

Towards the New Sciences of the Internet: Next Generation Search and the Power of Social Networks

Usama Fayyad, Ph.D.
Chief Data Officer & Executive VP
Yahoo! Inc.





Thanks to the following people at **Yahoo! Research** for some of material I drew on in this talk:

- SDS team at Yahoo!
- Yahoo! Research people:
 - Andrew Tomkins, Chief Scientist, Search
 - Prabhakar Raghavan, Head of Yahoo! Research
 - Raghu Ramakrishnan, Chief Scientist, Audience
 - Ravi Kumar, Manager of Search Team, US
 - Ricardo Baeza-Yates, Director of Y!R Barcelona and Santiago



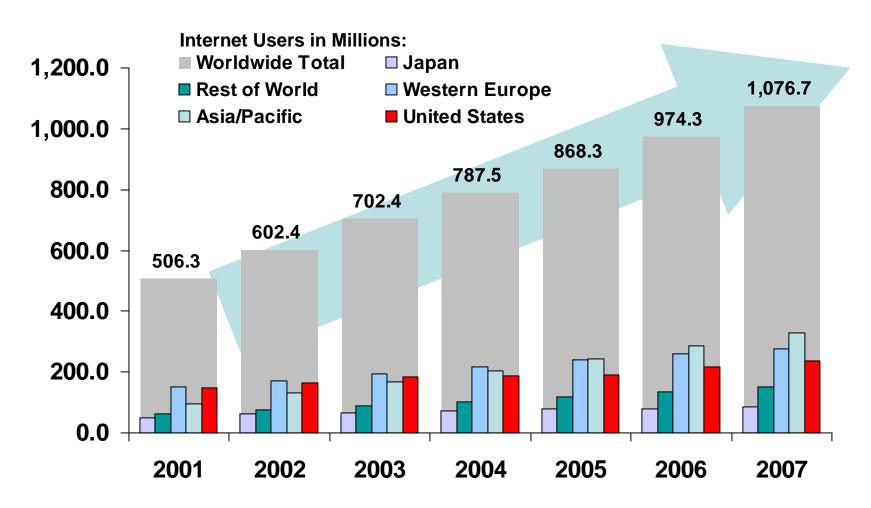


- About Yahoo! and its business
- Search and Search Advertising
- Targeting to Users for Relevance
- The Evolution of "Social Media"
- Understanding the new Phenomena of the Internet and Social Media
- Yahoo! Research
- Overview of Some Research Challenges
- The Evolution of Search
- Concluding Thoughts



Globally, I

Globally, Internet Users Will Number Over <u>1 Billion</u> by 2007









73% of the U.S. Internet population uses Yahoo! – Over 500 million users per month globally!

- Global network of content, commerce, media, search and access products
- 100+ properties including mail, TV, news, shopping, finance, autos, travel, games, movies, health, etc.
- 25 terabytes of data collected each day... and growing
 - Representing thousands of cataloged consumer behaviors

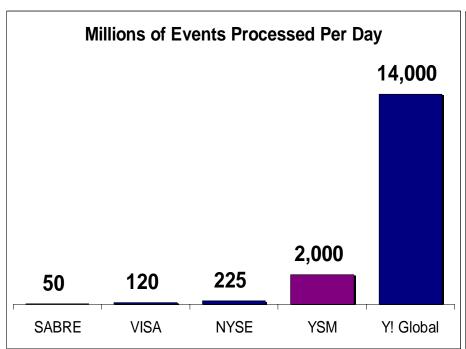
More people visited Yahoo! in the past month than:

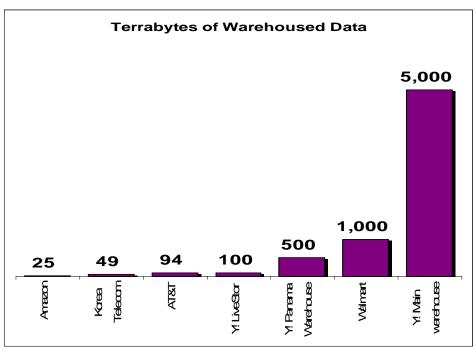
- Use coupons
- Vote
- Recycle
- Exercise regularly
- Have children living at home
- Wear sunscreen regularly

Data is used to develop content, consumer, category and campaign insights for our key content partners and large advertisers



Yahoo! Data - A league of its own...





GRAND CHALLENGE PROBLEMS OF DATA PROCESSING

TRAVEL, CREDIT CARD PROCESSING, STOCK EXCHANGE, RETAIL, INTERNET

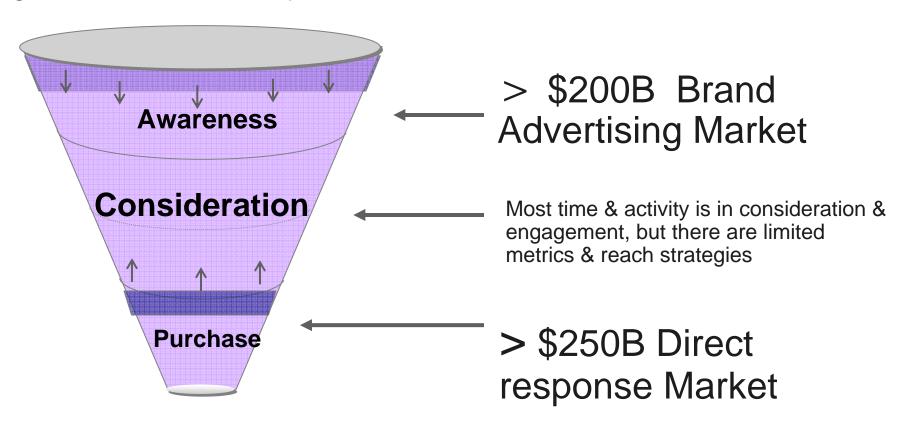
Y! Data Challenge Exceeds others by 2 orders of magnitude





Advertising: Brand and DR

Knowledge of users & their behavior throughout the purchase funnel can grow brand & direct response revenue







A Tale of Two Search Engines



Home Hotmail

=Audience

Spaces

OneCare

Sign

latin canon

Search

Advertisements Advanced · Options

=Monetization

Sponsored sites

Web results 1-10 of 13,300,000

See also: Images, Video, News, Maps, MSN, More Algorithmic results

CANON LATIN AMERICA, INC.

canon latin america, inc. ...

www.canonlatinamerica.com · Cached page · Translate this page

Canon law (Catholic Church) - Wikipedia, the free encyclopedia

This code was promulgated on 27 May 1917 as the Code of Canon Law (

Latin: Codex Iuris Canonici) by his successor, Pope Benedict XV, who set 19 May 1918 as the date on which it came into ...

en.wikipedia.org/wiki/Canon law %28Catholic Church%29 · Cached page

Canon of the Mass - Wikipedia, the free encyclopedia

This article incorporates text from the public-domain Catholic Encyclopedia of 1913. Canon of the Mass (Latin : Canon Missæ , Canon Actionis) is the name used in the Roman

en.wikipedia.org/wiki/Canon of the Mass · Cach Show more results from en.wikipedia.org

About Canon: Corporate Overview > Office Locations

Canon Latin America, Inc. 703 Waterford Way, Suite 400 Miami, FL 33126

Phone: 305-260-7400

www.usa.canon.com/templatedata/AboutCanon/ciwofloc.html · Cached page

Muratorian Canon in Latin

skesearch

Canon Is at Amazon

Low prices on Canon Is. Qualified orders over \$25 ship free amazon.com

Want a Free DVD Camcorder

Get great new products for Free. Join our member. incentive program. ProductTestPanel.com

Buy Canon Products

Shop at Staples.com and Save on Canon Products & Supplies Today! www.Staples.com

Best Camera Deals Online

Incredible prices! Up to 80% off on select items www.topdeallsonweb.com

Canon

Name Brand Photo products FIND, REACH, KNOW,

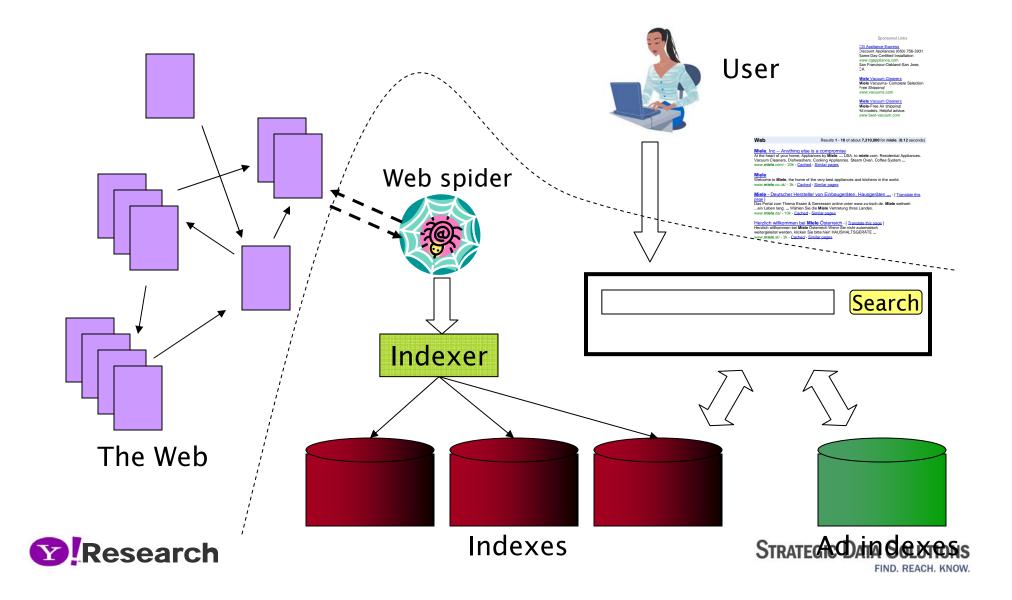


Algorithmic vs. Ad Search

- Analogous to classical separation of editorial vs commercial content
- Technical underpinnings:
 - Some commonalities (IR, ML)
 - Many differences (incentives, spam, mechanism design)
- Will cover both algo and ads



The two engines





1995: The Yahoo! Directory

- Apply human expertise and editorial to organize web sites
- What worked
 - Practical, Navigable
 - Trustworthy, Authoritative
- What didn't
 - Scalability
 - Granularity
 - Etc.







1995 : Altavista (Inktomi, Lycos, etc.)

- Automate the process of acquiring pages; use "information retrieval" techniques to return pages that contain a particular term
- What worked
 - Scalable (query for "IBM" returns 40M pages)
 - Simple
 - Granular
- What didn't
 - Scalability a double-edged sword
 - Ranking and relevance poor
 - Not authoritative (spam, irrelevance, etc.)







c. 1999-2006: PageRank (Google, Yahoo)

- Use topology (link structure) of the web to confer authority
- What works



- Relevance is greatly improved
- Navigational query is born (query for "IBM" gets me to ibm.com)
- What doesn't
 - Homogeneity of results (no personalization) means no "subjective" queries –
 webmasters vote by proxy for everyone and their answer is the only answer
 - System easily "gamed" by spammers leads to arms race



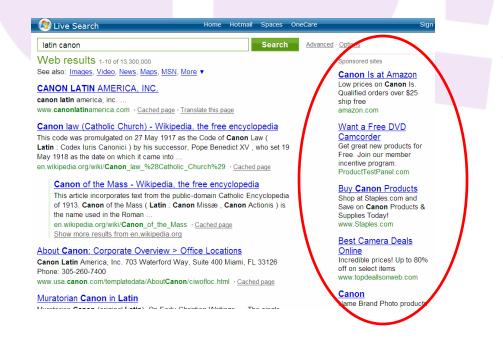


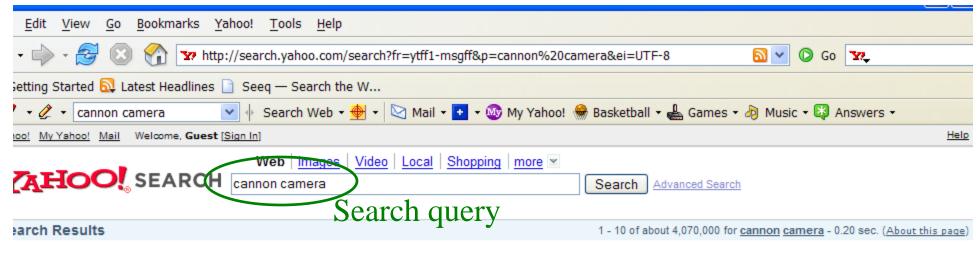
Meanwhile, On the Money Front...

- Sponsored search ranking: Goto.com (morphed into Overture.com → Yahoo!)
 - Your search ranking depended on how much you paid
 - Auction for keywords: <u>casino</u> was expensive!
- 1998+: Link-based ranking pioneered by Google
 - Blew away all early engines except Inktomi
 - Great user experience in search of a business model
 - Meanwhile Goto/Overture's annual revenues were nearing \$1 billion
- Result: Google added sponsored search "ads" to the side, independent of search results
 - 2003: Yahoo follows suit, acquiring Overture (for paid placement) and Inktomi (for search)
- The Monetization Mechanisms... Conversion of marketplace machanisms in 2007



Search advertising





Did you mean: canon camera

Canon Camera at Circuit City

www.CircuitCity.com - Circuit City - Official Site. Free Shipping on Orders \$24 and Up.

