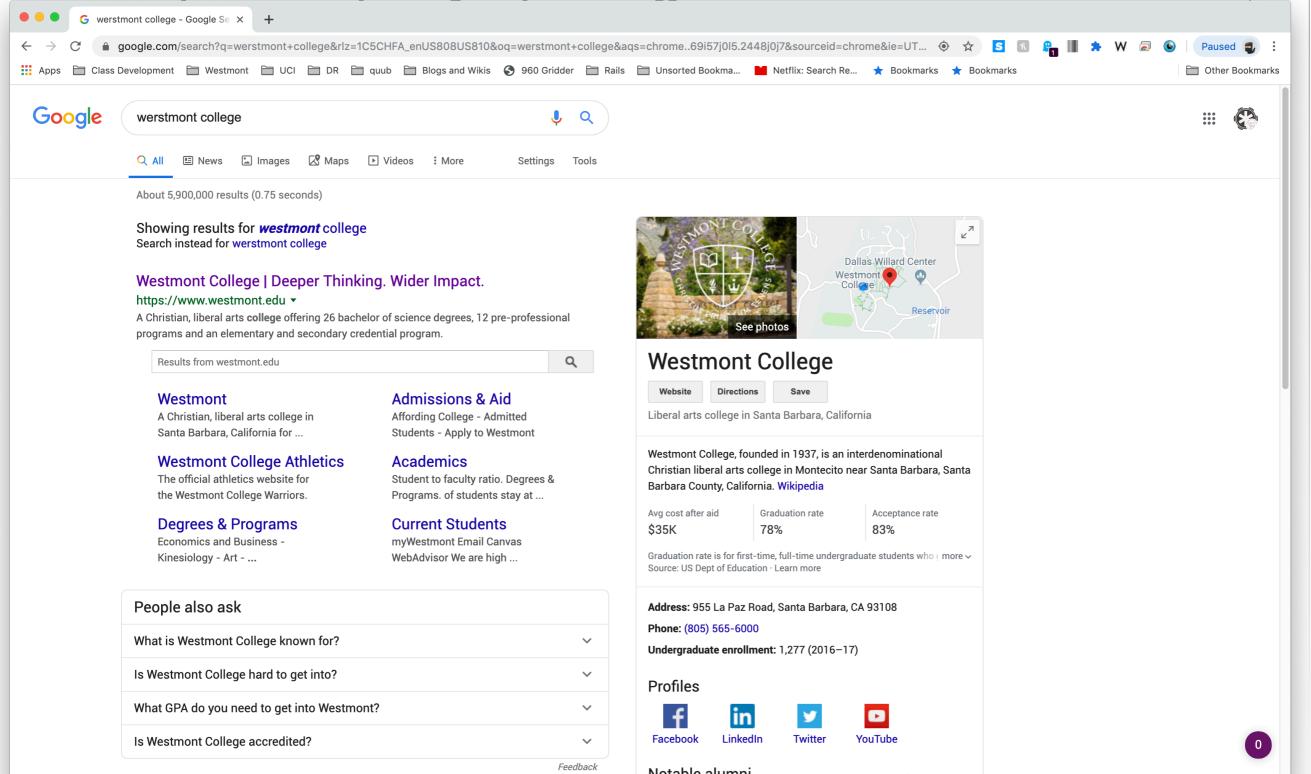
THE 10 BEST Restaurants Near Kitakurihama Station in Yokosuk...

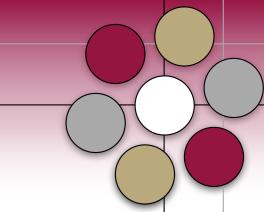
https://www.tripadvisor.com › ... › Yokosuka Restaurants ▼ このページを訳す

Restaurants near **Kitakurihama** Station, Yokosuka on TripAdvisor: Find traveler reviews and candid photos of dining near **Kitakurihama** Station in Yokosuka, Kanagawa Prefecture.

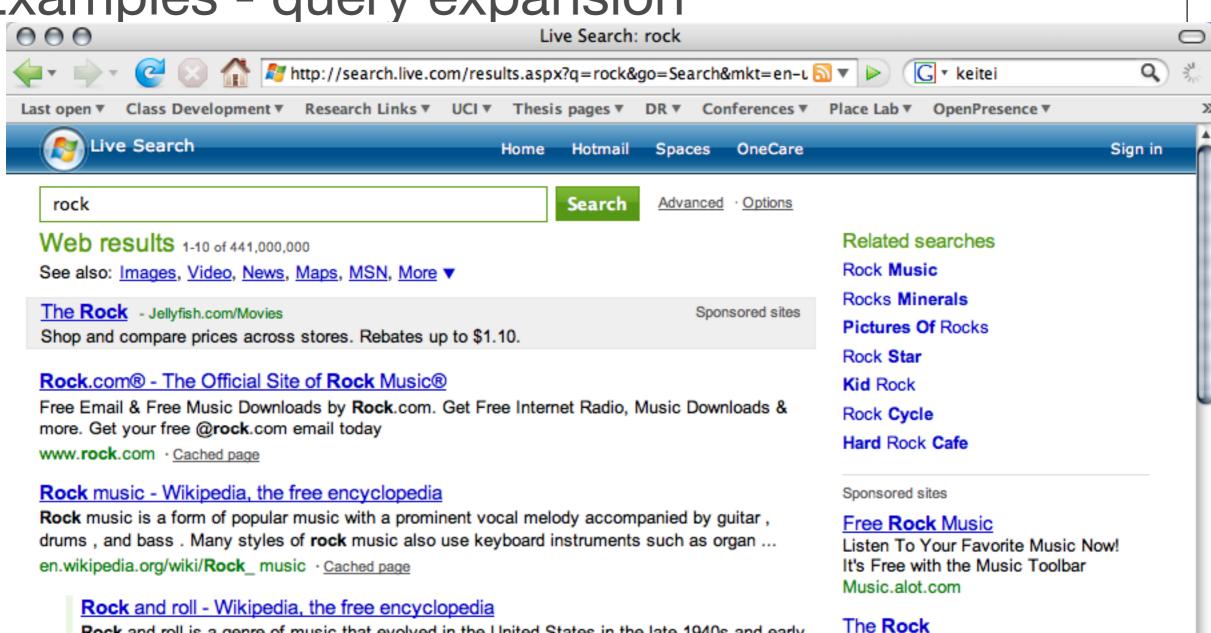


Examples - query spelling





Examples - query expansion



Rock and roll is a genre of music that evolved in the United States in the late 1940s and early 1950s, and quickly spread to the rest of the world.

en.wikipedia.org/wiki/Rock_and_roll · Cached page Show more results from en.wikipedia.org

Rock - Australia's Climbing Magazine

Subscription information. Overview of current issue. Classifieds. Back issues and guide books for

Interviews with Stars, Newsmakers & Icons at The Biography Channel®. www.biography.com

Rock Videos

Enjoy Savings & Selection on Rock Videos! Shopzilla.com



rock

Web

rock - Bing

مر

Maps

My saves

241,000,000 RESULTS

Images

Any time ▼

Videos

Rock.com: Band T-Shirts Rock Music Apparel Rock Merchandise

News

www.rock.com -

Rock band t-shirts, apparel, music posters, & more by Rock.com. Shop for rock and roll band merchandise, accessories, plus hard to find rock band t-shirts

Music · Men's T-Shirts · Login · Search at Rock · Kiss · Beatles

Rock - definition of rock by The Free Dictionary

www.thefreedictionary.com/rock -

rock 1 (rŏk) n. 1. Relatively hard, naturally formed mineral or petrified matter; stone. 2. a. A relatively small piece or fragment of such material. b. A relatively ...

