

Googling

I'm Feeling (un)Lucky

Gregory Conti
United States Military Academy
West Point, New York
gregory-conti@usma.edu



The views expressed in this presentation are those of the author and do not reflect the official policy or position of the United States Military Academy, the Department of the Army, the Department of Defense or the U.S. Government.



With regards to Google...

- I am not stating that they are doing anything malicious
- This is a meta analysis of potential capability and does not attempt to divine their intent beyond what they have stated publicly
- Assuming a 100% non-malicious Google, they are still subject to legal pressure, such as subpoenas, to reveal personal information
- I have used only publicly available resources, e.g. no insider information

Google and other information services are amassing an unprecedented amount of personal information on you and I as well as the companies we work for.

Is this trust misplaced?

Do you care?

Ask Yourself...

- **Have you or your employees ever searched on something that you would not want your mother to know about?**
- Do you use Gmail?
- Have you or your employees ever vanity surfed?
- Have you or your employees ever used a search engine to seek out corporate intelligence?
- Have you or your employees ever searched for sensitive personal information? (SSN?)

What if these footprints never went away?

Related News...

- Google.cn granted license for operation in China
(censorship required to do "legal business" in China)
http://news.xinhuanet.com/english/2006-02/22/content_4213904.htm
- Google moving search records out of China
<http://www.networkworld.com/news/2006/030106-google-china-search.html>
- American Airlines subpoenas Google, YouTube
(AA demanding that Google reveal the name of the person who posted airline's training videos on their Web sites.)
http://news.com.com/American+Airlines+subpoenas+Google,+YouTube/2100-1030_3-6048032.html
- Google Subpoena Woes Double
(DOJ, Child Pornography)
<http://www.webpronews.com/topnews/topnews/wpn-60-20060220GoogleSubpoenaWoesDouble.html>
- Yahoo 'helped Jail China writer'
<http://news.bbc.co.uk/1/hi/world/asia-pacific/4221538.stm>
- Phishing with Google Desktop
(remote code execution via malicious website)
http://www.theregister.co.uk/2005/12/03/google_desktop_vuln/

Assumptions

- In theory, terms of service offer us protection, in some instances Google may change the terms or be legally forced to circumvent them.
- Google never throws away information
- “Reasonable” amounts of information storage and processing power are available to Google
- We’ll focus on the technically feasible capabilities of Google

Zeitgeist (the spirit of the time)

Search patterns, trends, and surprises according to Google

“We should note that in compiling the Zeitgeist, no individual searcher's information is available or accessible to us. What you see here is a cumulative snapshot of interesting queries people are asking – some over time, some within country domains, and some on Google.com – that perhaps reveal a bit of the human condition. We appreciate the contribution all Google users make to these fascinating bits of information.”

Google Zeitgeist 2005

Google.com

Top Gainers of 2005

1. Myspace
2. Ares
3. Baidu
4. wikipedia
5. orkut
6. iTunes
7. Sky News
8. World of Warcraft
9. Green Day
10. Leonardo da Vinci

Google News

Top Searches in 2005

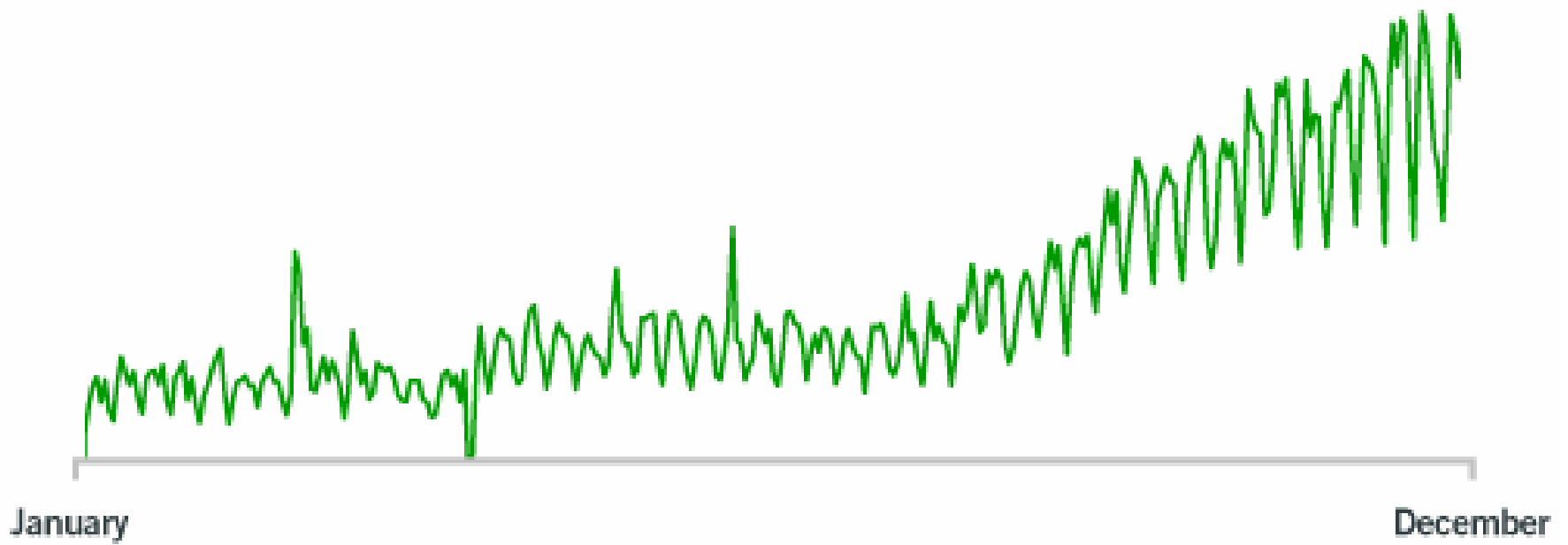
1. Janet Jackson
2. Hurricane Katrina
3. tsunami
4. xbox 360
5. Brad Pitt
6. Michael Jackson
7. American Idol
8. Britney Spears
9. Angelina Jolie
10. Harry Potter