Cannon Camera

RitzCamera.com - Huge Selection of Canon Cameras. Free Shipping & No Tax. Buy Today.

Canon (NYSE: CAJ)

Global manufacturer of copy machines, fax machines, cameras, computer peripherals, and optical products.

www.canon.com - 23k - Cached - More from this site

2. Canon Camera Museum

Showcasing camera history, technology, and design. www.canon.com/camera-museum - 22k - <u>Cached</u> - <u>More from this site</u>

Canon Digital Cameras

Official Canon site for its line of PowerShot and EOS digital cameras, photo printers, and film scanners.

www.powershot.com - 104k - Cached - More from this site

Canon USA

Manufacturer of professional and consumer imaging equipment and information systems including copiers, printers, image filing systems, cameras and lenses, and more.

SPONSOR RESULTS

Authorized Canon Cameras Pro Dealer

SPONSOR RESULTS

Buy Canon **Cameras** here. Imageologists: Professional photographic...

www.imageologists.com

Canon Cameras

We Offer 3,500+ Digital Cameras. Discover canon cameras. www.BizRate.com/canon

Camera Cases and Bags

To know Bogen Imaging Inc, just take a look at the premium brands.

Cannon Camera Battery

Accessory 📜

Spring Sale. 80% off. Valid till Apr-30. Free Ship coupon over \$30. www.cellphoneshop.net

Ad(

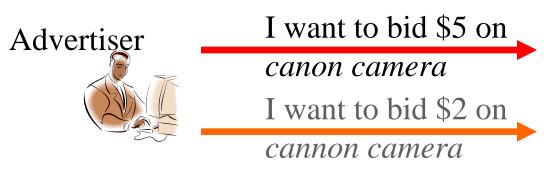


A Question for the audience

- Do you think an "average" user, knows the difference between sponsored search links and algorithmic search results
- Do you think an "average" user knows there are sponsored links on the page?
- Do you think a user knows where a sponsored link would navigate to upon a click?



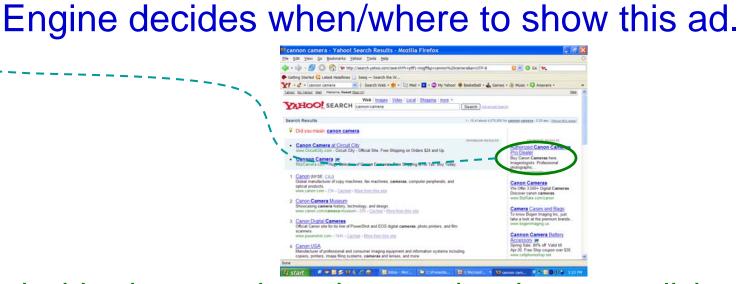




Ad Index
Sponsored search engine



Landing page



Engine decides how much to charge advertiser on a click.





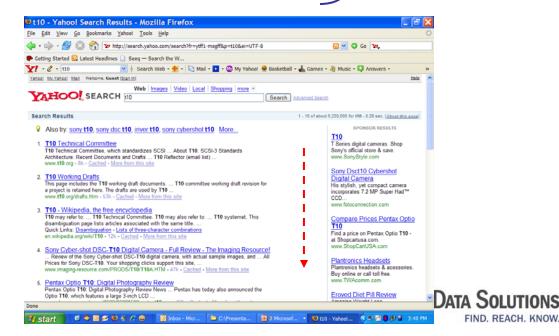
- When a searcher clicks on a displayed ad, advertiser pays the search engine
 - E.g., a Ford dealership may place an ad for the search
 Ford
 - Out of all such ads, engine chooses to display some on each *Ford* search
 - When a searcher clicks an ad, the advertiser who placed it pays





Engine: Three sub-problems

- 1. Retrieve ads matching query
- 2. Order the ads
- 3. Pricing on a click-through







1. Retrieve ads matching query

- Advertisers bid on keywords/phrases
- Match query to "nearby" bidded keywords
 - Retrieve ads for these bidded keywords
- Not a standard IR problem
 - Bidded keywords skewed to commerce
 - Advertisers and engine (trying to) co-operate
 - Marketplace of keywords
- Venezia hotels





- Standard tricks from IR: phrasing, spell-correction, ranking ...
 - Use both ad snippet and landing page features
 - Advertisers are trying to help! (Antispam)



SPONSOR RESULTS

250 **Hotels** in Venice - Italy

Save up to 70% on 1 to 5 star **hotels** in Venice. With guest-reviews.

www.booking.com

Hotel Venezia

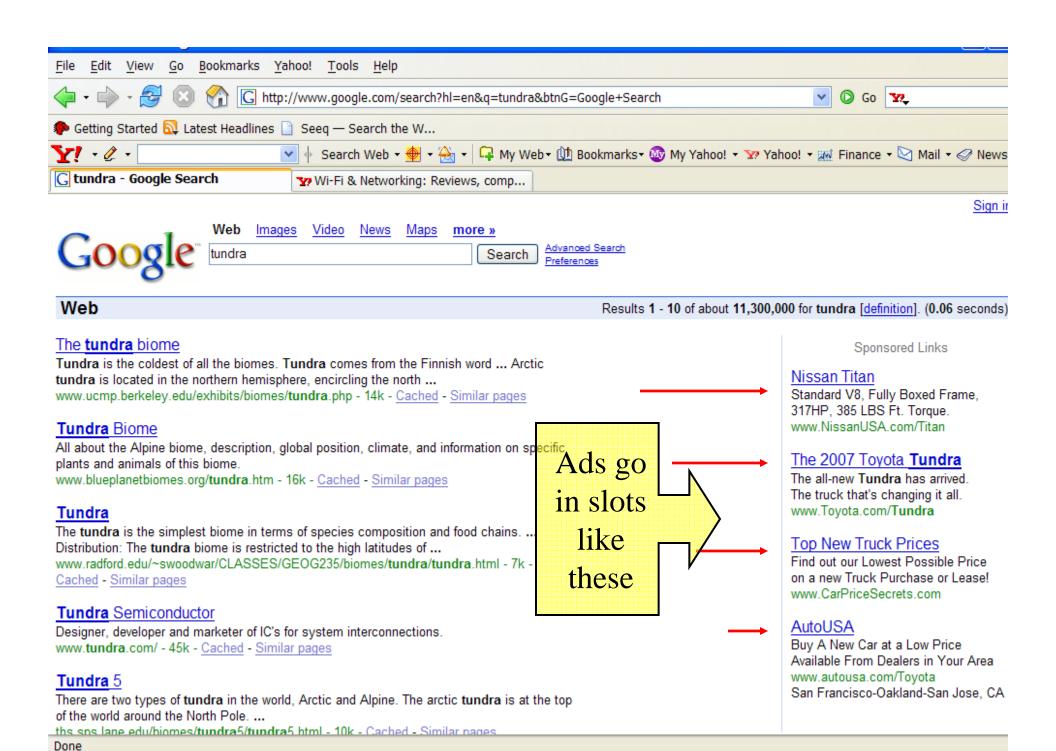
Compare **hotel** prices from over 120 top websites and save up to 70%.

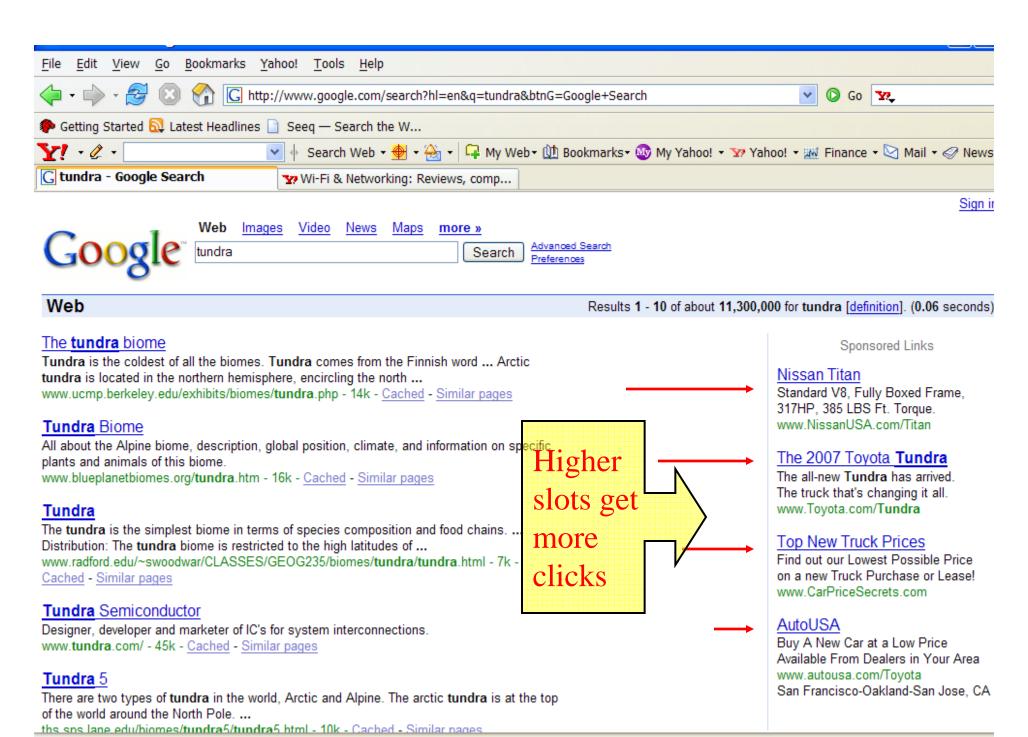
Hotels.SideStep.com

Charming Hotels & Resorts

Venice **Hotel** So Unique, So Very Distinctive! www.**TheCharmingHotels.** com





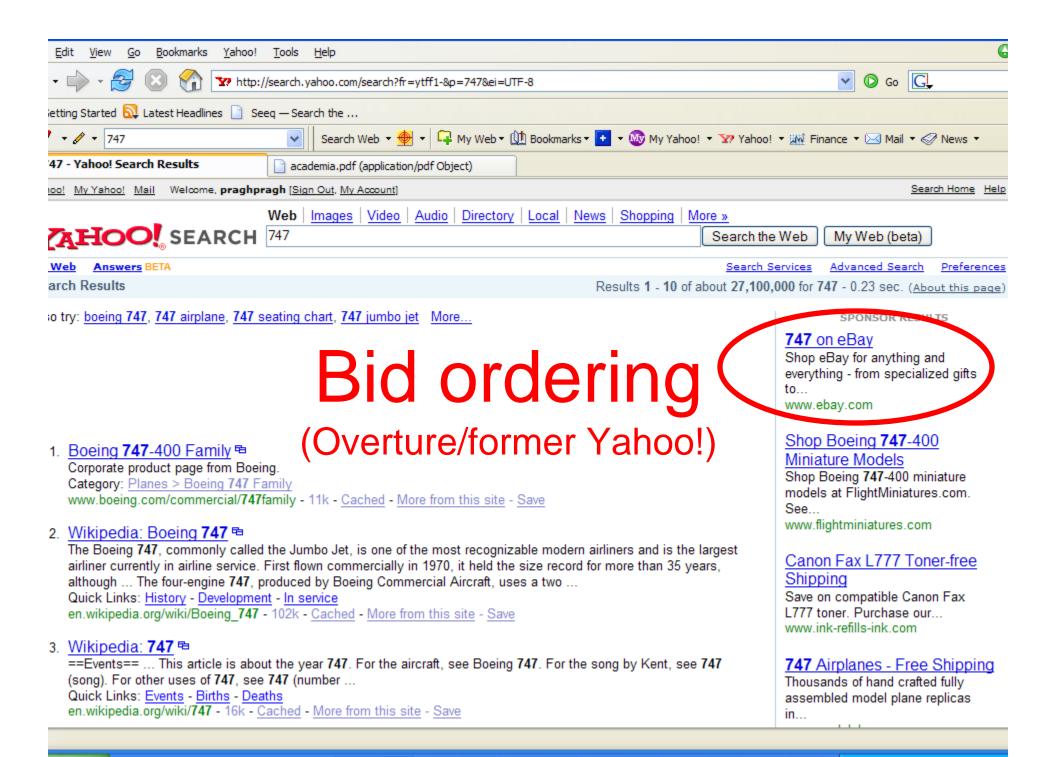


Done



- Most generally, composite IR+Econ score ... for today's talk, focus on Econ
- Original GoTo/Overture scheme:
 - Order by bid



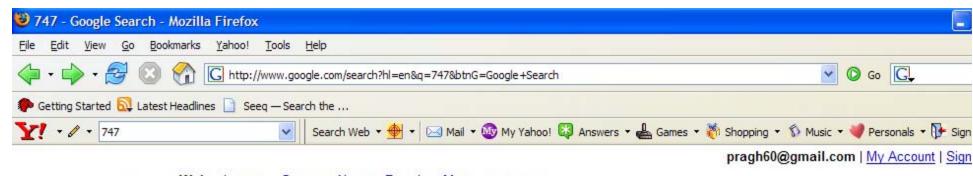




Revenue ordering, simplified

- Ad j's bid is denoted b_j
- Then, expected revenue from this ad is R_j
 = b_j x CTR_{jq}
- Order ads by R_j







 Web
 Images
 Groups
 News
 Froogle
 Maps
 more »

 747
 Search
 Advanced Search Preferences

Revenue ordering?

