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.
Acknowledgement

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
Overview

- 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, Internet Users Will Number Over 1 Billion by 2007

Yahoo! is the #1 Destination on the Web

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

Sources: Mediamark Research, Spring 2004 and comScore Media Metrix, February 2005.
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
Knowledge of users & their behavior throughout the purchase funnel can grow brand & direct response revenue.

> $200B Brand Advertising Market

Most time & activity is in consideration & engagement, but there are limited metrics & reach strategies.

> $250B Direct response Market
A Tale of Two Search Engines
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

User

Search

Indexer

Indexes

Indexes

Ad indexes

The Web

Web spider

Research
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: `casino` 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 mechanisms in 2007
Search advertising
Search query:

- Did you mean: **cannon camera**

1. **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 - Cached - More from this site

3. **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

4. **Canon USA**
   Manufacturer of professional and consumer imaging equipment and information systems including copiers, printers, image filing systems, cameras and lenses, and more.
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?
How it works

Advertiser

I want to bid $5 on

*canon* camera

I want to bid $2 on

*cannon* camera

Ad Index

Sponsored search engine

Engine decides when/where to show this ad.

Landing page

Engine decides how much to charge advertiser on a click.
Monetization

- 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
Retrieving ads

• Standard tricks from IR: phrasing, spell-correction, ranking ...
  - Use both ad snippet and landing page features
  - Advertisers are trying to help! (Anti-spam)
Research Ads go in slots like these.
Higher slots get more clicks
2. Order the ads

• Most generally, composite IR+Econ score ... for today’s talk, focus on Econ

• Original GoTo/Overture scheme:
  – Order by bid
Bid ordering
(Overture/former Yahoo!)

1. Boeing 747-400 Family
   Corporate product page from Boeing.
   Category: Planes > Boeing 747 Family
   www.boeing.com/commercial/747family - 11k - Cached - More from this site - Save

2. Wikipedia: Boeing 747
   The Boeing 747, commonly called the Jumbo Jet, is one of the most recognizable modern airliners and is the largest airliner currently in airline service. First flown commercially in 1970, it held the size record for more than 35 years, although... The four-engine 747, produced by Boeing Commercial Aircraft, uses a two...
   Quick Links: History - Development - In service
   en.wikipedia.org/wiki/Boeing_747 - 102k - Cached - More from this site - Save

3. Wikipedia: 747
   ==Events== This article is about the year 747. For the aircraft, see Boeing 747. For the song by Kent, see 747 (song). For other uses of 747, see 747 (number ...
   Quick Links: Events - Births - Deaths
   en.wikipedia.org/wiki/747 - 16k - Cached - More from this site - Save

SPONSOR RESULTS
747 on eBay
Shop eBay for anything and everything - from specialized gifts to...
www.ebay.com

Shop Boeing 747-400 Miniature Models
Shop Boeing 747-400 miniature models at FlightMiniatures.com. See...
www.flightminiatures.com

Canon Fax L777 Toner-free Shipping
Save on compatible Canon Fax L777 toner. Purchase our...
www.ink-refills-ink.com

747 Airplanes - Free Shipping
Thousands of hand crafted fully assembled model plane replicas in...
Revenue ordering, simplified

• Each ad $j$ has an expected click-through denoted $\text{CTR}_{jq}$ on query $q$
• Ad $j$’s bid is denoted $b_j$
• Then, expected revenue from this ad is $R_j = b_j \times \text{CTR}_{jq}$
• Order ads by $R_j$
Revenue ordering?
Economic ordering

- Bid and revenue ordering: two forms of ordering by an \textit{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

Edelman, Ostrovsky, Schwarz 2006
A new convergence

- 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
  - Shop

- Gray areas
  - Find a good hub
  - Car rental Finland
  - Exploratory search “see what’s there”

Broder 2002, A Taxonomy of web search
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

1. User searches on the word “Credit Card”

2. Sees that “Credit Card” falls under the category: “Finance/Credit and Credit Services”

3. Serves “Finance/Credit and Credit Services” banners to User anywhere on the Yahoo! network within 1 HOUR

All within 1 HOUR

25% - 261% higher CTR
Way Impulse Works

- 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
Moving Down the Funnel

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

Behavioral Targeting (BT)
• 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:

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 drove **139% more clicks** on algorithmic and sponsored links…

...specifically driving **249% more sponsored search clicks**…

...and driving **91% more activity** on the HarrisDirect.com website.
Example of challenges of inventing the new sciences of the Internet
Evolution of Social Media

• 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...
Photos from your contacts
Or, only show photos from your family and friends
What makes Flickr special?