Acronyms · Legal · Financial · Idioms

Images of rock

bing.com/images









See more images

Rock | Define Rock at Dictionary.com

www.dictionary.com/browse/rock -

Rock definition, a large mass of stone forming a hill, cliff, promontory, or the like. See more.

RockAuto

www.rockauto.com -

RockAuto ships auto parts and body parts from over 300 manufacturers to customers' doors worldwide, all at warehouse prices. Easy to use parts catalog.

Videos of rock

bing.com/videos











MSN

Beached baby shark gets rock pool refuge

MSN

Beached baby shark gets rock pool refuge

MSN

Rock Me Amadeus -Falco

VEVO

See more videos of rock

See results for

Dwayne Johnson (Actor)

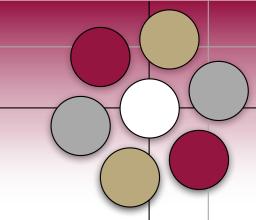
Dwayne Douglas Johnson, also known by his ring name The Rock, is an American actor, producer and professional wrestler. Johnson was a college football player for the University of Miami, where he won a national champions...



Rock (Geology)

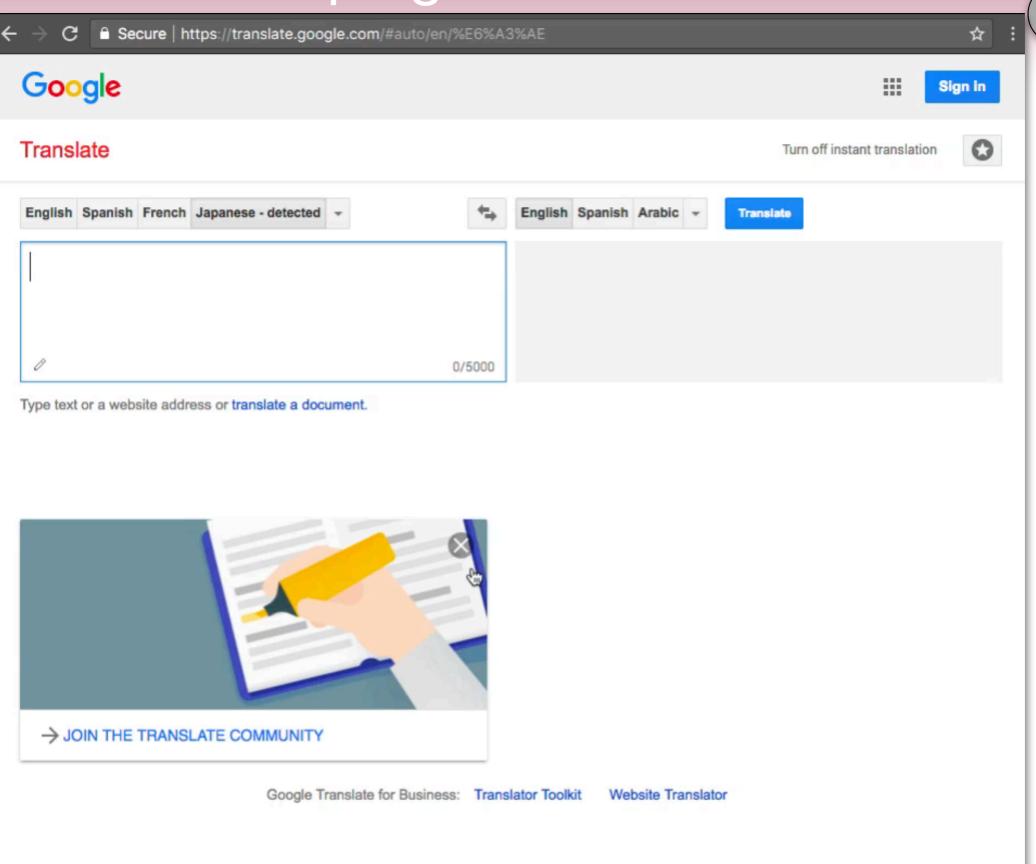
Rock or stone is a natural substance, a solid aggregate of one or more minerals or mineraloids. For example, granite, a common rock, is a combination of the minerals quartz, feldspar and biotite. The Earth's outer solid layer, the lith...