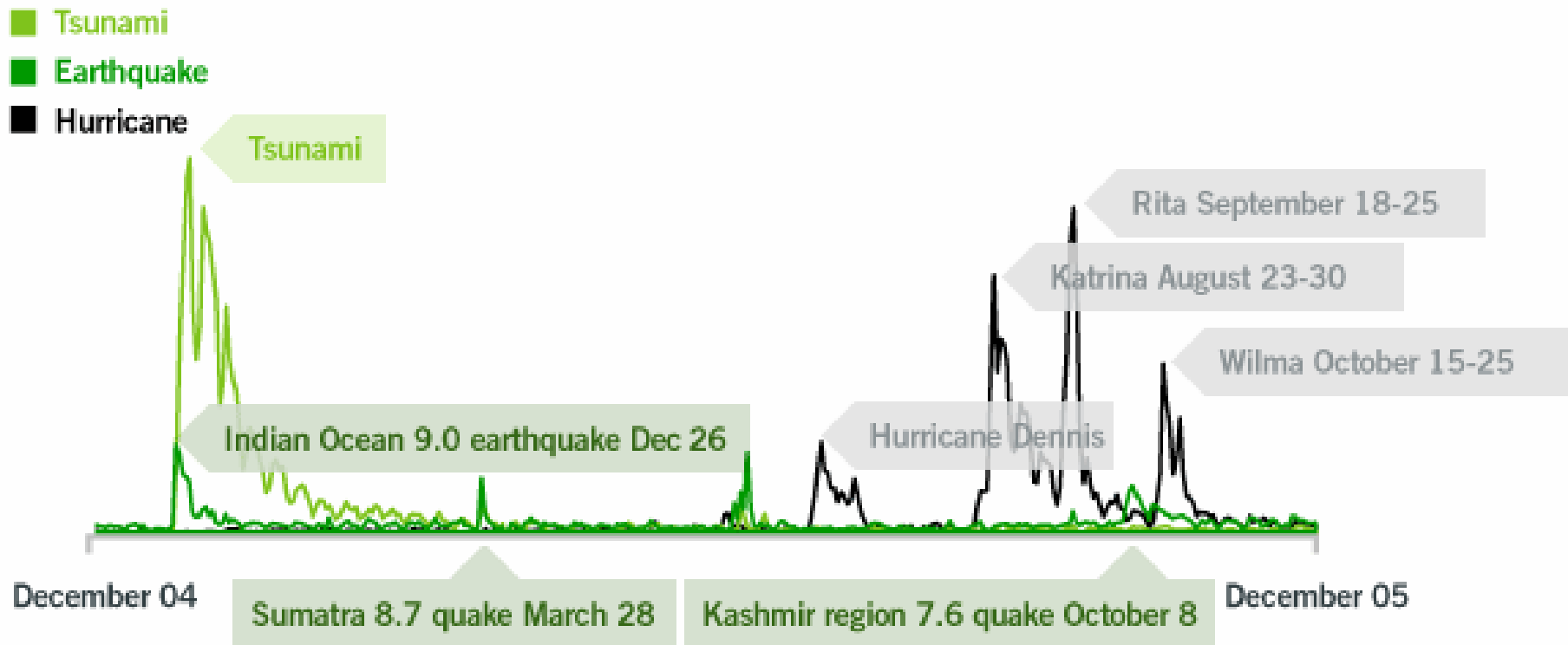
Froogle

Top Searches in 2005

1. ipod
2. digital camera
3. mp3 player
4. ipod mini
5. psp
6. laptop
7. xbox
8. ipod shuffle
9. computer desk
10. ipod nano