Web

Results 1 - 10 of about 83,900,000 for 747. (0.10 secon

Boeing: Commercial Airplanes - 747 Home

Boeing Announces First Engineering Releases for New 747-8 Family - Boeing Delivers Malaysia Airlines System's First 747-400 Freighter - Cargolux Expands ... www.boeing.com/commercial/747family/ - 12k - May 3, 2006 - Cached - Similar pages

Boeing: Commercial Airplanes - 747 - About the 747 Family

The Boeing 747 Family -- The Right Choice for the Large Airplane Market. About the 747 Family. Instantly recognized by passengers around the world, ... www.boeing.com/commercial/747family/background.html - 26k - Cached - Similar pages [More results from www.boeing.com]

Boeing 747 - Wikipedia, the free encyclopedia

The Boeing 747, commonly called the Jumbo Jet, is one of the most ... The 747SP was the longest-range airliner available until the 747-400 entered service ... en.wikipedia.org/wiki/Boeing 747 - 101k - May 3, 2006 - Cached - Similar pages

Airliners.net: Boeing 747-400

The 747-400 is the latest, longest ranging and best selling model of the 747 family. Boeing launched the 747-400 in October 1985 and the first development ... www.airliners.net/info/stats.main?id=100 - 27k - May 3, 2006 - Cached - Similar pages

Boeing 747SP Website

Boeing 747SP Production List, Photo Gallery, Incident Reports, Current Status, News and much more.

www.747sp.com/ - 21k - May 3, 2006 - Cached - Similar pages

Sponsored Links

747 On Sale

Chairs, Desks & Cabinets Everyday Low Prices. OfficeFurniture.BettyMills.com

Boeing™ **747**-400 Model

Accurately detailed Boeing™ 747-40 desktop display models www.FlightMiniatures.com

Boeing's Photo Collection

The world's leading source of still and motion aerospace imagery! www.boeingphotostore.com

747

Everything to do with 747 items. eBay.com



- Bid and revenue ordering: two forms of ordering by an econ score
- Does revenue ordering maximize revenue?
- No advertisers react to ordering scheme, by changing their bid behavior!
- Lahaie+Pennock ACM EC 2007
 - Family of schemes bridging Bid and Revenue ordering
 - Game-theoretic analysis







- Monetization and economic value an intrinsic part of system design
 - Not an afterthought
 - Mistakes are costly!
- Computing meets humanities like never before sociology, economics, anthropology ...



A question for the Audience:

Why is search-related advertising so powerful?



Classical web search user needs

Informational – want to learn about something (~25%)

Low hemoglobin

- Navigational want to go to that page (~40%)
 United Airlines
- Transactional want to do something (web-mediated) (~35%)
 - Access a service

Mendocino weather

Downloads

Mars surface images

Shop

Nikon CoolPix

- Gray areas
 - Find a good hub

Car rental Finland

Exploratory search "see what's there"

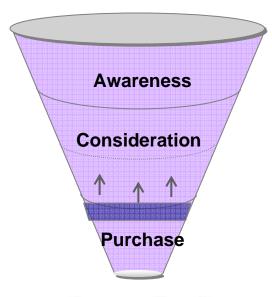




Moving Customers up the Funnel

Impulse Banners

- Target users based on their activity both search and property -- within the NEXT HOUR
 - Behavioral Categories Apparel, Computers, Home Appliances all the same categories that you can use for regular behavioral targeting!

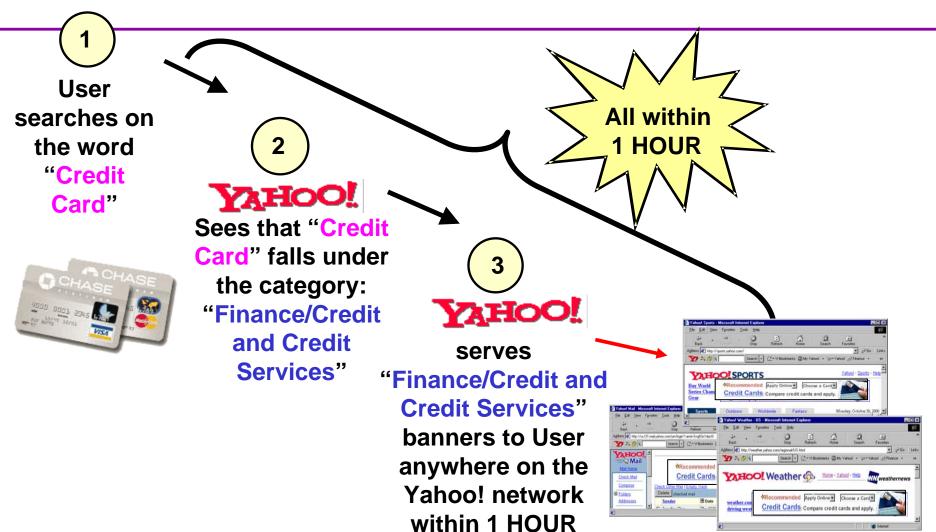








Impulse Banner Example





25% - 261% higher CTR

STRATEGIC DATA SOLUTIONS FIND. REACH, KNOW.



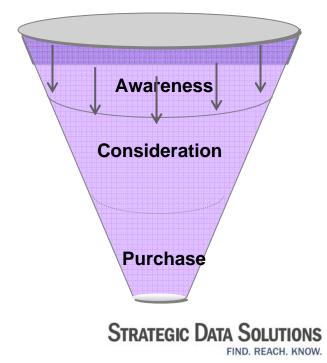
- Searches are not at all associated or tracked through personally identifiable information
- No long-term memory of search terms, all stored on client cookie.
- We generalize the category is targeting is at generic category: e.g. Financial Services, not "credit card"
- All targeting done in anonymous mode





 New generation marketing solutions to take brand advertisers down the marketing funnel

Behavioral Targeting (BT)

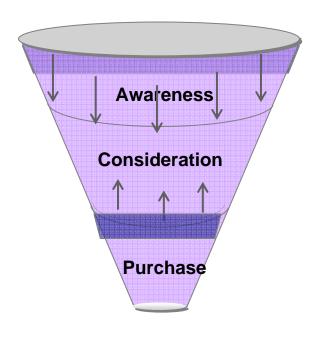






Brand Ads and Search Ads Interact!

- Is ad search strategy enough for a direct marketer?
- Do brand ads play a role in search advertising?
- Harris Direct Case Study









Case Study: Harris Direct

Viewing These Ads:



On: YAHOO! MAIL

MYYAHOO!

Had This Effect On:

- Aided Brand Awareness
 - Up 7%



- Brand Favorability
 - Up 32%



- Purchase Intent
 - Up 15%

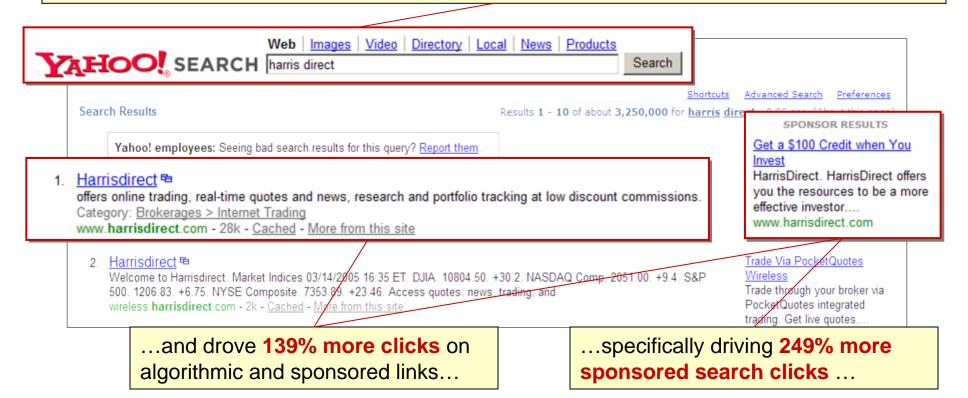






Case Study: Harris Direct

People who saw display ads were 61% more likely to search on related topics...





...and driving **91% more activity** on the HarrisDirect.com website.





Social Media and Yahoo! Research

Example of challenges of inventing the new sciences of the Internet





- Although the "traditional notion" of portal and web content is still attracting growing audiences
- The original notion of "publishing content" to attract audiences is changing fast
 - As people discover the fact that the Internet is an Interactive Medium
 - The uses of the Internet enter areas we could not imagine a short time ago
- A new notion of "publishing" is fast emerging
 - The opportunity of user-generated content

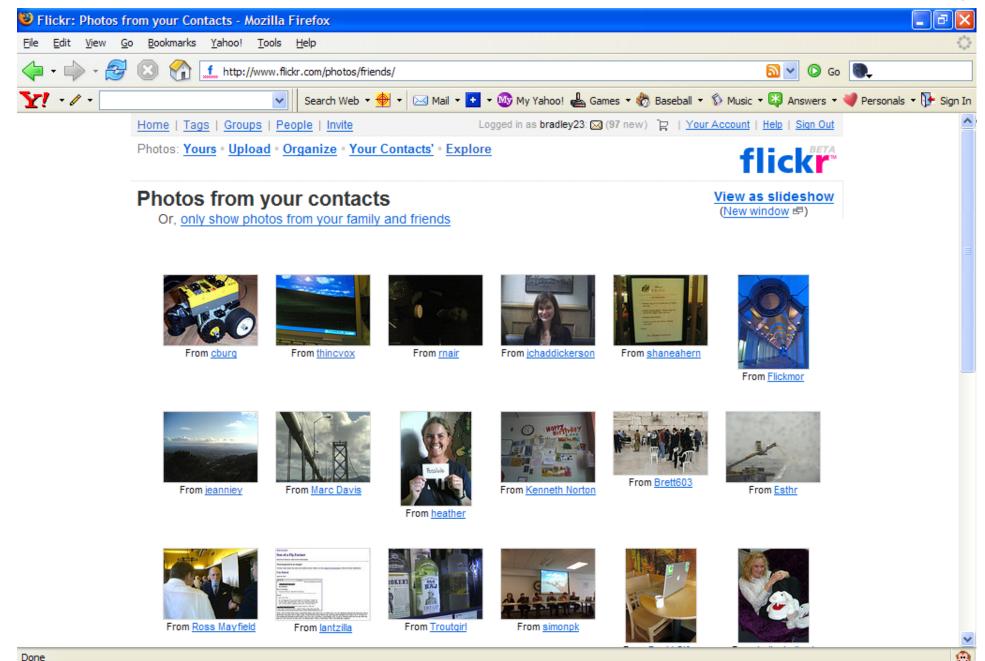




Examples of Social Media

- Blogs
 - The individual as publisher
 - Comments and tags part of the process
- Sharing Photos: e.g. Flickr
- Social Search
 - My Web 2.0
 - Yahoo! Answers
 - Del.icio.us
- Web communities:
 - Yahoo! Groups
 - Individual web presence: Facebook, MySpace, Yahoo! 360, Friendster, ...
- Video sharing: You Tube, Yahoo! Video, etc...









1. User Generated Content

Content not licensed from providers such as Corbis or Getty, but rather contributed by users.



sometimes it snows in april From Joü





Africa Masai boy From housden photos



slippers From benjaminhamilton





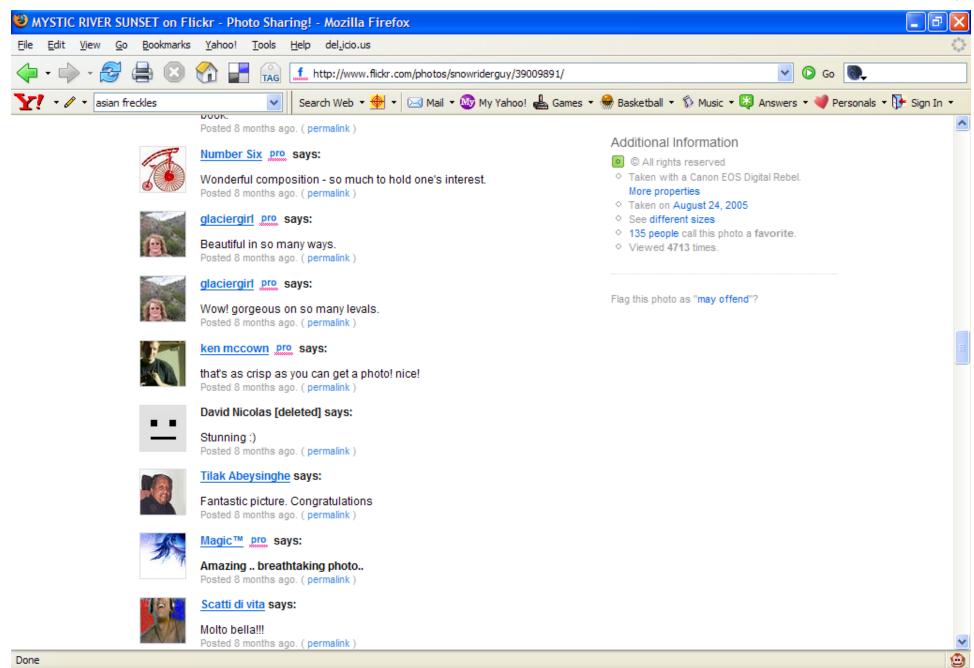




2. User Organized Content

Content is tagged, described, organized, discovered, etc. not by "editors" but by the users themselves.