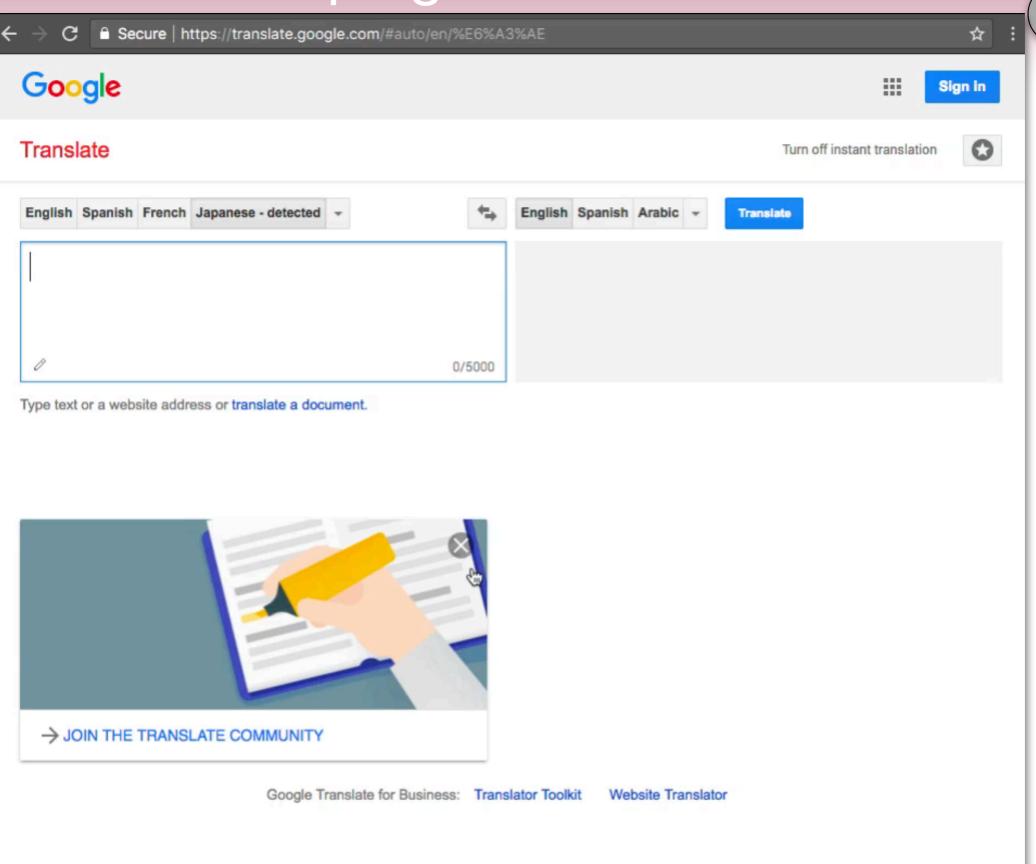




Query shortcuts

- Map: "santa barbara, ca 93108" vs the wrong zip code
- Calculation: "5+4"
- Flight Info: "american airlines 715"
- Stock price: "msft"
- Unit conversion: "1 dollar in euro"
- Music: "White Stripes"
- "Instant Answers" -> "animal noises"

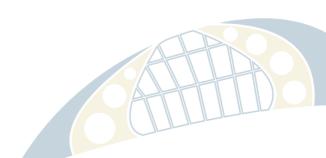




EXERCISE

WORK IN TEAMS

Try the queries on the next page in Google



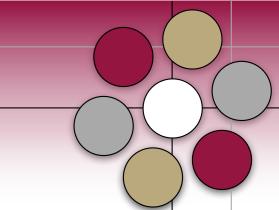
Query shortcuts

- roll a die
- flip a coin
- tip calculator
- recursion
- 88921384914 in english
- atari breakout
- google pacman
- do a barrel roll
- askew

- google gravity
- googler in 1998
- anagram
- zerg rush
- Bacon Number Matthew

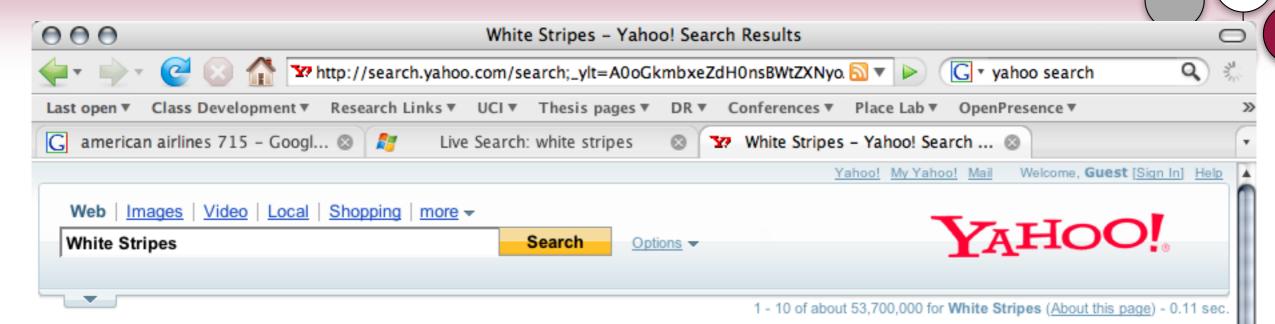
Broderick

			г			Т
f	'		'	1	,	
	?	?!	?!:	?!:г	?!: Г ј	3 i : L 1 :



Query shortcuts taken to the extreme

	the next solar e	,		
nput Inter	pretation:			
next	solar eclipse	after today		
annula				
		nursdav. January 14. 20)10 (7.344 days fr	om now)
11:06	pm PST Th	nursday, January 14, 20		om now) Orthographic projection





The White Stripes

Official site for The White Stripes, the Detroit-based garage/blues duo featuring Jack and Meg White.

www.whitestripes.com - 2k - Cached

White Stripes v2

Jack and Meg White fansite includes news and tour information, forum, links, release details, and more.

www.whitestripes.net - 72k - Cached

SPONSOR RESULTS

White Stripes Ringtones

Complimentary White Stripes Ringtones. Get Them Instantly. www.TheBombRingtones.com

White Stripes Ringtones

White Stripes Ringtones. Get Them Complimentary Now.
ChartRingers.com/whitestripes

The White Stripes

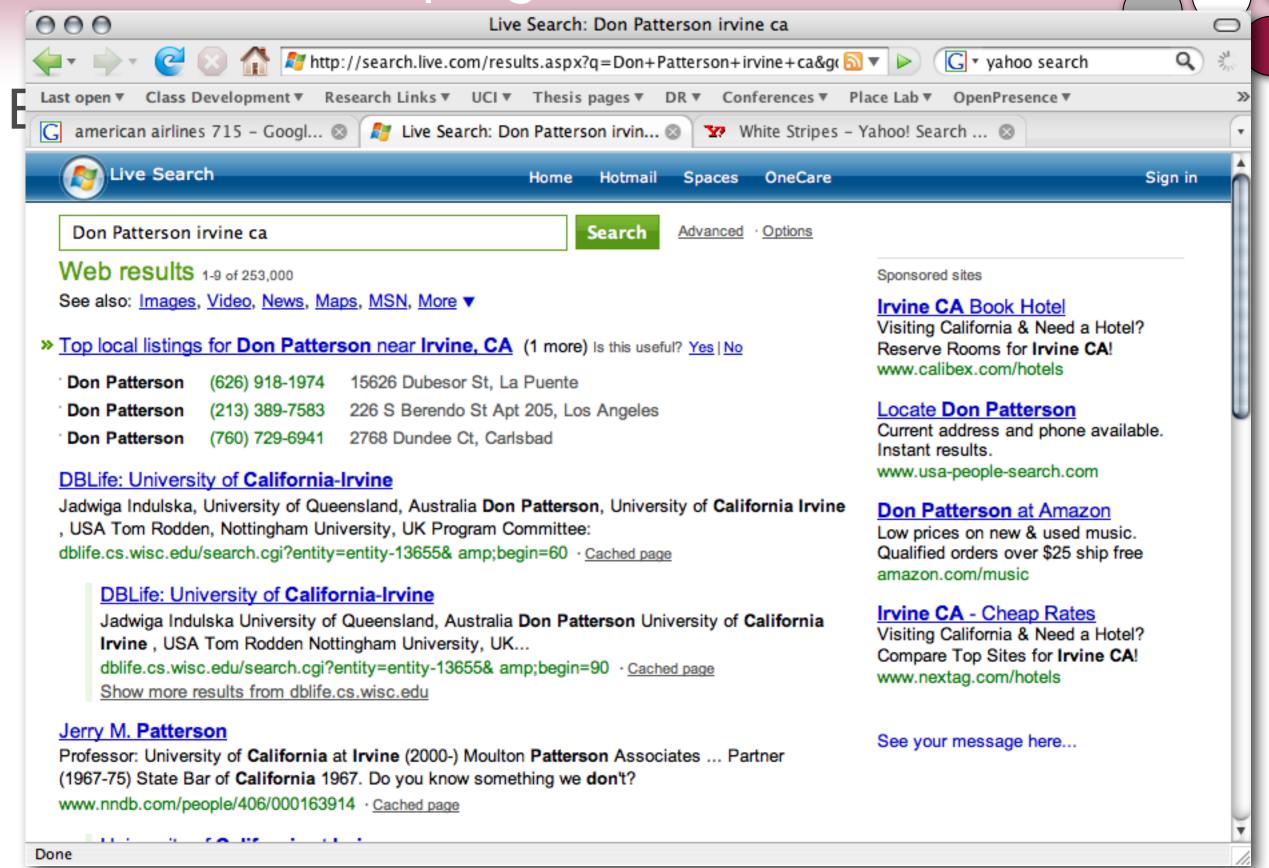
Find, compare & buy. Compare & Buy from 1000's of Stores. www.Dealtime.com