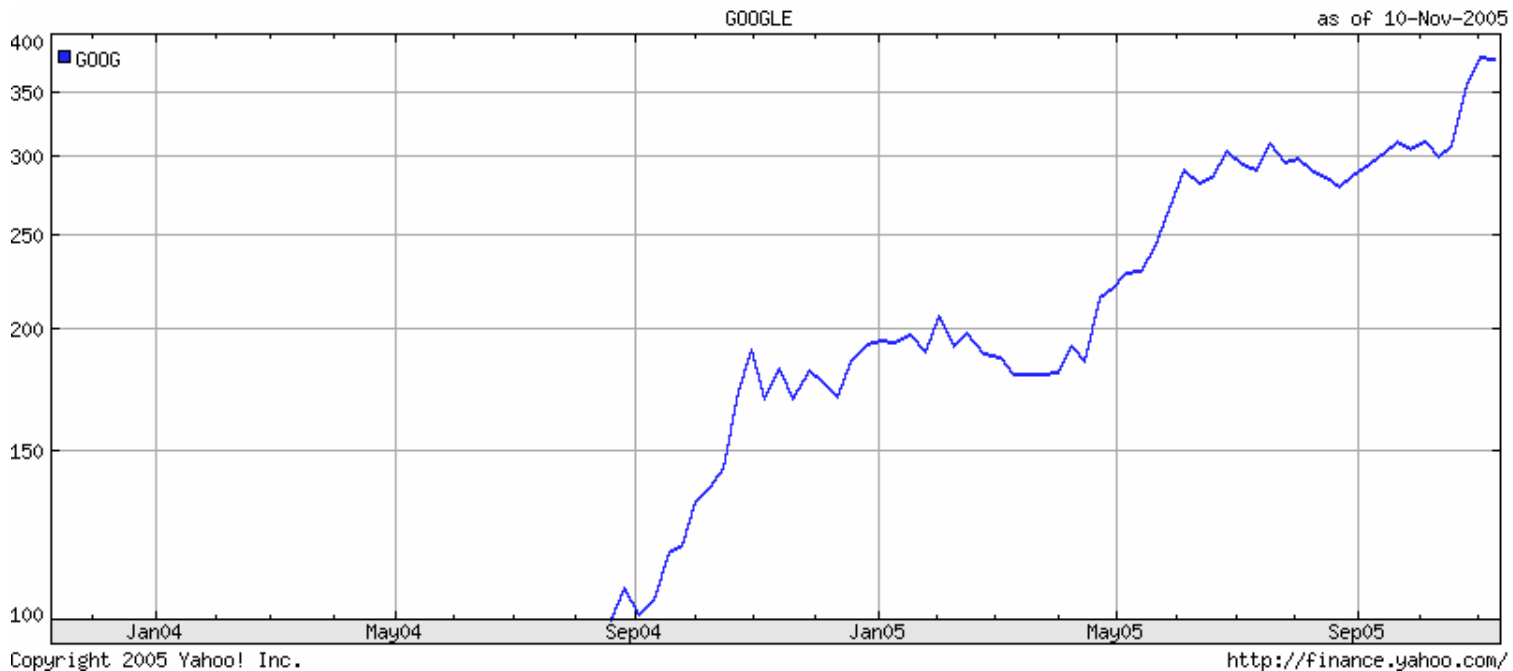
■ Wikipedia





What if...

“We should note that in compiling the Zeitgeist, each individual searcher’s information was available to us. What you see here is a precise snapshot of queries you’ve been asking some over time, some within country domains, and some on Google.com – that perhaps reveal a great deal about your public and private life. We appreciate the contribution all Google users make to these fascinating bits of information.”



Google's mission is to organize the world's information and make it universally accessible and useful.



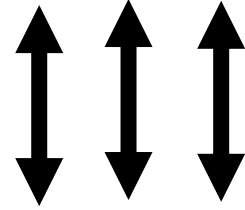
Evil



Google generates revenue by providing advertisers with the opportunity to deliver measurable, cost-effective online advertising

Good

Information others provide or retrieve

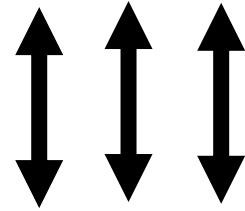


Information
you provide
or retrieve



Information Google retrieves
(think Googlebot)