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 Joë

- *Mum and...*
  From Chrissie64

- *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.
**Number Six** says:
Wonderful composition - so much to hold one's interest.
Posted 8 months ago. (permalink)

**glaciergirl** says:
Beautiful in so many ways.
Posted 8 months ago. (permalink)

**glaciergirl** says:
Wow! gorgeous on so many levels.
Posted 8 months ago. (permalink)

**ken mccown** says:
that's as crisp as you can get a photo! nice!
Posted 8 months ago. (permalink)

**David Nicolas [deleted]** says:
Stunning :
Posted 8 months ago. (permalink)

**Tilak Abeyesinghe** says:
Fantastic picture. Congratulations
Posted 8 months ago. (permalink)

**Magic™** says:
Amazing .. breathtaking photo..
Posted 8 months ago. (permalink)

**Scatti di vita** says:
Molto bella!!
Posted 8 months ago. (permalink)
What makes Flickr special?

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
New Science?

• 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...
Areas of Research

• **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!
How we get there

- 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

- **Andrei Broder**: inventor of key search, web spam, technologies

- **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
The ESP Game: Labeling the Web - Mozilla Firefox

Welcome, BRADLEY23
(Not BRADLEY23? click here)

How to Play
Play NOW!

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

Funded in part by the National Science Foundation (NSF)
© 2008 Carnegie Mellon University, all rights reserved. Patent Pending
The ESP Game

Time Left: 0:36
Score: 0140

Taboo Words: SMILE
Your Guesses: WOMAN

Agreed on: WOMAN
Type your next guess:

Your partner has entered a guess

Flag
Pass

© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.
Score: 140

Time Bonus: 0

Total for this Game: 140

Percentile: 22

Cumulative Score: 33285

Rank: Gifted

Points needed for next rank: 166715
2. User Organized Content

Content is tagged, described, organized, discovered, etc. not by “editors” but by the users themselves.
We found 18,149 photos about golden and retriever.
Search

Photos  Groups  People

Everyone's Photos  golden retriever  SEARCH  Advanced Search

We found 18,149 photos about golden and retriever.
View: Most relevant  Most recent  Most interesting

From AndrewMorrel...
From bitterlyswae...
From spojano
From AndrewMorrel...

From AndrewMorrel...
From AndrewMorrel...
From AndrewMorrel...
From AndrewMorrel...

Sponsored Results

Golden Retriever Guide
Learn how to make your Golden Retriever happy, healthy and obedient. eBook $19.95 with...
www.all-about-puppies.com

Golden And Retriever at Amazon.com
Shop at Amazon.com for low prices on Golden And Retriever. Free Super Saver Shipping...
www.amazon.com

Information on Golden Retrievers
Shocking article exposed the truth about Golden Retriever training and how to train...
www.dog-e-news.com

Golden Retriever Memorial
The Sweet-Natured Golden Retriever...

...caught in mid-play with his brother Sammy. Rest assured - though they look vicious, they are indeed playing. They'll get rough for a few moments, make a few growling noises, then
**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)

**dijooyboy** 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 / jaguar / clusters

- **car, cars, auto, etype, automobile, classic, vintage, autoshow, red, show**
  - 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
Subjective Queries

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.
Search Results

atlanta buckhead

1 - 10 of 22 results for "atlanta buckhead"

1. Where is the best place to go in Atlanta?
   Asked by sanpetre - 3 weeks ago - in Atlanta - 2 Answers - Resolved

2. Does anyone know a place in Atlanta that has good sushi?
   Asked by mbcrain25 - 2 months ago - in Ethnic Cuisine - 4 Answers - Resolved