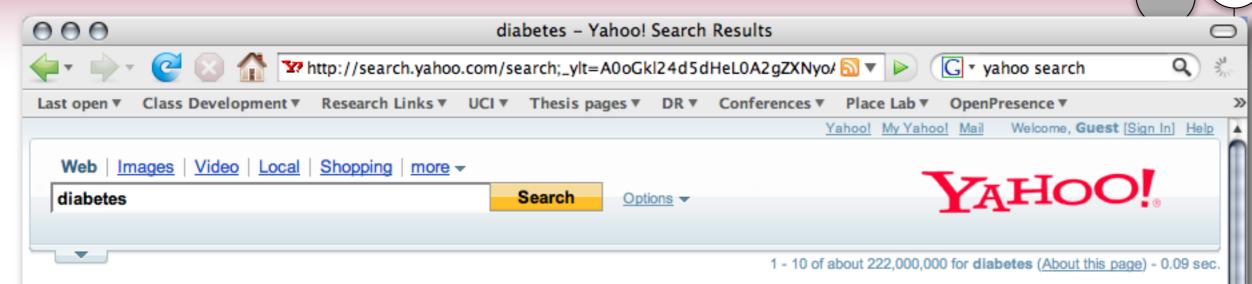
The White Stripes at

Amazon.com

Low prices on new & used music. Qualified orders over \$25 ship free. Amazon.com/music

See your message here...





SPONSOR RESULTS

Free Diabetes Supplies

www.LibertyMedical.com - Medicare Diabetics - Your supplies Could be 100% covered. Learn more.

Diabetes and Pain

www.Cymbalta.com - Diabetic peripheral neuropathic pain (DPNP) -learn about managing it.

Learn About Diabetes

www.Type2Diabetes-Info.com - Learn more about type 2 diabetes including treatment options.

Diabetes Information - Health Results

Articles Community Other Resources

Symptoms Yahoo! Groups National Institute of Diabetes &...

Treatments Ask a Question Medications Read Q & A MedlinePlus: Diabetes

Yahoo! Shortcut - About

CDC: Basics About Diabetes

Fact sheet about **diabetes** covering its common symptoms, risk factors, and management options. www.cdc.gov/diabetes/faq/basics.htm - 36k - <u>Cached</u>

American Diabetes Association

Provides extensive news and information about diabetes including symptoms, nutrition, weight

SPONSOR RESULTS

Diabetes Learning Center

Educational site explains how to manage type 1 and type 2 diabetes. www.bddiabetes.com

Manage Diabetes Naturally

Sugar Free Psyllium Fiber Products for Type 2 **Diabetes**. Order Online. www.konsyl.com

Diabetes

Get a Free Meter from Liberty if You have **Diabetes** & on Medicare. www.**Diabetic-Testing-Supplies.** com

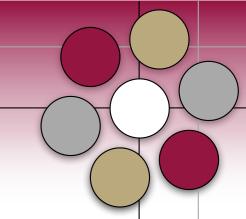
Diabetes Is Info

Get Info on **Diabetes** Is from 14 search engines in 1. www.info.com

Losing Income Due to Diabetes?

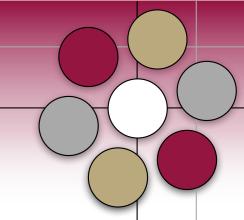
Unable to work due to Diabetes?

Done



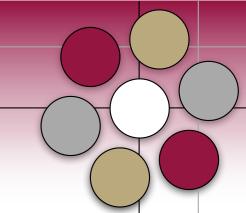
Context

- Context determination
 - Spatial
 - User Location
 - IP geocoding
 - cell phone localization
 - User input ("make this my home")
 - Web page location
 - Target Location
 - "SNA" maps to a latitude and longitude



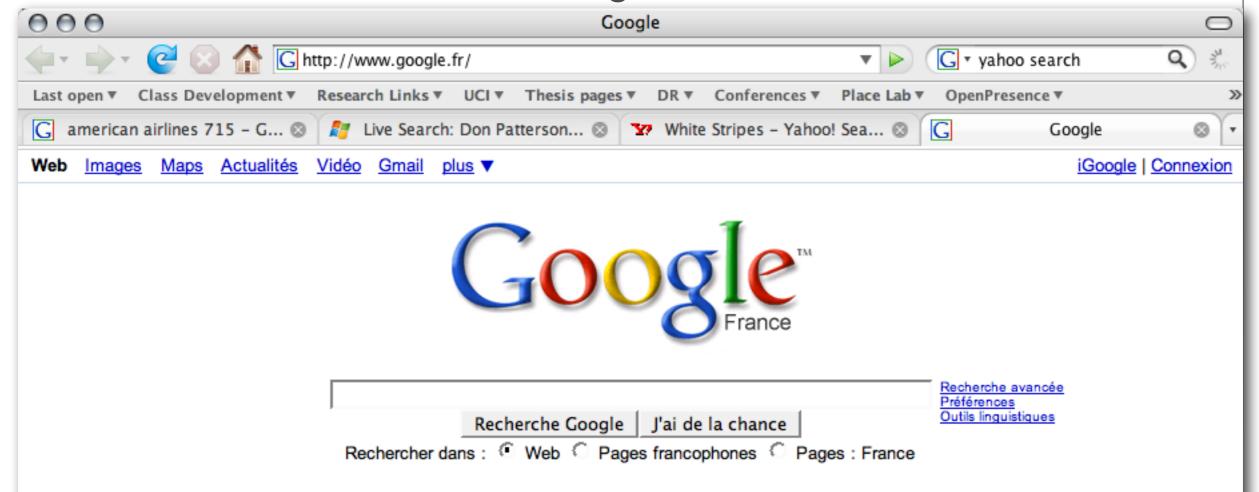
Context

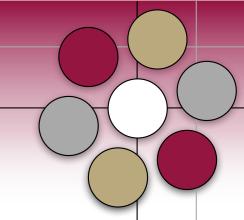
- Query Stream
 - Previous queries have influence
- Personal Search
 - Your profile and history have influence
- Explicit
 - You choose to use a vertical search engine