Information others provide or retrieve

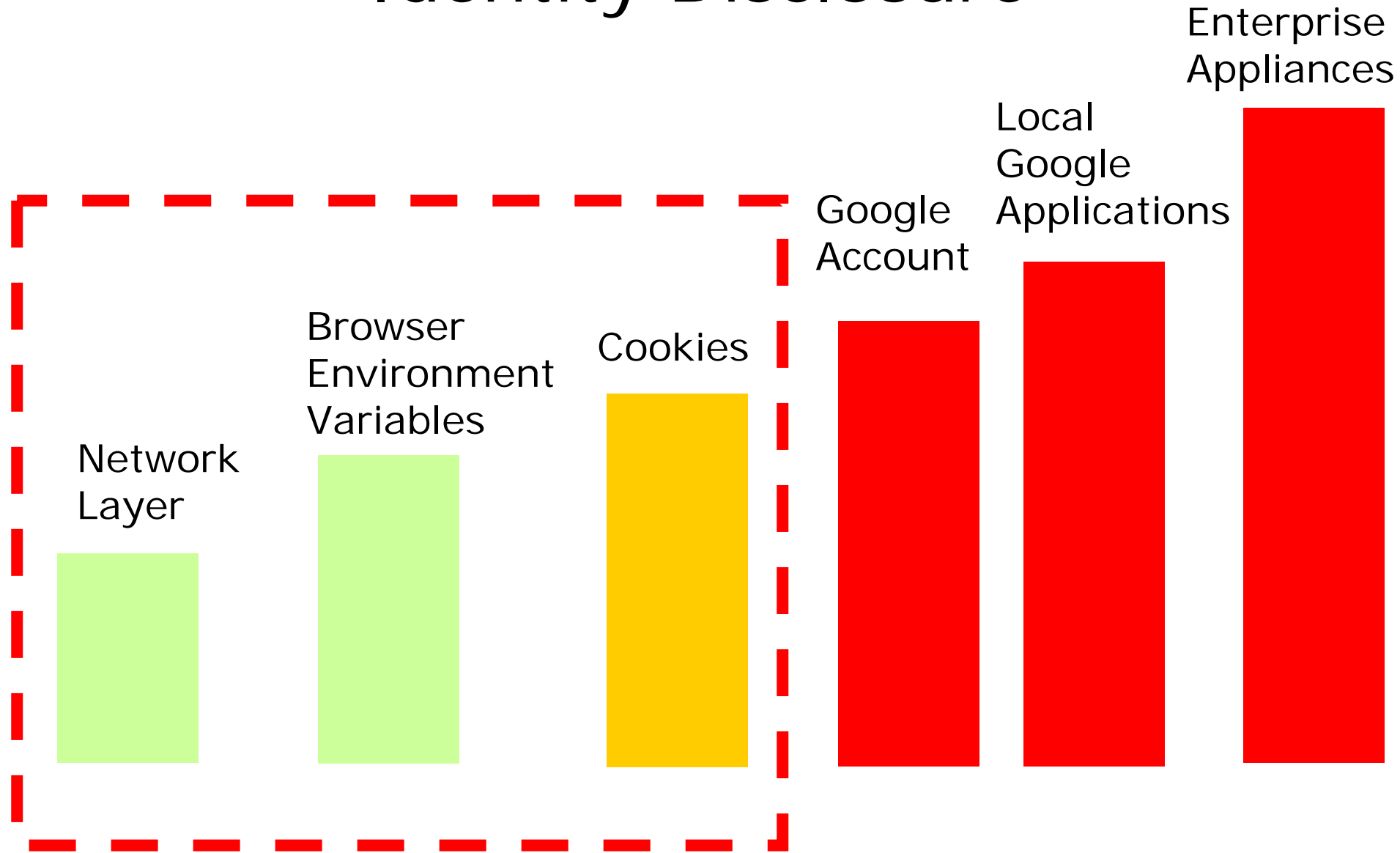


Information
you provide
or retrieve



Information Google retrieves
(think Googlebot)

Identity Disclosure



Google Services



[Alerts](#)

Receive news and search results via email



[Answers](#)

Ask a question, set a price, get an answer



[Blog Search](#)

Find blogs on your favorite topics



[Book Search](#)

Search the full text of books



[Catalogs](#)

Search and browse mail-order catalogs



[Directory](#)

Browse the web by topic



[Froogle](#)

Shop smarter with Google



[Groups](#)

Create mailing lists and discussion groups



[Images](#)

Search for images on the web



[Labs](#)

Try out new Google products



[Local](#)

Find local businesses and services



[Maps](#)

View maps and get directions



[Mobile](#)

Use Google on your mobile phone



[News](#)

Search thousands of news stories



[Scholar](#)

Search scholarly papers



[SMS](#)

Use text messaging for quick info



[Special Searches](#)

Search within specific topics



[University Search](#)

Search a specific school's website



[Web Search](#)

Search over billions of web pages



[Web Search Features](#)

Do more with search

Google Tools



[Blogger](#)

Express yourself online



[Code](#)

Download APIs and open source code



[Desktop](#)

Info when you want it, right on your desktop



[Earth](#)

Explore the world from your PC



[Gmail](#)

A Google approach to email



[Local for mobile](#)

View maps and get directions on your phone



[Picasa](#)

Find, edit and share your photos



[Talk](#)

IM and call your friends through your computer



[Toolbar](#)

Add a search box to your browser



[Translate](#)

View web pages in other languages

How much information do you give away? Can it be used to uniquely identify you?

Alerts

Maps

Answers

Mobile

Blog Search

News

Book Search

Scholar

Catalogs

SMS

Directory

Special Searches

Froogle

University Search

Groups

Video

Images

Web Search

Google Labs

Local

Blogger

Google Pack

Code

Picasa

Desktop Search

Google Talk

Google Earth

Toolbar

Gmail

Translate Tool

Local for mobile

Linking Information Disclosure w/Identity

- Do you care if Google is able to tie your online activities to your identity?
- Can “anonymous” browser-based Google interaction be uniquely identified to you?
 - If so, how long / how much data will it take?
- To what extent can Google tie other Google users’ activities to your own?
 - And how much does the information garnered from the activity of others provide about you?
- Can the you be identified if you change IP addresses?
- Are today’s online privacy technologies and best practices enough to prevent fingerprinting?
 - If not, how much do they slow down the process?
- Can your identity be tied to other online data (e.g. websites)?

Google Talent

“Passionate about these topics? You should work at Google.”

- algorithms
- artificial intelligence
- compiler optimization
- computer architecture
- computer graphics
- data compression
- **data mining**
- file system design
- **genetic algorithms**
- **information retrieval**
- **machine learning**
- natural language processing
- operating systems
- **profiling**
- robotics
- text processing
- user interface design
- web information retrieval
- and more!