3. is atlanta ga a good town to relax and have a nice short vacation? where is the best place to stay?
   Asked by dougie906 - 1 month ago - in Atlanta - 4 Answers - Resolved

4. What schools can i go to for fashion design?
   Asked by hayleyoravetz - 2 months ago - in Teen & Preteen - 2 Answers -
Yahoo! Answers - Does anyone know a place in Atlanta that has good sushi? - Mozilla Firefox

http://answers.yahoo.com/question/_ylt=A8oe5U2ArXTaZ5QK.FZSkjxi?qid=1006020504220

Ask Your Question | My Q&A

Ready to participate? - Get started!

Home > Food & Drink > Ethnic Cuisine > Resolved Question

Resolved Question

Closed to new answers

Does anyone know a place in Atlanta that has good sushi?

mbcrain25
2 months ago

I am looking for a place on the west side of Atlanta that has good sushi, any ideas?

Best Answer - Chosen by Asker

The question was specific to Atlanta as I recall! There is a Beni Hana in Atlanta and that is good all around. All the nice places like that are in a section of Atlanta called Buckhead so I suggest checking that out.

nietzsche754
2 months ago

Asker's Rating: ★★★★★

Beer, Wine & Spirits
Cooking & Recipes
Entertaining

Categories
Is there a low-fat, low-cal recipe for potato salad?

No definitive answer

Unverifiable answer

Community consensus
Where is the Science?

• 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?
Evolution of Search Today

- 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
I want to book a vacation in Tuscany.
Example 2

- Loved the vacation, want to make that sweet Italian coffee at home
  - Search: “making good espresso”
  - Browse: coffegeek.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
Things to notice

- 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 trends

<table>
<thead>
<tr>
<th>Content type</th>
<th>Amount of content produced per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published content</td>
<td>3-4 GB</td>
</tr>
<tr>
<td>Professional web content</td>
<td>(\sim 2) GB</td>
</tr>
<tr>
<td>User generated content</td>
<td>8-10 GB</td>
</tr>
<tr>
<td>Private text content</td>
<td>(\sim 3) TB (300x more)</td>
</tr>
<tr>
<td>Upper bound on typed content</td>
<td>(\sim 700) TB ((\sim 200)x more)</td>
</tr>
</tbody>
</table>

[Ramakrishnan and Tomkins 2007]
<table>
<thead>
<tr>
<th>Metadata type</th>
<th>Amount of metadata produced per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchortext</td>
<td>100 MB</td>
</tr>
<tr>
<td>Tags</td>
<td>40 MB</td>
</tr>
<tr>
<td>Pageviews</td>
<td>180 GB</td>
</tr>
<tr>
<td>Reviews</td>
<td>Around 10 MB</td>
</tr>
</tbody>
</table>

[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
What is a Library?
What is a book?
What is a reference?
But wait... there’s more!
Grand Haven Lighthouse
The Grand Haven Lighthouse on Lake Michigan.

Uploaded on February 12, 2006 by Ramaker
idli sambar and masala dosa belongs to the 'food bigger than your head' category - look at the half pint glass for scale. I have no idea how one is meant to eat this.

Mumbai, Maharashtra, churchgate, India...

Uploaded on June 1, 2006 by digitalia
fried flowers
You can get deep-fried day lilies at 六十石山. Covered with pepper, they're pretty good. They sort of taste like tempura.

Uploaded on September 4, 2006
by aldarbac
From a bridge in Paris

Walking up towards Les Invalides you can catch this view of the Eiffel Tower.

Eiffel, Paris, France, Architecture...

Uploaded on January 12, 2006 by Matito
Night shot during dinner near the Eiffel...  
Taken from the restaurant "Le ciel de Paris" on the top of the Montparnasse Tower.

Paris, Eiffel Tower, Eiffel, Tower...

Uploaded on May 29, 2005 by Fly For Fun
Areas of Research

- **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
Concluding Thoughts (1)

- 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 is this science?

- 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
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