3. User Distributed Content

Flickr achieved distribution across the internet, not through "business deals" per se, but rather through the Flickr community which distributed Flickr content on 3rd-party blogs.

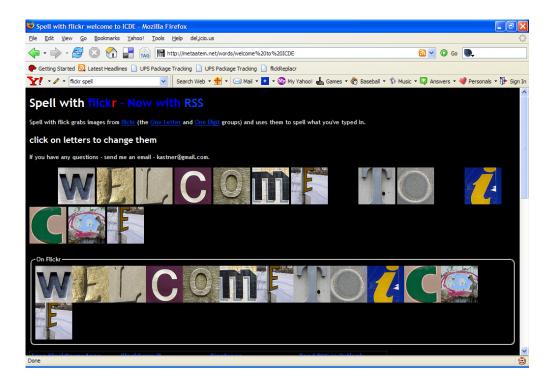






4. User Developed Functionality

Flickr exposed APIs (PHP, Perl, etc.) that allowed the community of developers to build against the Flickr platform.







1. User Generated Content

Content not licensed from providers such as Corbis or Getty, but rather contributed by users.

2. User Organized Content

Content is tagged, described, organized, discovered, etc. not by "editors" but by the users themselves.

3. User Distributed Content

Flickr achieved distribution across the internet, not through "business deals" per se, but rather through the Flickr community which distributed Flickr content on 3rd-party blogs.

4. User Developed Functionality

Flickr exposed APIs (PHP, Perl, etc.) that allowed the community of developers to build against the Flickr platform.

Entire ecosystem created by less than ten employees... aided by millions in the Flickr community.





Yahoo! Research

Inventing the new sciences of the Internet





- The Internet touches all of our lives: personal, commercial, corporate, educational, government, etc...
- Yet many of the basic notions we talk about:
 - Search, Community, Personalization,
 Engagement, Interactive Content, Information
 Navigation, Computational Advertising
 - Are not at all understood, or well-defined
 - These are not disciplines that academia or any industry research labs focus on...





Community:

- How do you know what to believe on the Internet?
- Trust models on-line and trust propagation
- What makes communities thrive? Whither?
- Social media, tagging, image and video sharing
- Microeconomics: a new generation of economics driven by massive interactions
 - Auction marketplaces
 - The web as a new LEI of activities and economies

Information Navigation and Search

- We are in the early days of search and retrieval

Computational Advertising

- Targeting and matching sciences, Inferring user intent
- Pricing models (CPM, CPC, CPA, CPL, etc...)
- Large-scale optimization and yield management





Vision: Where the Internet's future is invented

 with innovative economic models for advertisers, publishers and consumers.

Mission: NEXT -- Invent the

Next generation Internet by defining the future media to

Engage consumers and

eXtend the economics for advertisers and publishers through new sciences that establish the

Technical leadership of Yahoo!





Scientific excellence

- World-recognized leadership through Business impact
- Build the Largest, Deepest and Smartest Research Organization focused on a few chosen areas
- Explore areas that nobody else is exploring
- Open model with strong emphasis on publication, peer review, and real problems





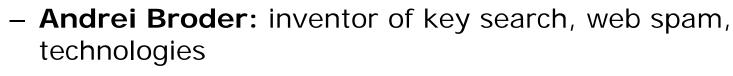
A sampling of the Top Researchers now at Yahoo! (Search)



 Prabhakar Raghavan: CTO Verity, Web Research architect at IBM, Head of Y!R



 Rajeev Rastogi: Well-known expert in databases/data mining, head of Bell Labs India, now head of Yahoo Labs, Bangalore





 Andrew Tomkins: chief Scientist of WebFountain, inventor of key algorithms, structure of Web graph



 Ricardo Baeza-Yates: renowned expert in text and query mining, authored seminal texts in IR, Director of Y!
 Research Barcelona and Santiago





A sampling of the Top Researchers now at Yahoo! (Community, Social Media, and User Experience)



 Raghu Ramakrishnan: world authority on data mining, database systems, and community/social search



 Duncan Watts: world expert in social networking, lead Yahoo!'s research in human social dynamics. Columbia University, director of the Collective Dynamics Group. Author of Six Degrees: The Science of a Connected Age.



- Malcom Slaney: multimedia analysis and user models, organized the Stanford CCRMA Hearing Seminar, IBM's Almaden Research Center, Interval Research, Apple's Advanced Technology Group, Schlumberger's Palo Alto Research Laboratory, and Bell Labs. Coauthor of "Principles of Computerized Tomographic Imaging," and "Computational Models of Auditory Function."





A sampling of the Top Researchers now at Yahoo! (Microeconomics)



Preston McAfee: VP and Research Fellow at Yahoo! Research in Burbank, CA, leads Microeconomics research. Prof at Caltech. Author *Introduction to Economic Analysis*, and *Incentives in Government Procurement*. Author of *Competitive Solutions: The Strategist's Toolkit*. One of 4 editors of the *American Economic Review*, the most prominent economics journal



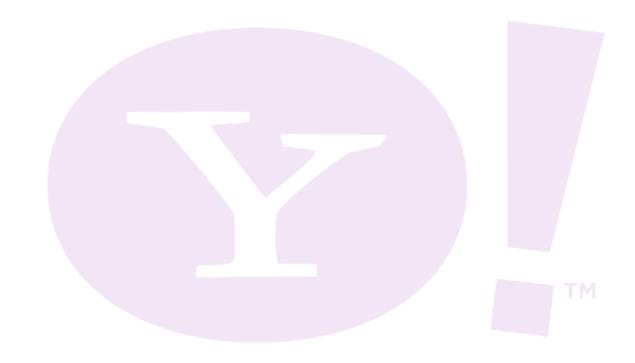
 Michael Schwarz: renowned economist, auctions, web, Faculty at Harvard and U.C. Berkeley, was a National Fellow at the Hoover institution at Stanford, a Robert Wood Johnson Foundation Scholar at UC Berkeley



David Pennock: world expert on expertise is the design and analysis of prediction markets. Very well known experiments with on-line marketplaces.

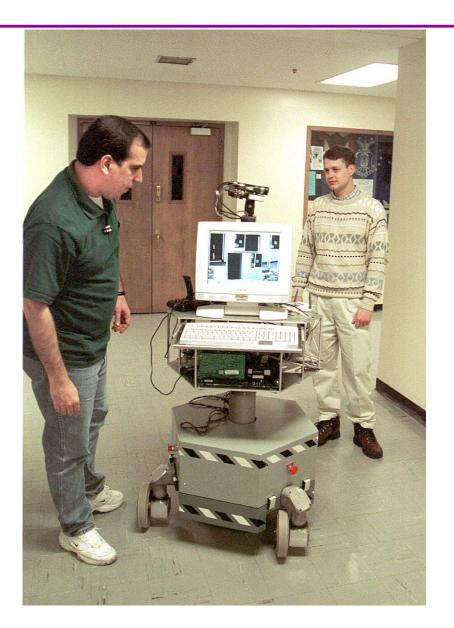


"Social" Search



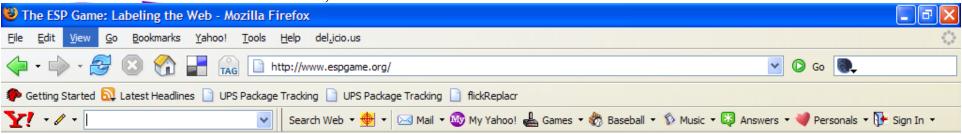
Is the Turing test always the right question?

A Digression: Computer Vision is hard





Luis Von Ahn, et al.





Terms of Service | FAQ | ESP Image Search | Contact Us | Credits

Funded in part by the National Science Foundation (NSF).

© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.

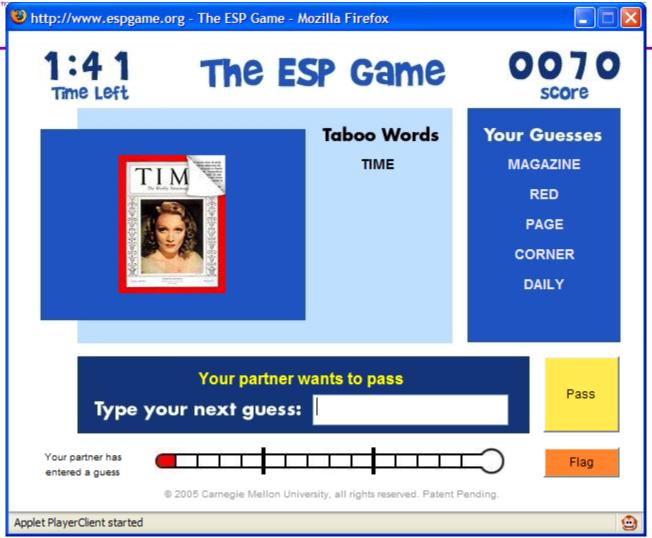










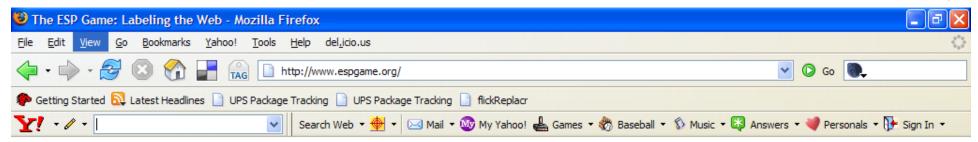














Terms of Service | FAQ | ESP Image Search | Contact Us | Credits

Funded in part by the National Science Foundation (NSF).

© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.



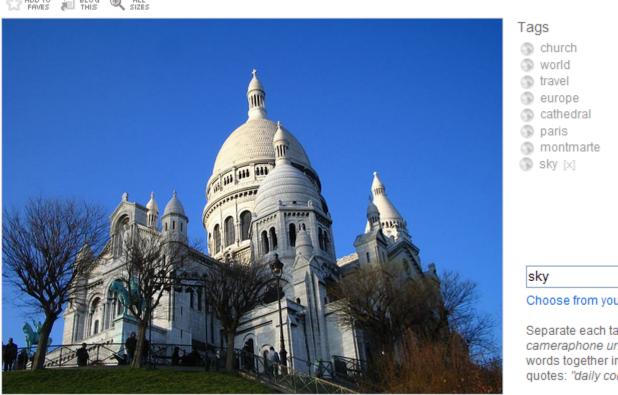
Done





2. User Organized Content

Content is tagged, described, organized, discovered, etc. not by "editors" but by the users themselves.

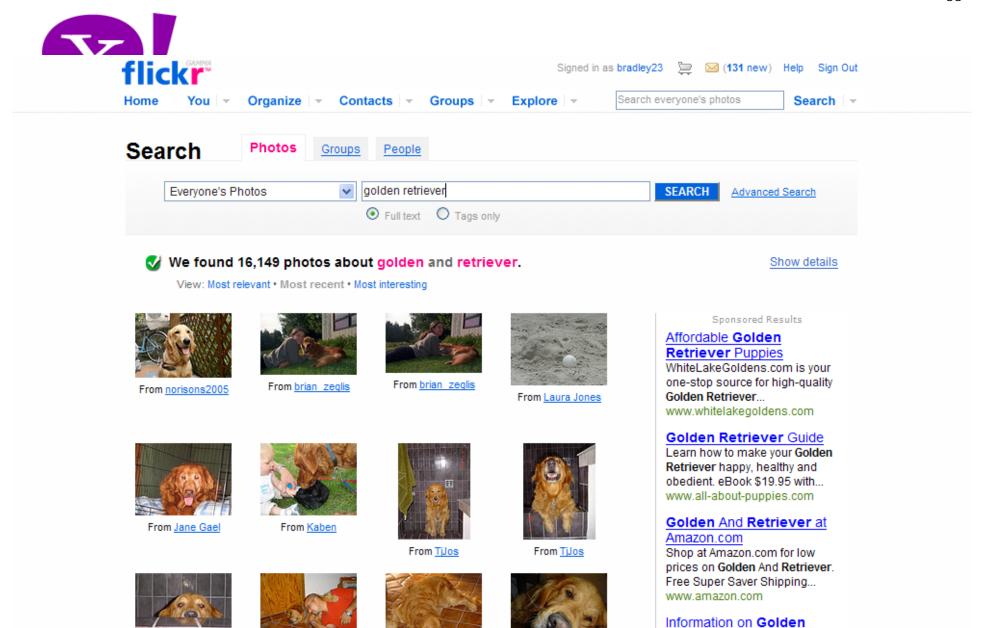


ADD

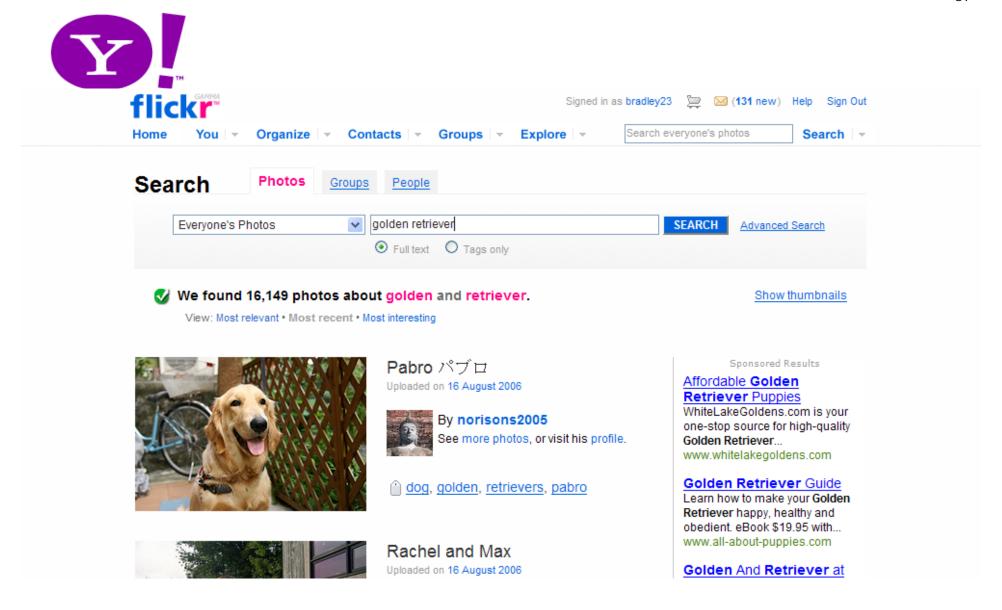
Choose from your tags

Separate each tag with a space: cameraphone urban moblog. Or to join 2 words together in one tag, use double quotes: "daily commute".