Uniquely Identifying Users

- I'm Feeling Lucky
- Words you can't spell
- Words you don't know the meaning of
- Google as address book
- Is Google your home page?



Google Search Statistics

- mork file format
(firefox)
- formhistory.dat
- all form data
- ~12 months usage
- repeated use not
known
- >1962 entries

Google Research...

google

google api

google autolink

google base

google blacklist

google chat

google contest

google hacked

google intervened

google isp

google keyhole

google labs

google maps

google scholar

google screen

google sets

google toolbar

google wallet

google watch

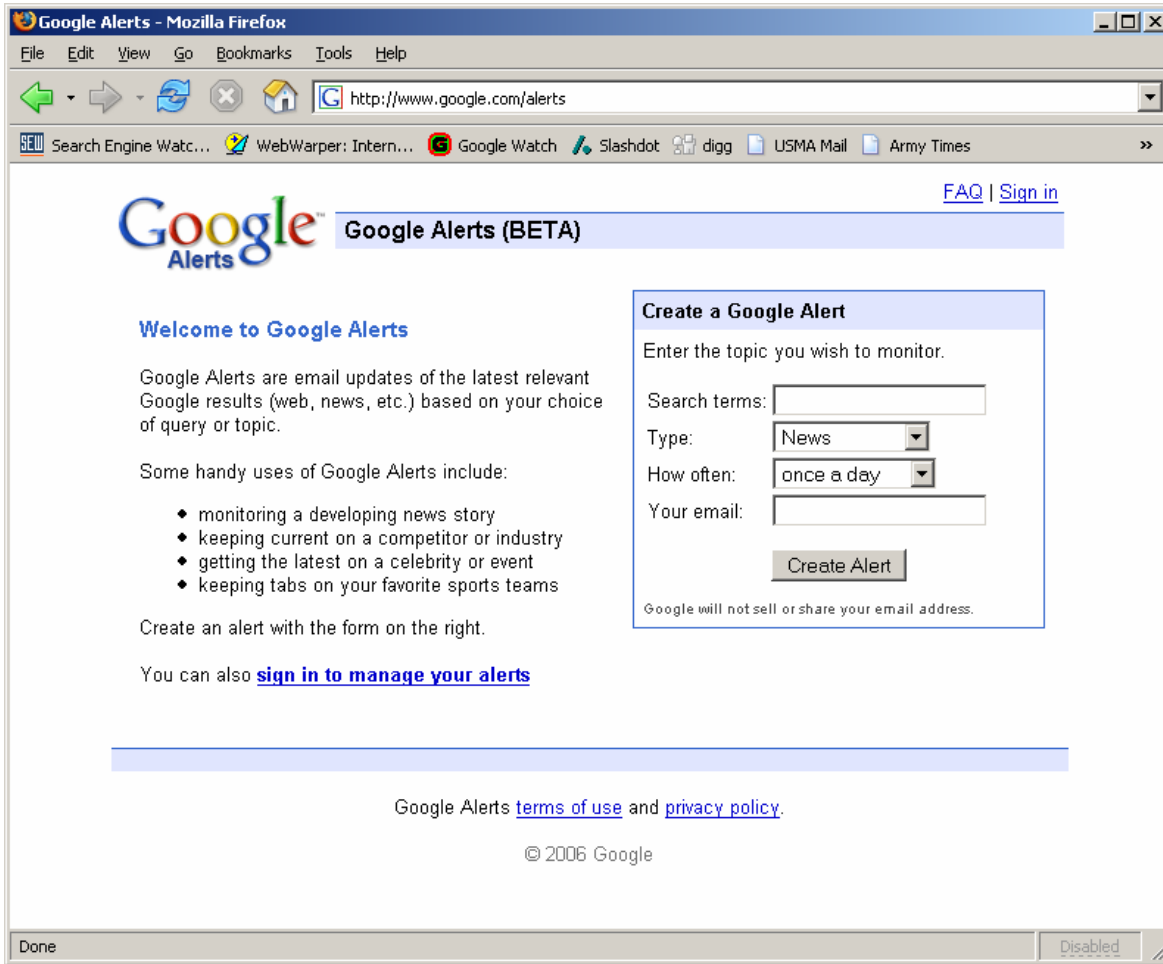
google wireless

googlebot

Quoted Searches...