Context

- Implicit Context
 - google.fr is the default in france
 - User Location and Target Location





Context Use

- Restrict Results / Delete inappropriate results
- Rerank results based on context
- Change the content presentation
 - Language etc.
 - Specifying a context is built into some of the search APIs

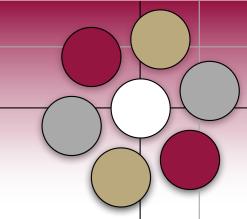
EXERCISE

WORK IN PAIRS

Search the Lumen database to find a site that has been removed as a result of a government removal order

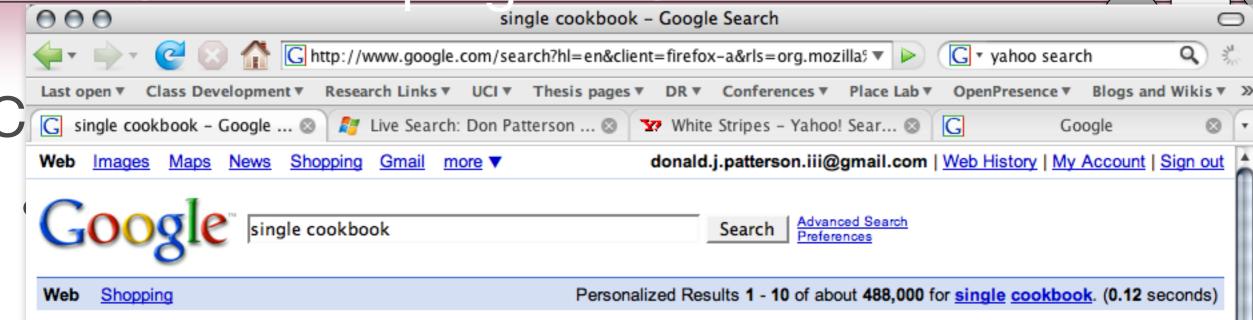
https://www.lumendatabase.org





Context Transfer

Move Search Context with the User



Product search results for single cookbook



Eating in: The Official Single Man's ...

Kristian Whitney Page: The Single Mother ...

Eating in: The Official Single Man's ...

\$6.90 - Desktop Music, LLC

\$24.52 - Eruditor.com

\$6.90 - A1Books

See single cookbook results available through Google Checkout

Salvatore's Single Fathers' Kitchen, recipe, cooking, kitchen ...

Salvatore's **Single** Fathers' Kitchen. A Free Online **Cook Book** designed specifically for **Single** Fathers. Get lots of recipies that are Fast, Fun, ... www.singlefatherskitchen.com/ - 22k - Cached - Similar pages - Note this

A 19th-century cookbook gives new twist to 'soul food ...

Nov 22, 2007 ... The evidence of a **single cookbook** is not enough to rewrite culinary history. Still, Russell's book suggested that a more nuanced view might ... www.iht.com/articles/2007/11/22/arts/**cookbook**.php - 68k - Cached - Similar pages - Note this

New Module: phpwsCookBook - phpWebSite Support Forums

I'm not completely satisfied with the layout of the pages for viewing the list of cookbooks, a single cookbook, or a section. But I'm not sure how I want to ... phpwsforums.com/showthread.php?t=130 - 56k - Cached - Similar pages - Note this

Amazon.com: B. Marold's review of Joy of Cooking: 75th Anniversary ...

IF I were limited to a **single cookbook**, my personal choice between these two is 'Joy of Cooking', simply because it's recipes are just as good as 'Better ... www.amazon.com/review/R199EBZOY1WQZM - 116k - Cached - Similar pages - Note this

Sponsored Links

Borders Book Stores

Get A Coupon For 20% Off Any Item Find A Borders **Book** Store Near You! www.VisitBorders.com

Serious Singles Only 30+

Tired of online dating games ?
Meet real, quality singles now!
www.SeriousSinglesOnly.com
San Francisco-Oakland-San Jose, CA

Cookbook

Get Recipes, Menus, Cooking Secrets & More at AOL® Food - Visit Now! Food.AOL.com

Single Cookbook at Amazon

Buy books at Amazon.com and save. Qualified orders over \$25 ship free Amazon.com/books

Design Your Own Cookbook

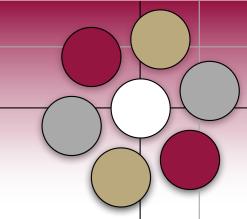
Create Your Own Cookbook & Send it to Family & Friends! www.Bookemon.com

Done



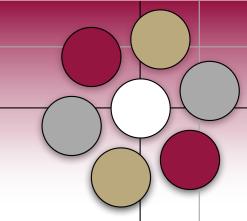
Eligible for FREE Super Saver Shipping.

Overview



- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam

Overview

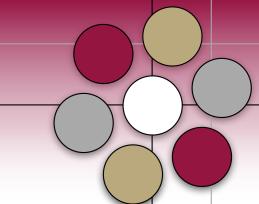


- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users



Spam

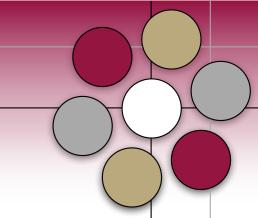
<u>Spam</u>



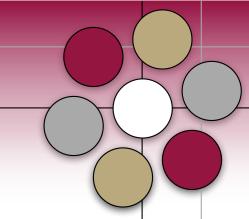
The trouble with paid placement (aka ads):

- It costs money... so instead
- Search Engine Optimization ("SEO")
 - define: "Tuning" your web page to rank highly in the search results for select queries
 - Alternative to paying for placement
 - It is marketing. Getting your content to your audience.