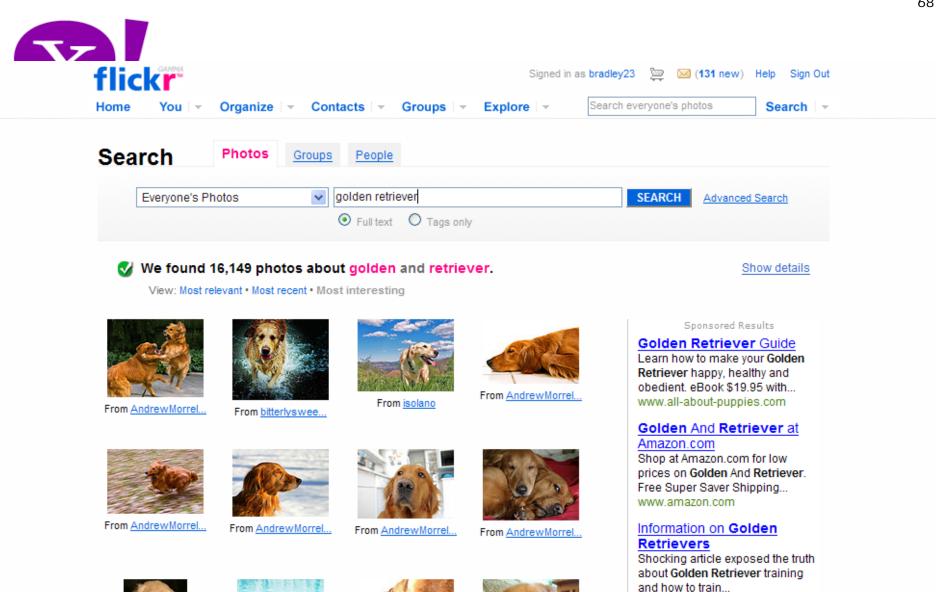








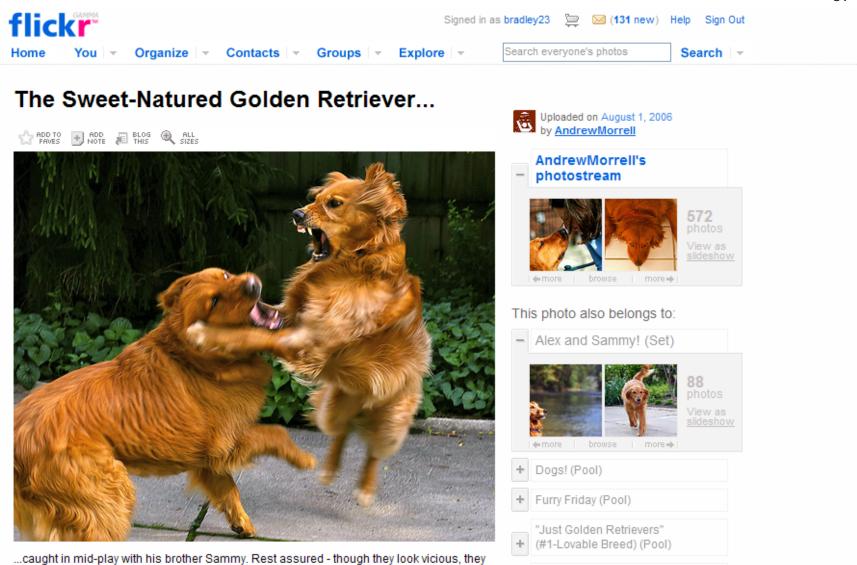






www.dog-e-news.com

Golden Retriever Memorial



are indeed playing. They'll get rough for a few moments, make a few growling noises, then



___ flickrDogs Blog: Submit photos





weaving major says:

wow, look at those fangs! great capture! Posted 2 weeks ago. (permalink)



Jeanne8 says:

As long as they are playing it looked like a mean fight at first. Blessings!

Posted 2 weeks ago. (permalink)



Amanda Clouse Photography pro says:

cute! Dale and Toby like to do this! Posted 2 weeks ago. (permalink)



djjoyboy says:

Amazing Catch Andrew, I wish you would make an EXPOSURE show in Pittsburgh sometime soon!

Posted 2 weeks ago. (permalink)



Photo Patty pro says:

Great action shot. I'm sure they were having a ball.

Posted 2 weeks ago. (permalink)



Boston Fan in Michigan pro says:

You tagged that 'gentle'?? Egads. Those teeth look awfully wolfish...:P Posted 2 weeks ago. (permalink)



Duogirl says:

HAHA!!!

Posted 2 weeks ago. (permalink)



saldous says:

Tags

- dog
- golden retriever
- morrell
- alex
- sammy
- borell
- docile
- S loving
- playful
- gentle
- top20dogpix
- interestingness
- 3 1500v60f
- SpecAnimal
- a exploretop20

Add a tag

Additional Information

- © All rights reserved
- Taken with a Canon EOS-1D Mark II.
 More properties
- Taken on June 23, 2006
- · See different sizes
- o 215 people call this photo a favorite
- Viewed 8,339 times

Flag this photo as "may offend"?

Home | Sign Up Sign In | Help

Photos: Explore Flickr • Learn More



Tags / jaguar / clusters



SEARCH

(Or, try an advanced search.)











<u>car</u>, <u>cars</u>, <u>auto</u>, <u>etype</u>, <u>automobile</u>, <u>classic</u>, <u>vintage</u>, <u>autoshow</u>, <u>red</u>, <u>show</u>

> See more in this cluster...











zoo, animal, cat, animals, bigcat, seattle, woodlandparkzoo, sleep, edinburgh, caged

> See more in this cluster...











guitar, fender

See more in this cluster...











aircraft, raf

See more in this cluster...











mac, apple

See more in this cluster...



Challenges in social media

- How do we use these tags for better search?
- What's the ratings and reputation system?
- How do you cope with spam?
- The bigger challenge: where else can you exploit the power of the people?
- What are the incentive mechanisms?





Case Study: Yahoo! Answers & Y!R

Illustrating New Research Challenges





The kinds of queries that rely on domain expertise...

- "Do you know a reputable plumber in Atlanta?"
- "Where is the cool nightlife in Soho?"
- "What political blogs do you think I'd enjoy reading?"
- "Where can I buy a cool pair of boots?"

These kinds of queries are ill-served by today's search engines, but are ironically the most valuable (i.e. transactional queries.)

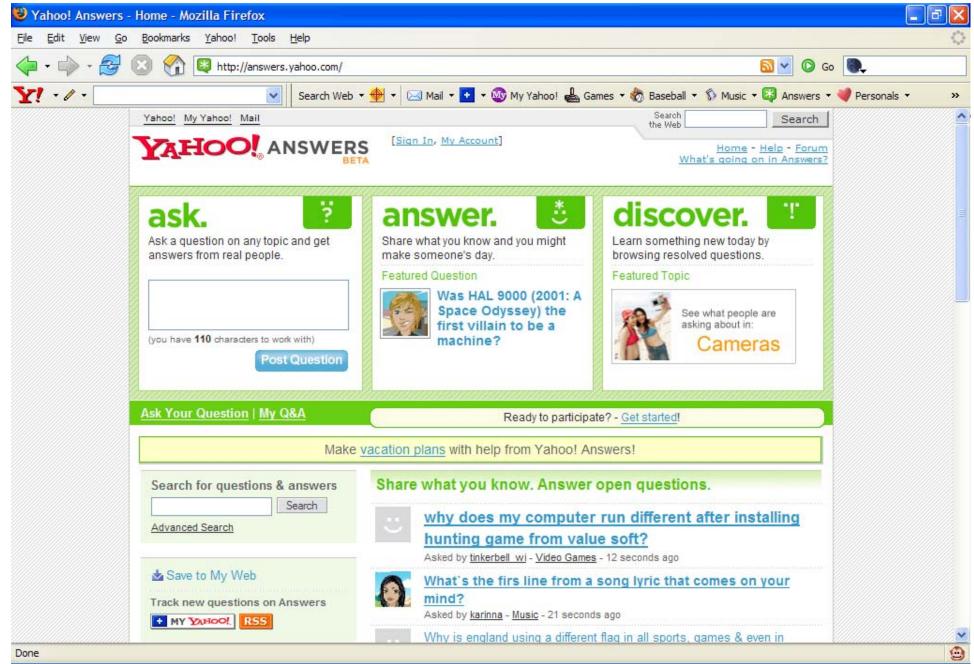


"Better Search Through People" Case Study: Yahoo! Answers

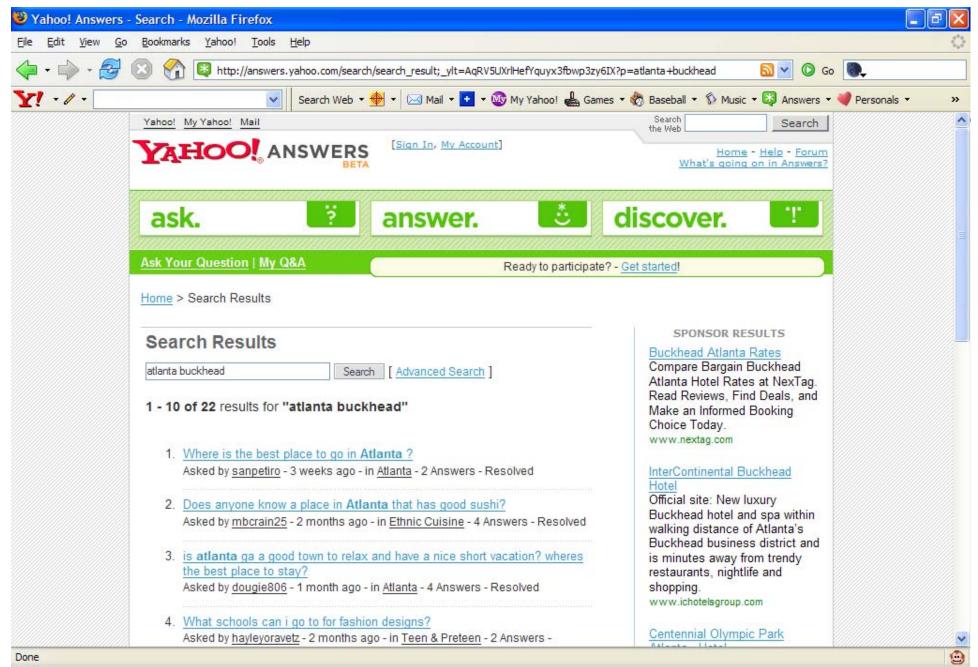


- Today's state-of-the-art web search (at best) connects a person with a query to a web page that contains those characters
- Even as we nobly try to index every page, (every book, every video, every podcast, etc.) it is worth noting that these are artifacts of human knowledge
- Yahoo! Answers addresses this problem. It connects a person with a question to a community of people best suited to answer the question.