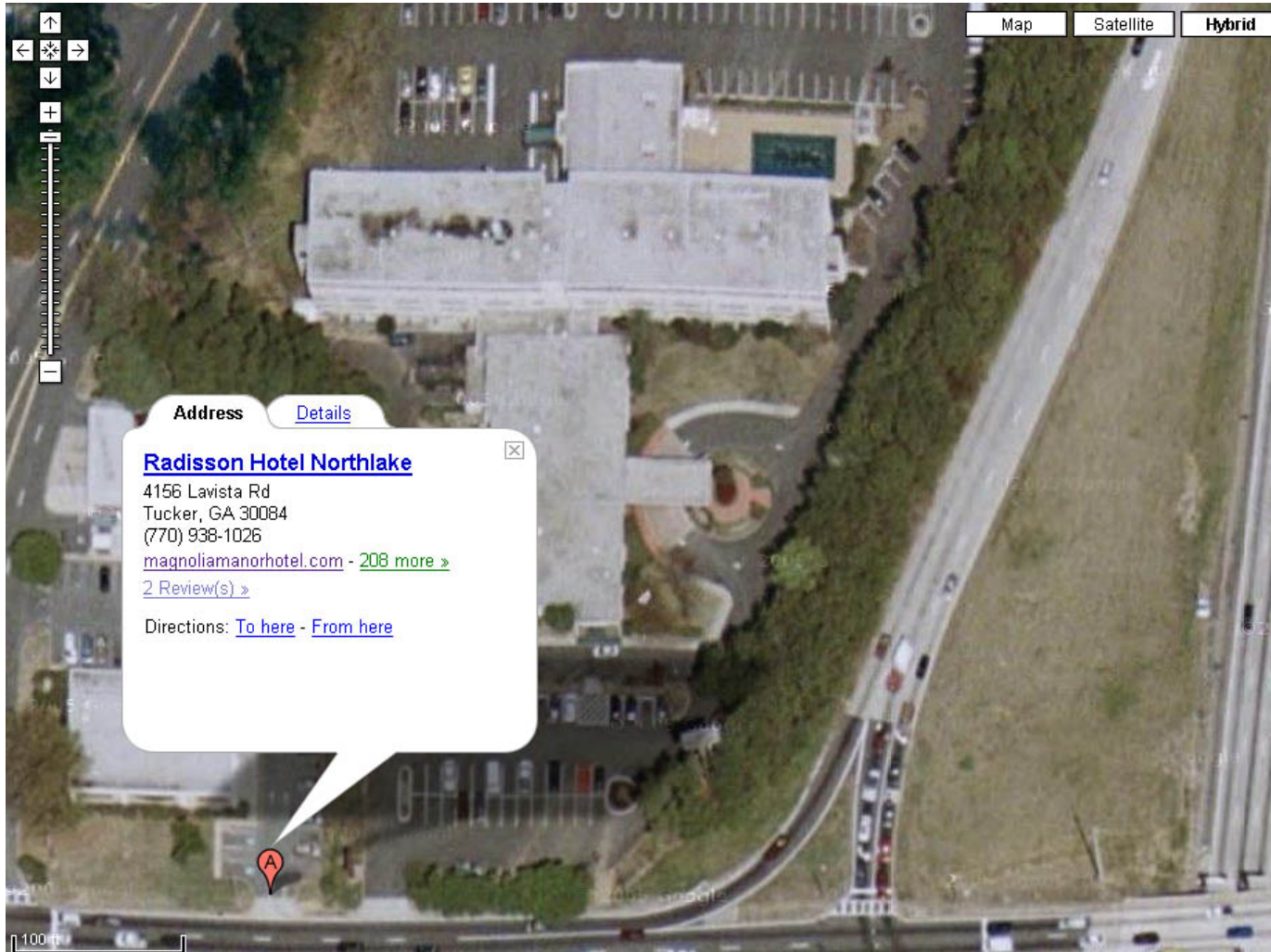
"informative art"	<name>
"installing pcap"	"kp ddk"
"installing pcap:"	<name>
<name>	<name>
<name>	"luddite t-shirt"
<name>	"man elf"
<name>	"man elf" file
<name>	"man elf" linux
<name>	"man elf(\"
<name>	<name>
<name>	<name>
<name> usma	<name>

Google Alerts



- search terms
- email address
- frequency
- category (news, web, news and web, groups)

Google Maps / Satellite Imagery



Locations
of interest
to you
(down to street
level)

What
have you
looked at?

Google Earth

(desktop application)

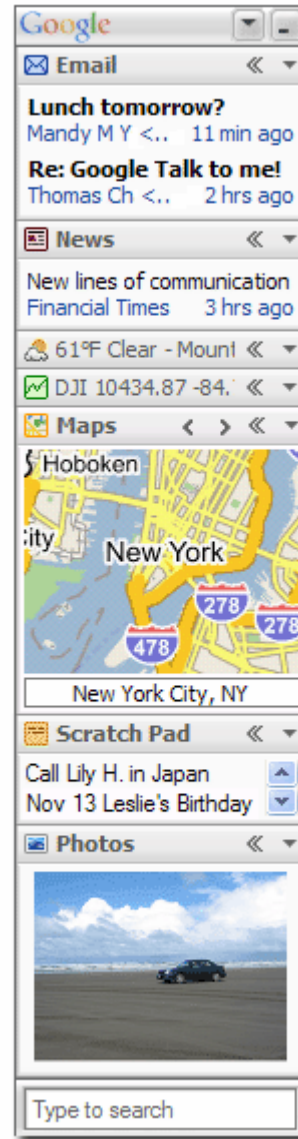


- measure distances
- fly to an address
- print save and email
- driving directions
- search
- save search results

Google Desktop

"Info when you want it, right on your desktop"