Spam

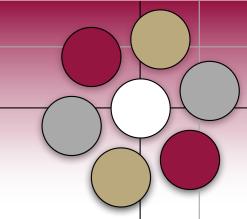


Search Engine Optimization

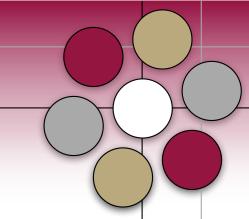


Search Engine Optimization

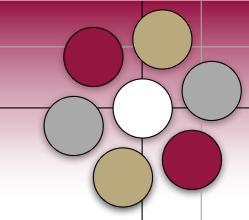
Motives



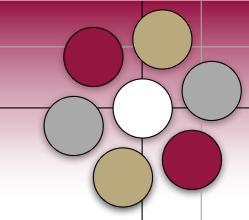
- Motives
 - Commercial



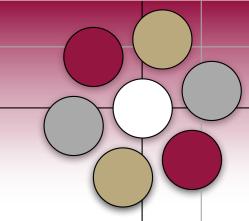
- Motives
 - Commercial
 - Political



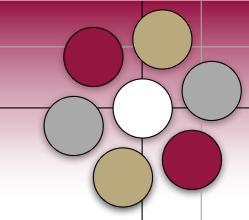
- Motives
 - Commercial
 - Political
 - Religious



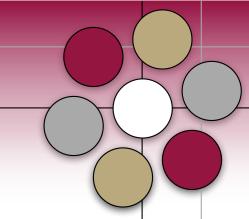
- Motives
 - Commercial
 - Political
 - Religious
 - Lobbying



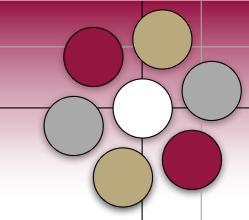
- Motives
 - Commercial
 - Political
 - Religious
 - Lobbying
- Who does this?



- Motives
 - Commercial
 - Political
 - Religious
 - Lobbying
- Who does this?
 - Internally: webmasters



- Motives
 - Commercial
 - Political
 - Religious
 - Lobbying
- Who does this?
 - Internally: webmasters
 - Commercially: companies, consultants



- Motives
 - Commercial
 - Political
 - Religious
 - Lobbying
- Who does this?
 - Internally: webmasters
 - Commercially: companies, consultants
 - Hosting services

Search Engine Optimization

- Learn more about how to do it online:
 - Web-Master World
 - http://www.webmasterworld.com
 - Search Engine Specific Tricks
 - Discussions about academic papers and results
 - Econtent creation vs link building
 Let's get the IncrediBills and the like to pitch in on this one

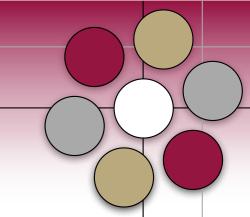
 Moved: 2 pages on site %55 similar duplicate content or not?

 Post was moved here: http://www.webmasterworld.com/google/3537427.htm

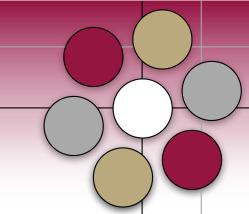
 Ideal length of content for SEO purposes?
 How long should an article be?

 Source of Competitor's Traffic from a Major.com?
 No apparent outlinks from the site to competitor's site

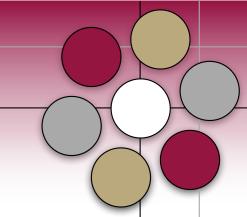
Moved, week ald website ranked #14 in goodle for main keyword



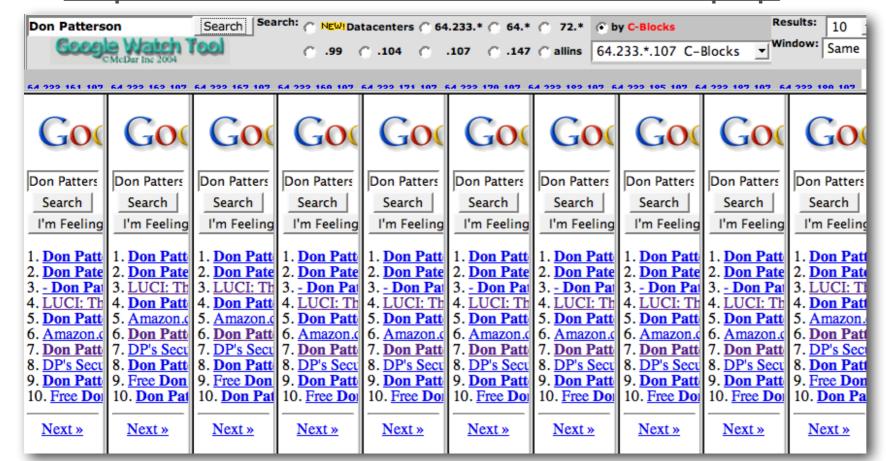
- There are ethical and inethical ways to approach SEO
- Legitimate approach is to:
 - create valuable content
 - make it widely accessible
 - clearly organize it
 - keep it up to date
 - use web standards
 - use web validation tools
 - get high visibility sites to link to your content

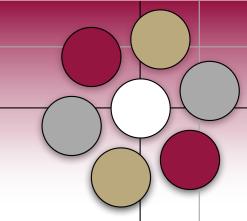


- Inethical approaches (aka spam):
 - lots of tricks
 - make lots of fake pages which point to your site
 - make lots of fake comments on sites which point to your site
 - In a nutshell, "lie"
- Sometimes legitimate and illegitimate techniques are hard to differentiate. It can be a fine line between them.



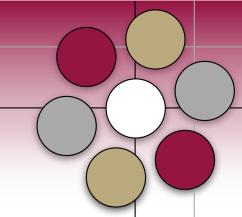
- Ranking depends on the data center
 - http://www.flickr.com/photos/the_impression_that_i_get/1321041609/
- Examine the different results:
 - http://www.mcdar.net/dance/index.php





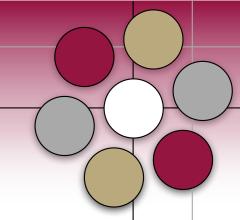
Keyword Stuffing

- First Generation Search Engines
 - Heavily relied on tf/idf ratio.
 - E.G. The highest ranking page for the query "brilliant computer scientist" had the most examples of those words.