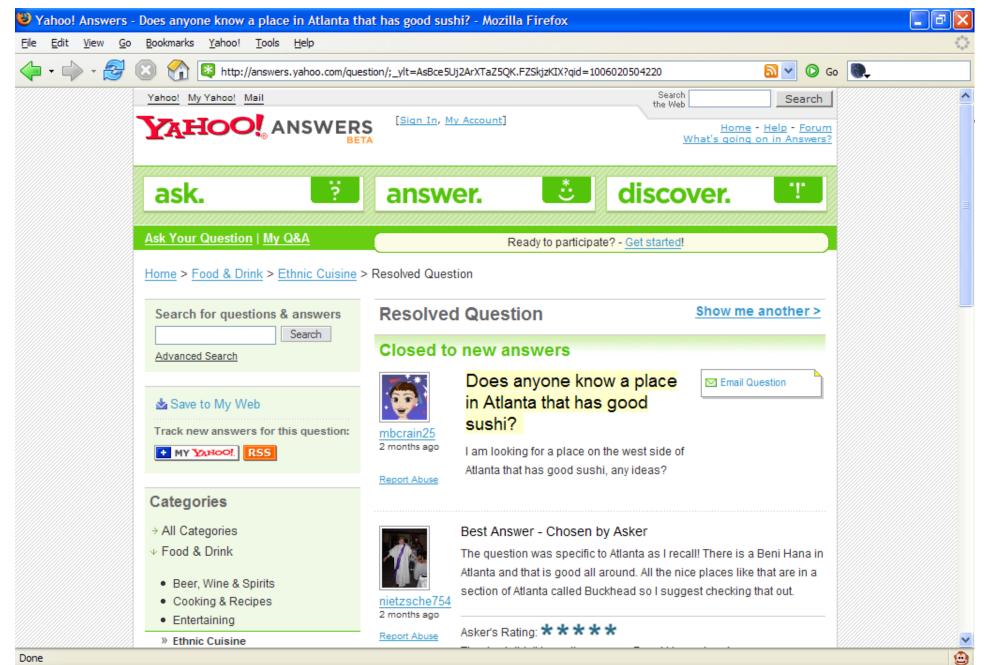


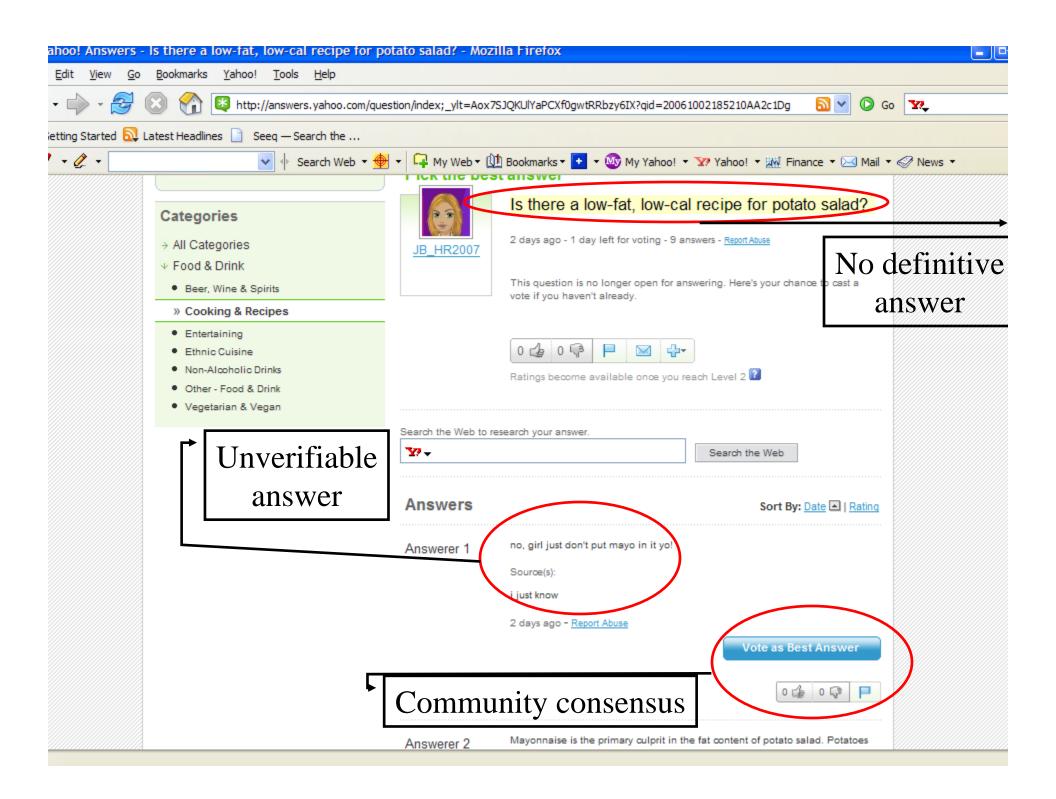


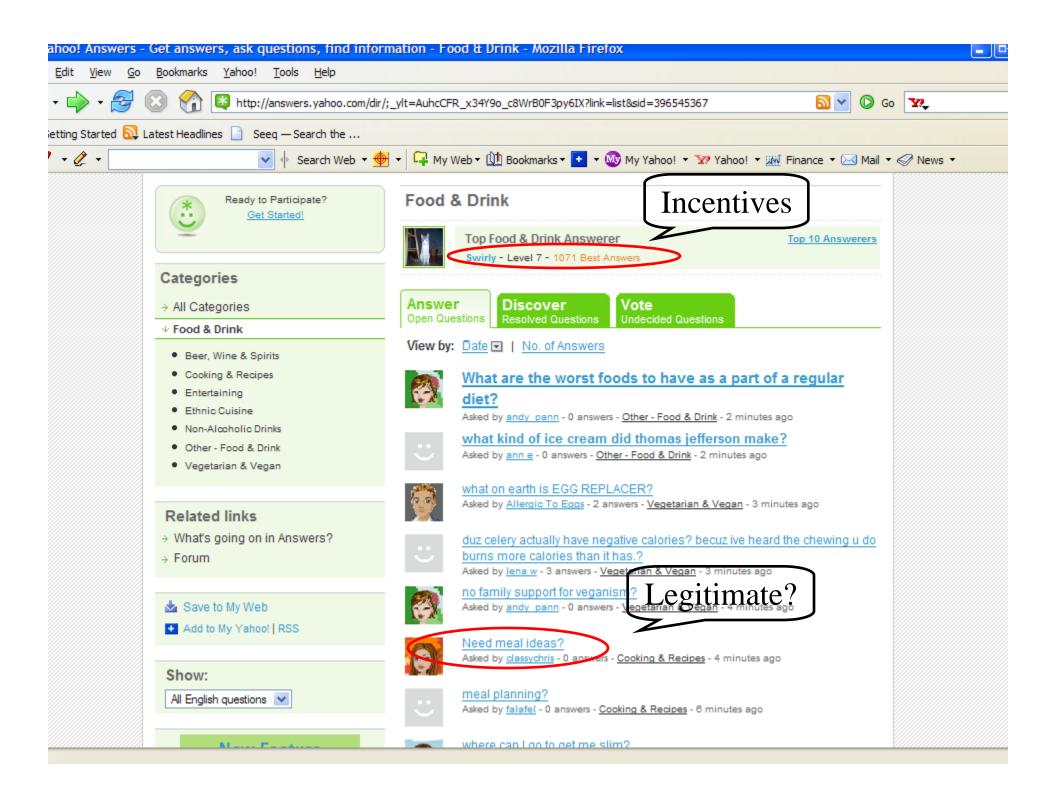














- Which questions are legitimate?
- What is the incentive system?
- How do we validate answers?
- What is the role of the community?
- What is the reputation system?





What are the challenges?

- Community of users
 - Social system
- Incentives and reputations
 - Economic system
- Poorly phrased, grammatically limited queries
 - Language analysis
- Improving user experience from past data
 - Data mining





Evolution of Today's Search

What changes are happening in the short term?





- Internet firmly moved from a curiosity to a substrate for life activity
- Content growing, changing, diversifying, fragmenting
- Search evolving in response
- Value migrating to ecosystem
- Semantics of content unlock the value in the ecosystem





Towards Getting Things Done... vs. Searching











- Loved the vacation, want to make that sweet Italian coffee at home
 - Search: "making good espresso"
 - Browse: coffeegeek.com
 - Study: Temperature surfing a Rancilio Silvia
 - Price comparison: nextag.com
 - Vendor comparison
 - Purchase from chriscoffee.com
 - Frothing milk tutorial
 - Cleaning and maintenance
 - Purchase grouphead brush and Urnex





Loved the vacation, want to make that sweet Italian coffee at home





Trends in task complexity

- Dawn of search:
 - Navigational queries
 - Pockets of information
- Today:
 - Increasing migration of content online
 - New forms of media only available online
 - Infrastructure for payments and reputation sufficient for many users





- Long-running user goals
- Search as hub:
 - start there
 - return for resource discovery and at task boundaries
 - traverse the web broadly to complete task
- Web services integrated into task





Content Growth





Content type	Amount of content produced per day
Published content	3-4 GB
Professional web content	$\sim 2~ ext{GB}$
User generated content	8-10 GB
Private text content	$\sim 3~{ m TB~(300x~more)}$
Upper bound on typed content	\sim 700 TB (\sim 200x more)

[Ramakrishnan and Tomkins 2007]





Metadata type	Amount of metadata produced per day
Anchortext	100 MB
Tags	40 MB
Pageviews	180 GB
Reviews	Around 10 MB

[Ramakrishnan and Tomkins 2007]





What does this mean to Search?

- Publishers and search engine collaborate
- Users see richer search experience
- Accomplish their tasks faster and more effectively
- Example: abstracts surfacing structured content

Back to the future?





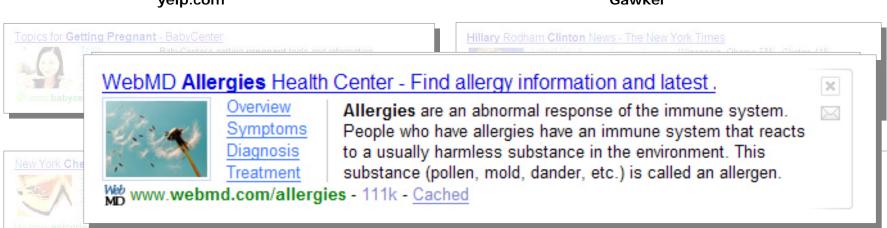
Search results of the future





WebMD Allergies Health Center - Find allergy information and latest

yelp.com Gawker



epicurious LinkedIn



answers.com webmd







Search results of the future



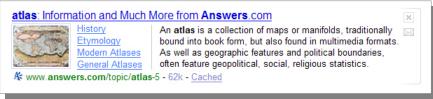
yelp.com



babycenter



epicurious



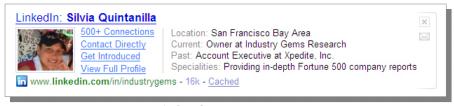
answers.com



Gawker



New York Times



LinkedIn



webmd

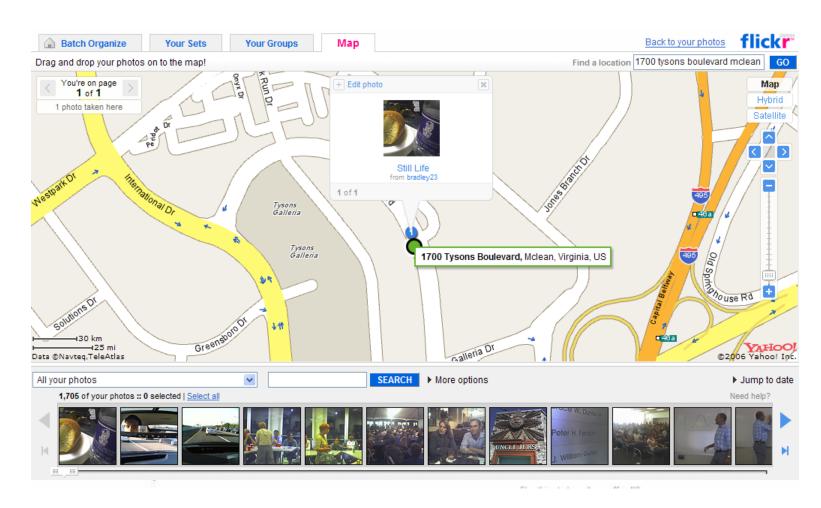




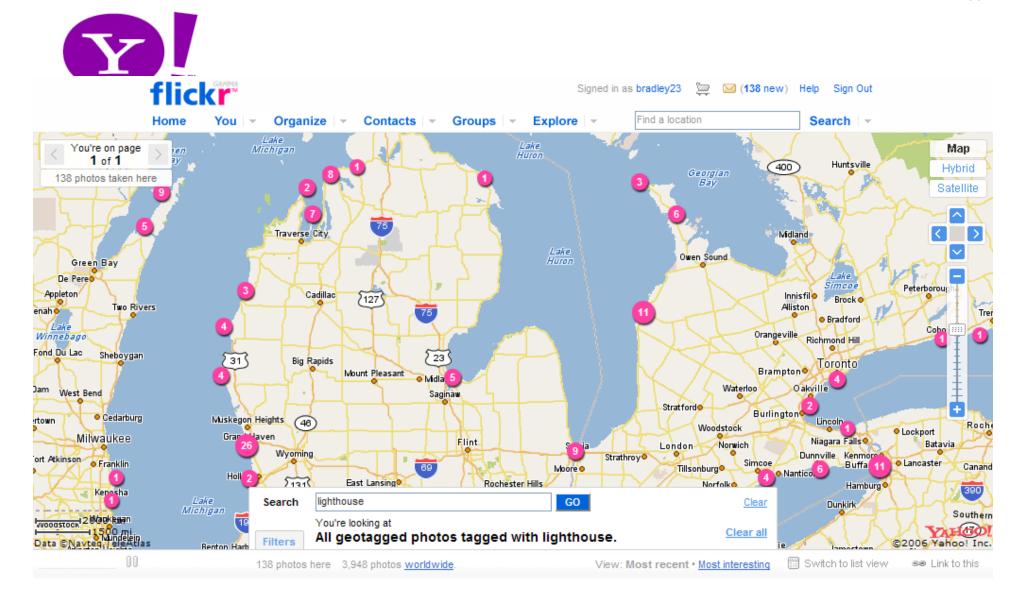
What is a Library?
What is a book?
What is a reference?