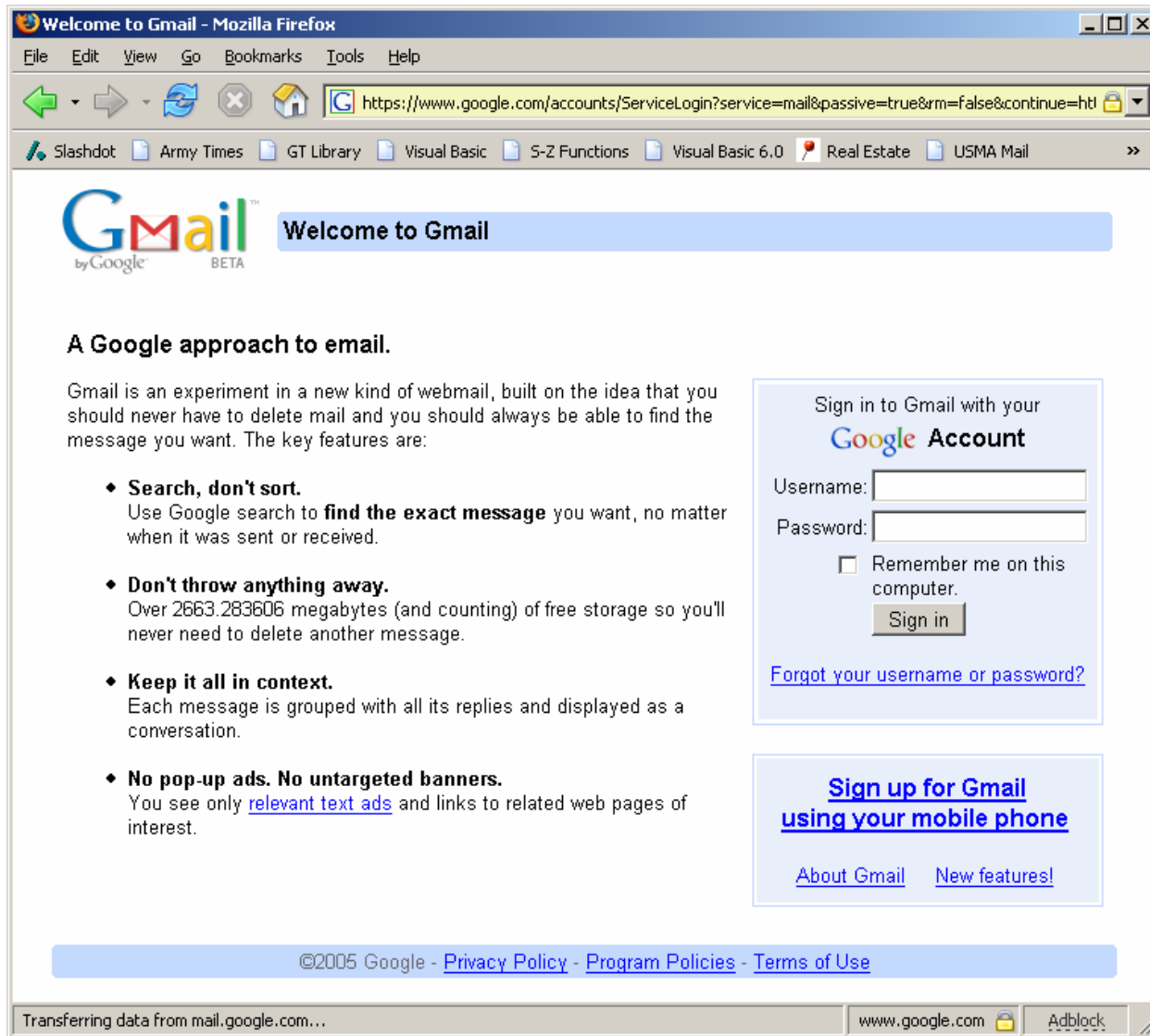
- "Get all your personalized info in one place with Sidebar"
- "Launch applications and search instantly with Quick Find"
- "Find all your email, files, photos, web history, Gmail, and more"
- "Search conveniently in Outlook with the Outlook Toolbar"



Plug-ins

- **gdMoney**: Monitor your credit card and bank transactions in real time.
- **Quick Lock plug-in**
- **Google Desktop Search** for Pocket PC Handhelds and Smartphones
- **Könguló web spider** crawls websites you specify, e.g. intranet websites
- **Laplink Everywhere** search remote PCs

Gmail



- email recipients
- email sources
- content
- N-order contact

Gmail

Welcome to Gmail - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://www.google.com/accounts/ServiceLogin?service=mail&passive=true&rm=false&continue=htl

Slashdot Army Times GT Library Visual Basic S-Z Functions Visual Basic 6.0 Real Estate USMA Mail

Gmail
by Google BETA

Welcome to Gmail

A Google approach to email.

Gmail is an experiment in a new kind of webmail, built on the idea that you should never have to delete mail and you should always be able to find the message you want. The key features are:

- ◆ **Search, don't sort.**
Use Google search to **find the exact message** you want, no matter when it was sent or received.
- ◆ **Don't throw anything away.**
Over 2663.283606 megabytes (and counting) of free storage so you'll never need to delete another message.
- ◆ **Keep it all in context.**
Each message is grouped with all its replies and displayed as a conversation.
- ◆ **No pop-up ads. No untargeted banners.**
You see only [relevant text ads](#) and links to related web pages of interest.

Sign in to Gmail with your
Google Account

Username:

Password:

Remember me on this computer.