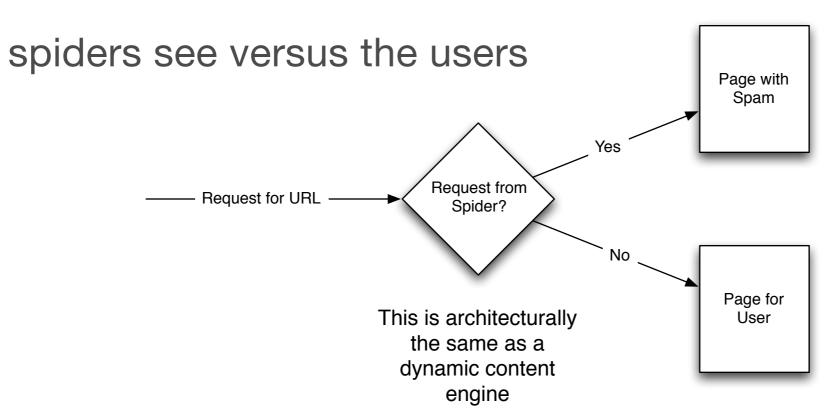
Keyword Stuffing

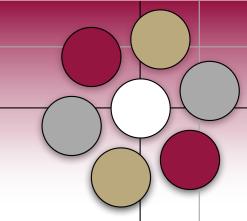
- So SEOs responded by screwing around with keywords
 - Misleading meta-tags
 - Repeating keywords over and over and over and....
 - Playing games with colors. (white on white keywords)
 - visible to spiders but not users in browsers



Keyword Stuffing

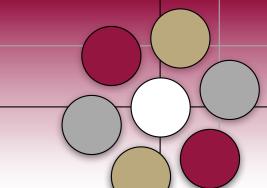
- Cloaking
 - define: Serving different content to a spider than to a user.
 - More sophisticated versions of differentiating what the





Other spam techniques

- Doorway pages
 - Like cloaking but using a redirect
 - Initial page is optimized for a keyword then a redirect takes the user to the "real" page
- Link spamming
 - Programs that search for blogs and automatically leave comments with links
- Robot Clicker-Fraud
 - Programs that "click" on query results to up their value.



Spam Inductor

Find out more

ORDER NOW

Get a first page listing on Google - GUARANTEED! For maximum search engine traffic - the best of SEO and search advertising. Visitors in just 48 hours from \$7/day. Discover the traffic potential!

> WARNING: This site contains sneaky, underhanded **Black Hat Seo tactics.**

> Black Hat Seo is responsible for more online fortunes than you'd care to imagine but it's NOT for everybody.

Make Money Blogging

See How I Earn Over Six Figures a year Blogging

I Will Get Your Website to the Top of

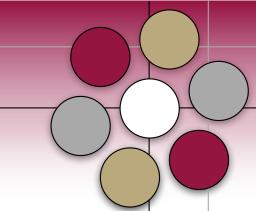
The art of search engine optimization...gaining top spots on Google...is no easy chore. I know...this is my I assist people in getting top positions for their websites on Google, Yahoo, MSN and all the other major

There are a few givens on the internet when it comes to trying to market goods and services:

No Traffic=No Sales!

End of story...that's it...bottom line!

If you have a make to



Spam Contest

Web Images Maps News Shopping Gmail more ▼

donald.j.patterson.iii@gmail.com | Web History | My Account | Sign out



nigritude ultramarine

Search

Advanced Search Preferences

Web

Personalized Results 1 - 10 of about 31,200 for nigritude ultramarine. (0.21 seconds)

Anil Dash: Nigritude Ultramarine

Do me a favor: Link to this post with the phrase **Nigritude Ultramarine**. I'd rather see a real blog win than any of the fake sites that show up on that ... www.dashes.com/anil/2004/06/nigritude-ultra.html - 155k - Cached - Similar pages - Note this

Nigritude Ultramarine FAQ - Jan 5

Nigritude Ultramarine FAQ - frequently asked questions about nigritude ultramarine and the realted SEO contest.

www.nigritudeultramarines.com/ - 57k - Cached - Similar pages - Note this

SEO contest - Wikipedia, the free encyclopedia

In the English-language world, the **nigritude ultramarine** competition by SearchGuild is widely acclaimed as the mother of all SEO contests [citation needed]. ... en.wikipedia.org/wiki/**Nigritude_ultramarine** - 35k - <u>Cached</u> - <u>Similar pages</u> - <u>Note this</u>

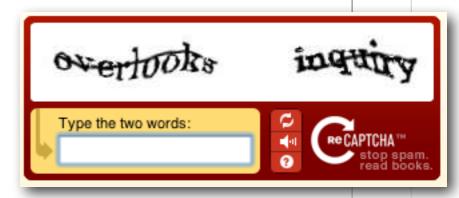
Slashdot | How To Get Googled, By Hook Or By Crook

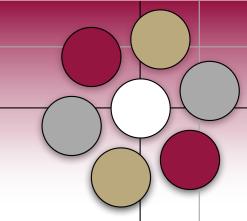
The current 3rd result showcases the "Nigritude Ultramarine Fighting Force" When discussing nigritude ultramarine [slashdot.org] it is important to ... slashdot.org/article.pl?sid=04/05/09/1840217 - 136k - Cached - Similar pages - Note this

nigritude ultramarine - Sriram's WebLog on ASP.NET VB.NET C#

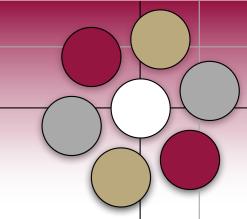
nigritude ultramarine. Wondering what it is? SEO Challenge held a contest for webmasters and site owners to come up with any search optimization technique ... weblogs.asp.net/sonyram/archive/2004/06/08/151375.aspx - 25k - Cached - Similar pages - Note this

- Quality Indicators
 - Statistical Analysis of Links (aka PageRank)
 - votes from authors
 - Usage indicators (users visiting a page)
 - votes from users
- Anti-Robot techniques
 - "Captchas"
 - Completely Automated Public Turing Test to Tell
 Computers and Humans Apart

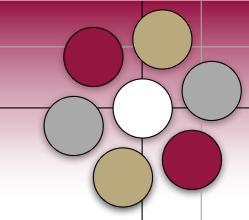




- Limits on meta keywords
- Spam Recognition by machine learning
- "no-follow" attribute
- Family Friendly filters
 - Automatic Detection of Pornography
 - Often the spammers desired landing page
 - Text Analysis
 - Look for keywords and variants



- Robust Link Analysis
 - Ignore statistically improbable links
 - Use link analysis to detect spammers
 - "Guilt by association"