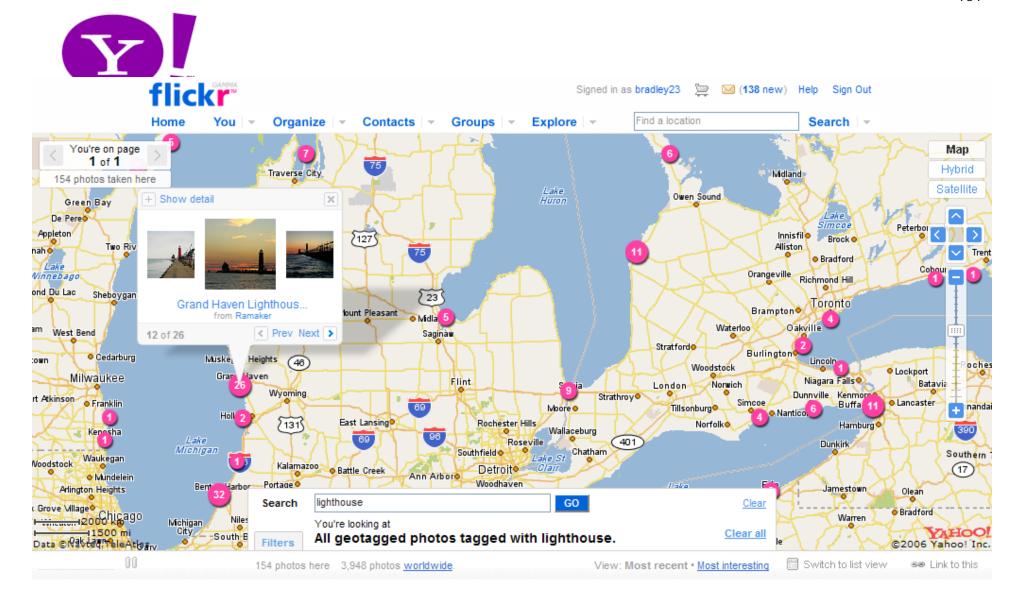
But wait... there's more!