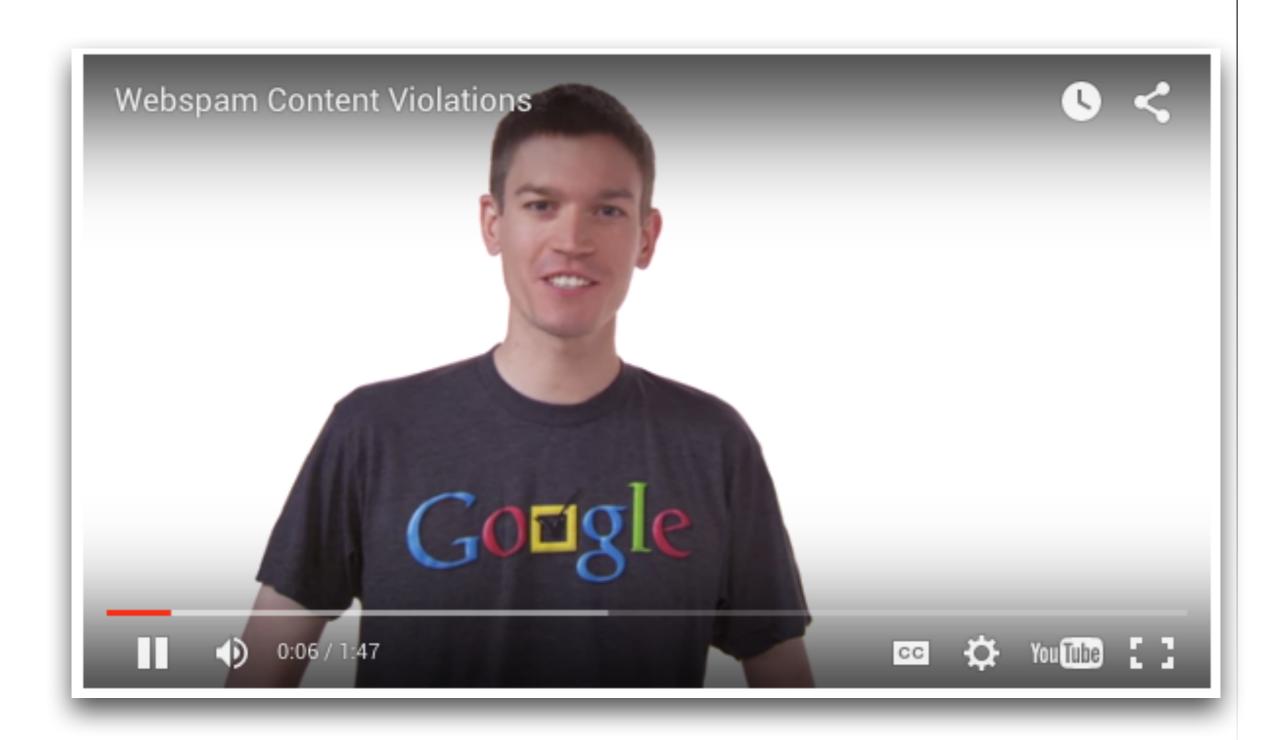


- Editorial Intervention
 - Blacklists
 - Query Reviews
 - Customer Complaints
 - Visualization Tools

Webmaster Guidelines

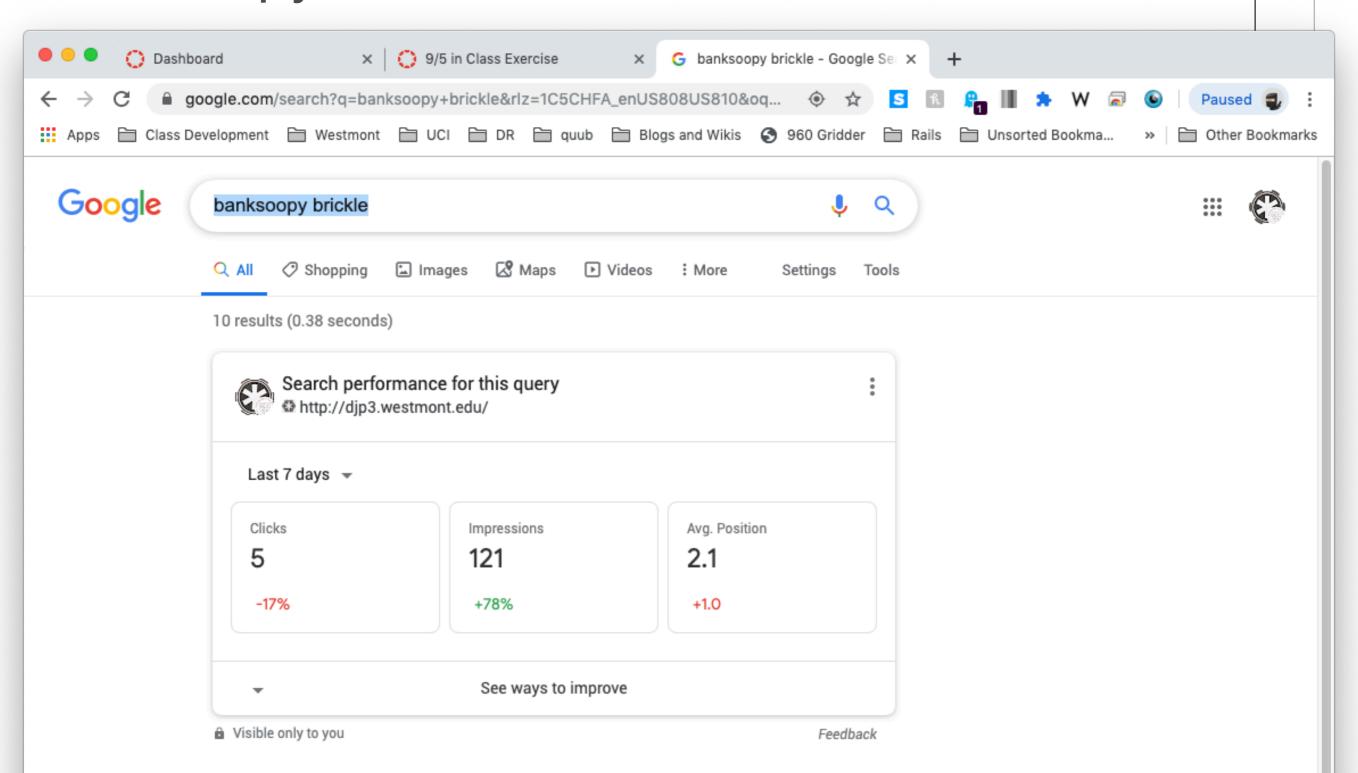
- Search Engines have SEO policies
 - What is allowed and not allowed
- Example: Search for "google webmaster guidelines" or "msn guidelines for successful indexing"
- Ignore them at your own risk
- Once you are blacklisted by a search engine you will disappear from the web
 - Remember how search engines enable scalability?
- Adversarial IR Research:
 - http://airweb.cse.lehigh.edu/

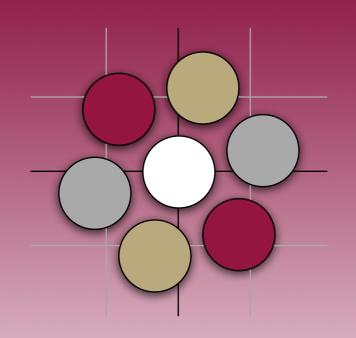
Webmaster Guidelines



Spam - Extra Credit

banksoopy brickle





WESTMONT COMPUTER SCIENCE