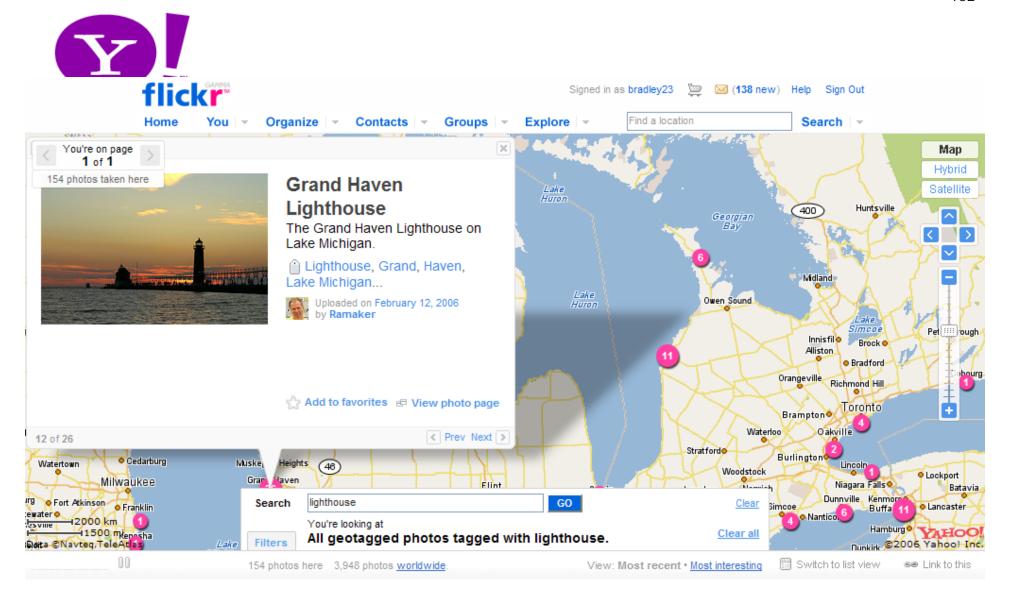




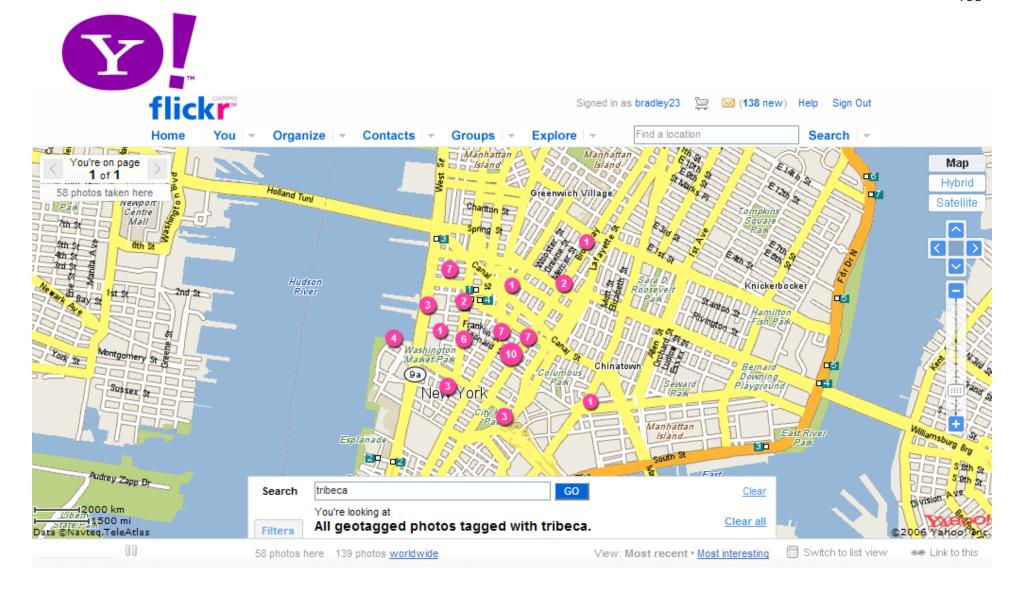




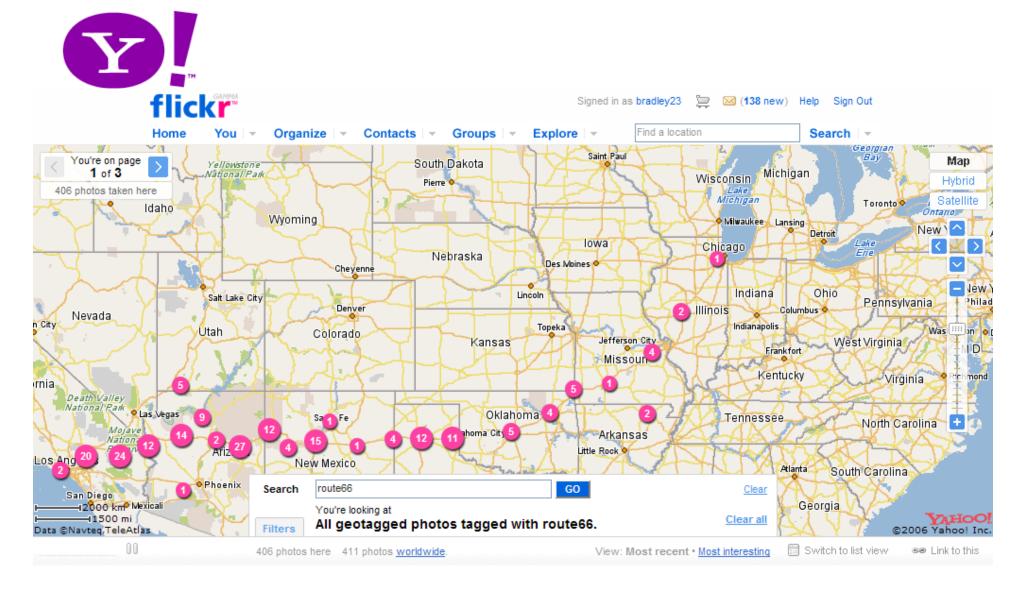




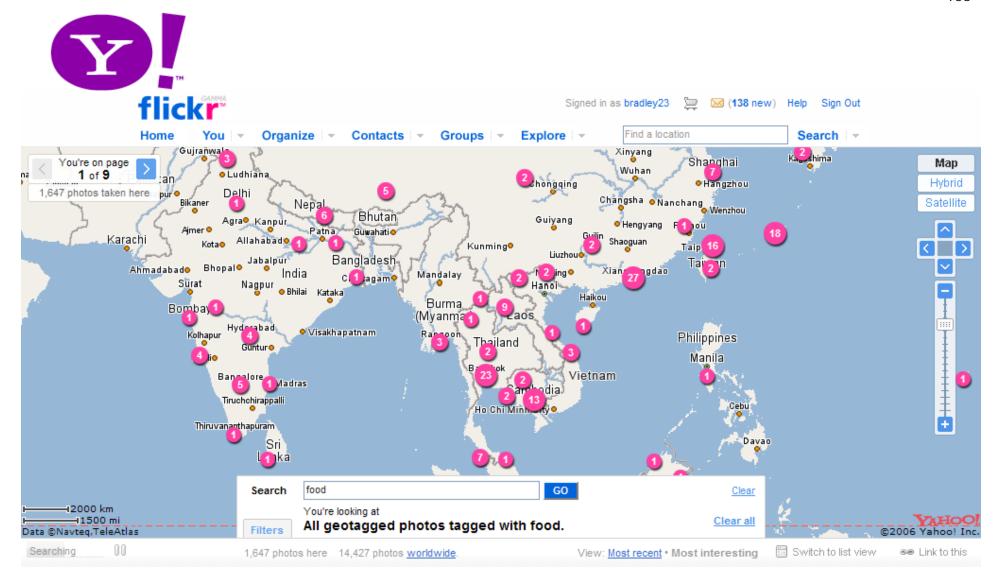




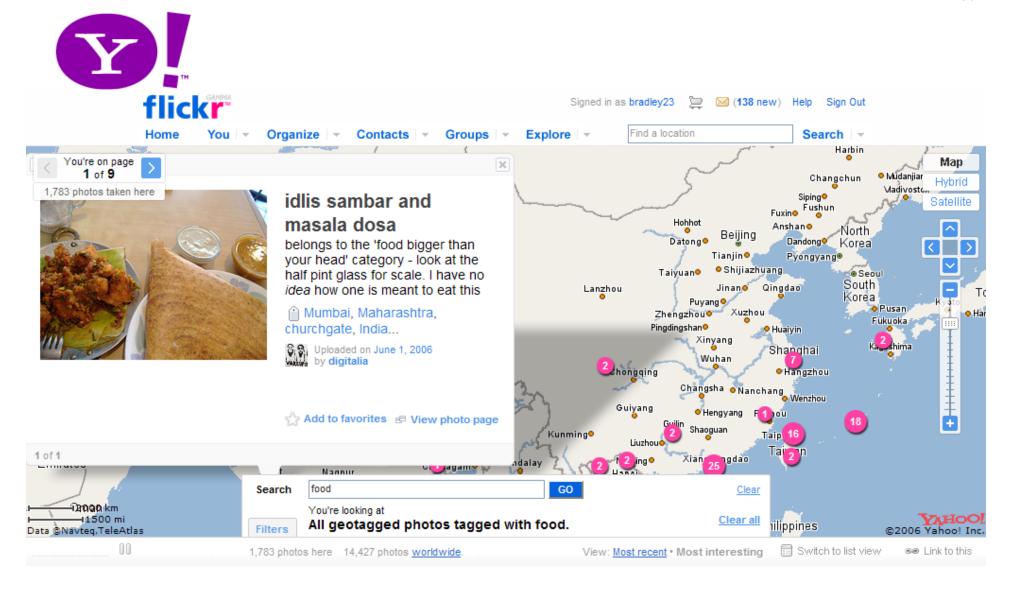




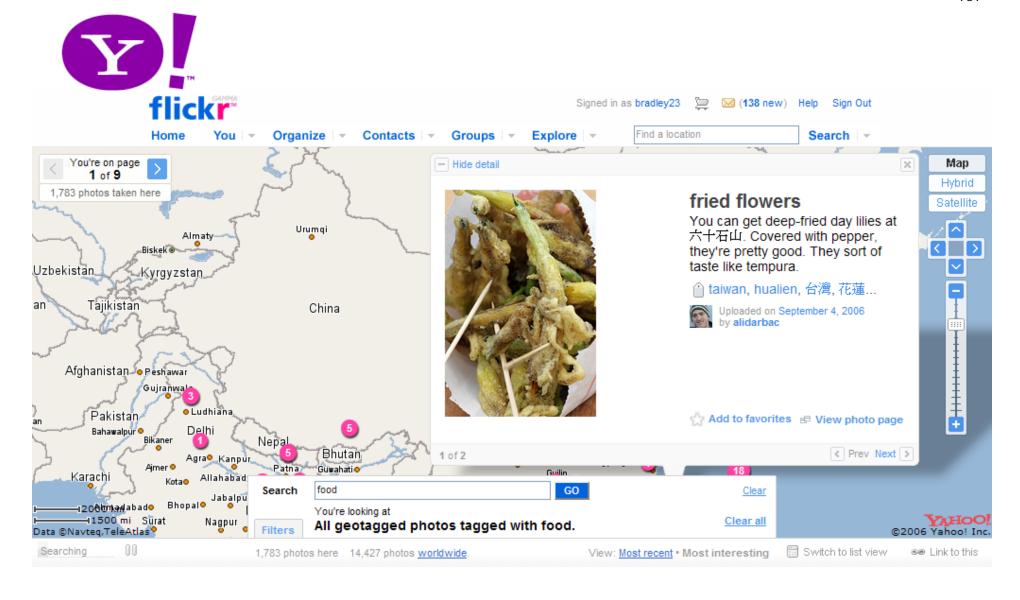




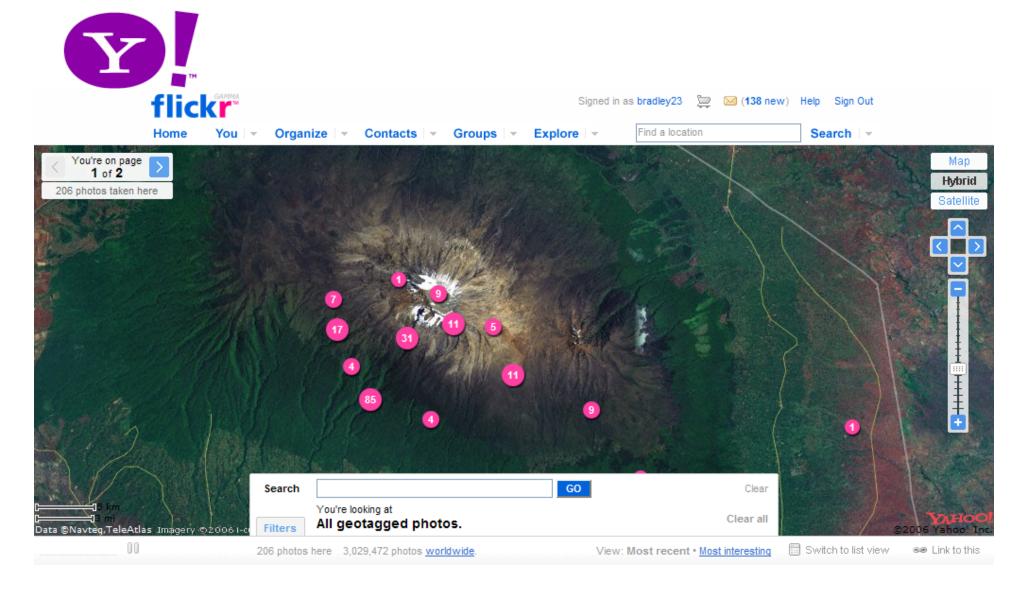




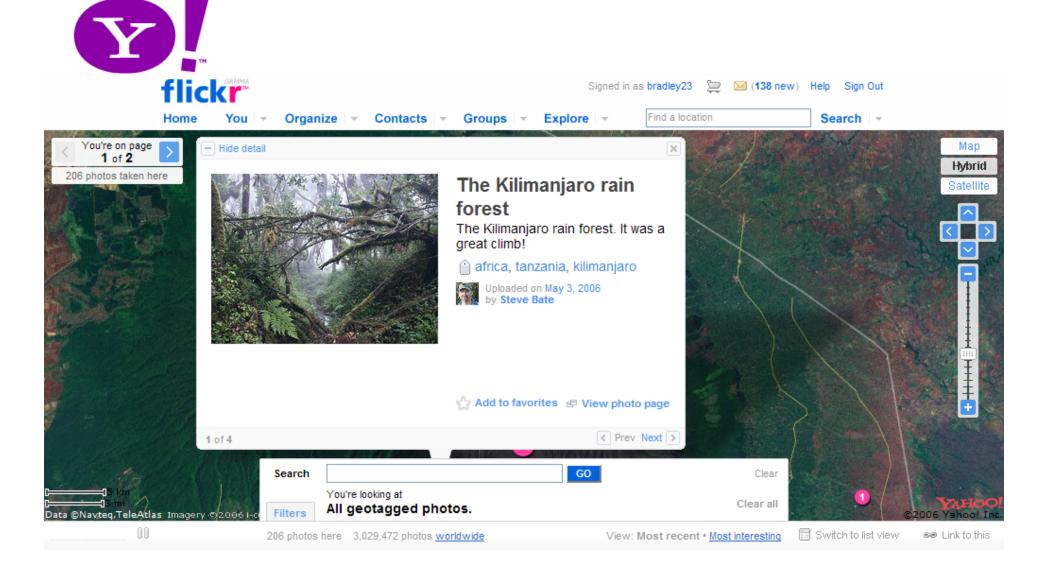




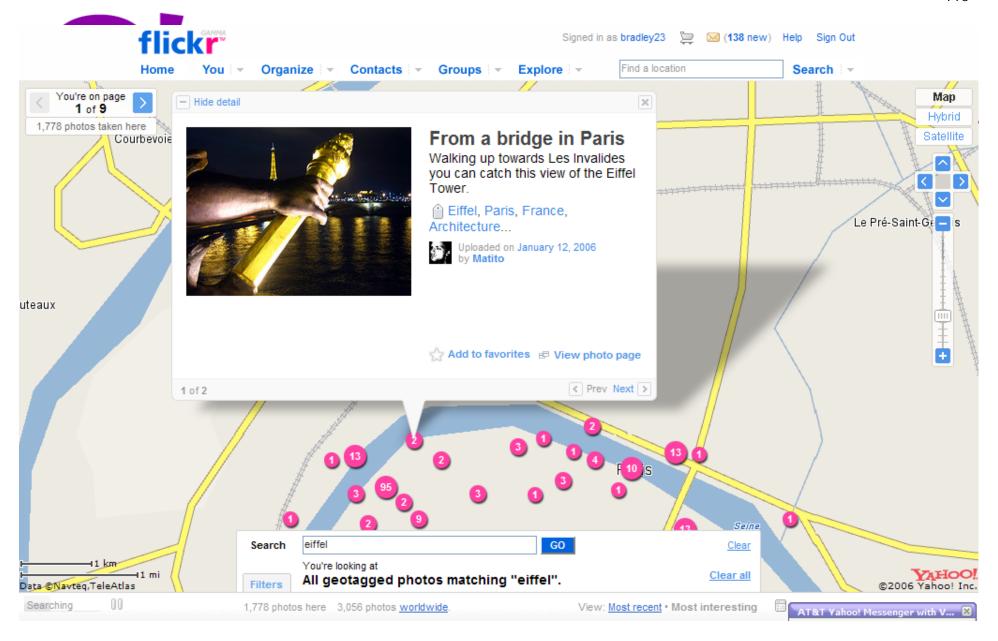




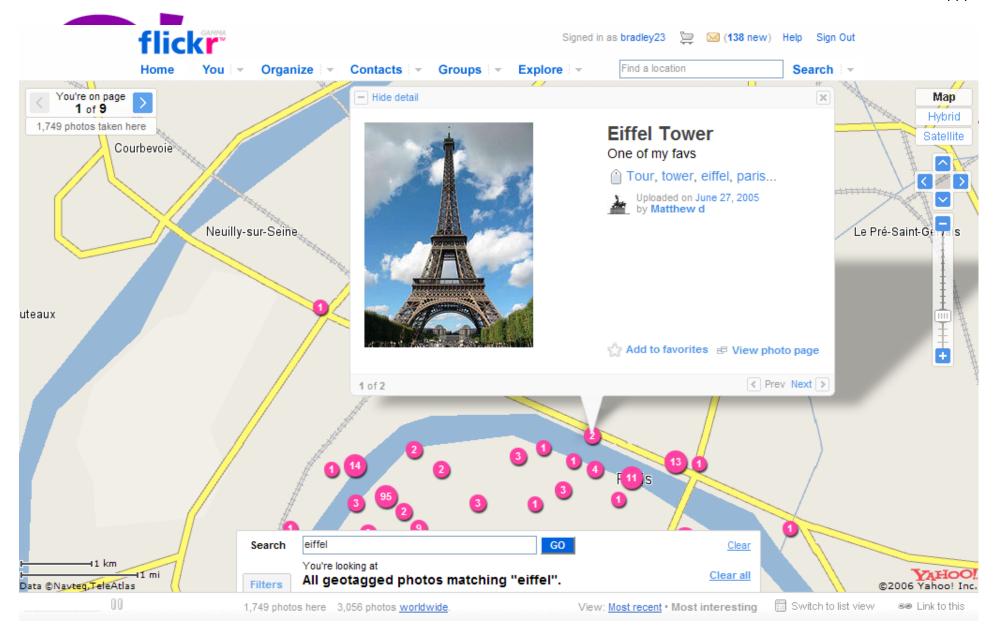




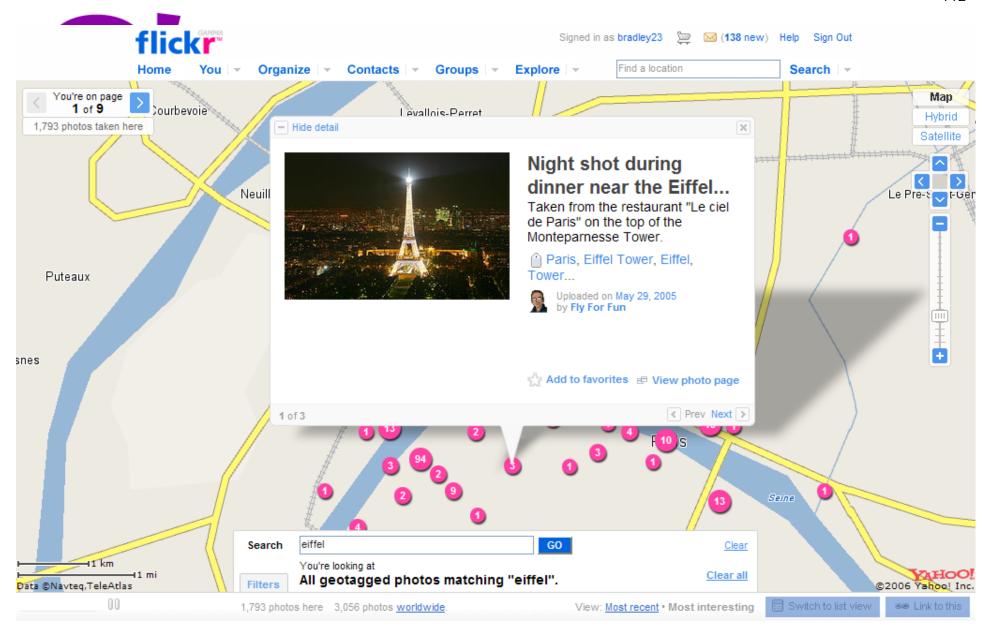
















Community:

- How do you know what to believe on the Internet?
- Trust models on-line and trust propagation
- What makes communities thrive? Whither?
- Social media, tagging, image and video sharing
- Microeconomics: a new generation of economics driven by massive interactions
 - Auction marketplaces
 - The web as a new LEI of activities and economies

Information Navigation and Search

- We are in the early days of search and retrieval

Computational Advertising

- Targeting and matching sciences, Inferring user intent
- Pricing models (CPM, CPC, CPA, CPL, etc...)
- Large-scale optimization and yield management





- The notion of "corpus" and publishing is changing fundamentally
- We still do not have the basic sciences to understand what is happening
- The web is changing, content sources are fragmenting and changing the source distribution is radically changing
 - Publisher consumer divide is becoming fuzzy
- Search engine interface is finally changing to adapt





Concluding Thoughts (2)

- The view that Search is everything is LIMITED (at best)
 - Economics of publishing and advertising
 - Users do not differentiate ad and content
 - Behavioral data is the most powerful
 - "Nothing predicts behavior like behavior"
- Monetization and economic value an intrinsic part of system design
 - Not an afterthought
 - Mistakes are costly!
- Computing meets humanities like never before sociology, economics, anthropology ...
- A more holistic view of Search and Information Navigation is needed





Thank You! & Questions?

Usama_fayyad@yahoo.com





No time to cover today

- Micro-Economics of the Web
 - Auction marketplaces
 - Marketplace and Exchange Design
 - The economics of Engineering IT Decisions
- Computational Advertising
 - Targeting and matching sciences
 - Inferring user intent
 - Pricing models (CPM, CPC, CPA, CPL, etc...)
 - Large-scale optimization and yield management





- What media experiences grow/thrive?
 - Data-driven stickiness
 - EBay reputations, Amazon recommendations
- Which ones die out?
- More global/youthful demographic
- An intrinsically data-driven science





- Micro-Economics of the Web
 - Auction marketplaces
 - Marketplace and Exchange Design
 - The economics of Engineering IT Decisions
- Computational Advertising
 - Targeting and matching sciences
 - Inferring user intent
 - Pricing models (CPM, CPC, CPA, CPL, etc...)
 - Large-scale optimization and yield management