[Forgot your username or password?](#)

[Sign up for Gmail using your mobile phone](#)

[About Gmail](#) [New features!](#)

©2005 Google - [Privacy Policy](#) - [Program Policies](#) - [Terms of Use](#)

Transferring data from mail.google.com... www.google.com Adblock

- email recipients
- email sources
- content
- N-order contact

Gmail

The screenshot shows the Gmail login page in a Mozilla Firefox browser window. The browser's address bar displays the URL: <https://www.google.com/accounts/ServiceLogin?service=mail&passive=true&rm=false&continue=htl>. The browser's menu bar includes File, Edit, View, Go, Bookmarks, Tools, and Help. The browser's toolbar shows various icons, including a home button and a search button. The browser's status bar at the bottom indicates "Transferring data from mail.google.com..." and shows the address "www.google.com" and an "Adblock" extension.

Welcome to Gmail - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://www.google.com/accounts/ServiceLogin?service=mail&passive=true&rm=false&continue=htl

Slashdot Army Times GT Library Visual Basic S-Z Functions Visual Basic 6.0 Real Estate USMA Mail

Gmail
by Google BETA

Welcome to Gmail

A Google approach to email.

Gmail is an experiment in a new kind of webmail, built on the idea that you should never have to delete mail and you should always be able to find the message you want. The key features are:

- **Search, don't sort.**
Use Google search to **find the exact message** you want, no matter when it was sent or received.
- **Don't throw anything away.**
Over 2663.283606 megabytes (and counting) of free storage so you'll never need to delete another message.
- **Keep it all in context.**
Each message is grouped with all its replies and displayed as a conversation.
- **No pop-up ads. No untargeted banners.**
You see only [relevant text ads](#) and links to related web pages of interest.

Sign in to Gmail with your
Google Account

Username:

Password:

Remember me on this computer.

[Forgot your username or password?](#)

[Sign up for Gmail using your mobile phone](#)

[About Gmail](#) [New features!](#)

©2005 Google - [Privacy Policy](#) - [Program Policies](#) - [Terms of Use](#)

Transferring data from mail.google.com... www.google.com Adblock

- email recipients
- email sources
- content
- N-order contact

Google Calendar (CL2)

(unconfirmed)

CL2

Create Event

Back Add

February 2006

Su	Mo	Tu	We	Th	F	Sa
22	23	24	25	26	27	28
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	1	2	3	4
5	6	7	8	9	10	11

Calendars

My Calendars

Other Calendars

Search public calendars

[Manage Calendars](#)

Search My Calendars [View Search Details](#)

[Back to Calendar](#)

What

When to All Day

Repeats:

This event does not repeat

Where

Calendar

Description

Options

Reminder:

Show me as: Available Busy

Privacy: This event is:

- Default
- Private
- Public

[View Event Details](#)

[Back to Calendar](#)

Guests

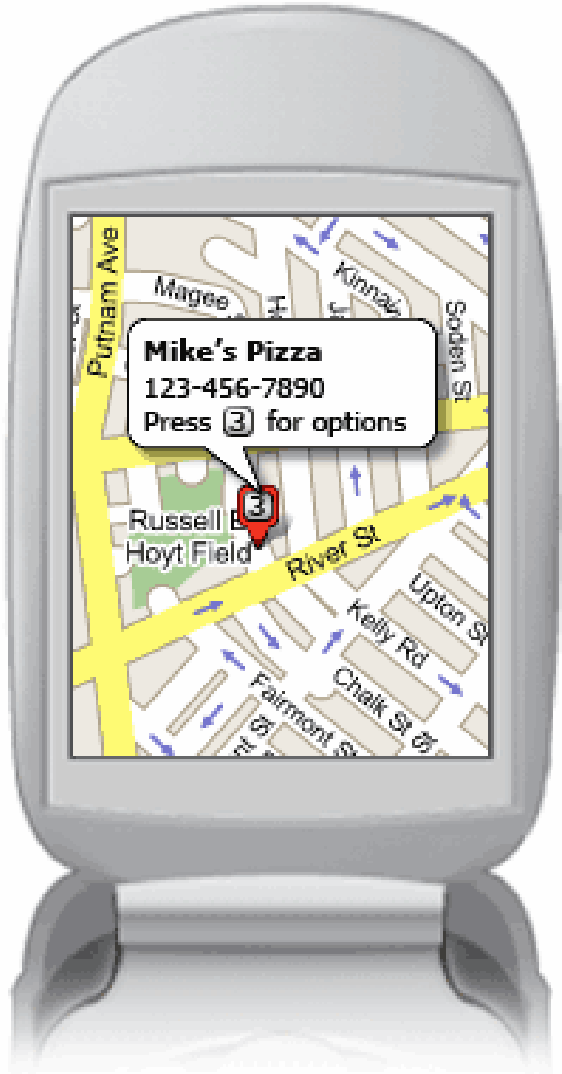
Add guests

Enter the email addresses of guests, separated by commas

Guests can:

- invite others
- leave comments
- see guest list

Local Information for Local Devices



- Local for Mobile
 - detailed directions
 - search results integrated with map
 - zoom in/out, drag maps
 - satellite imagery
- free download

Search Appliances

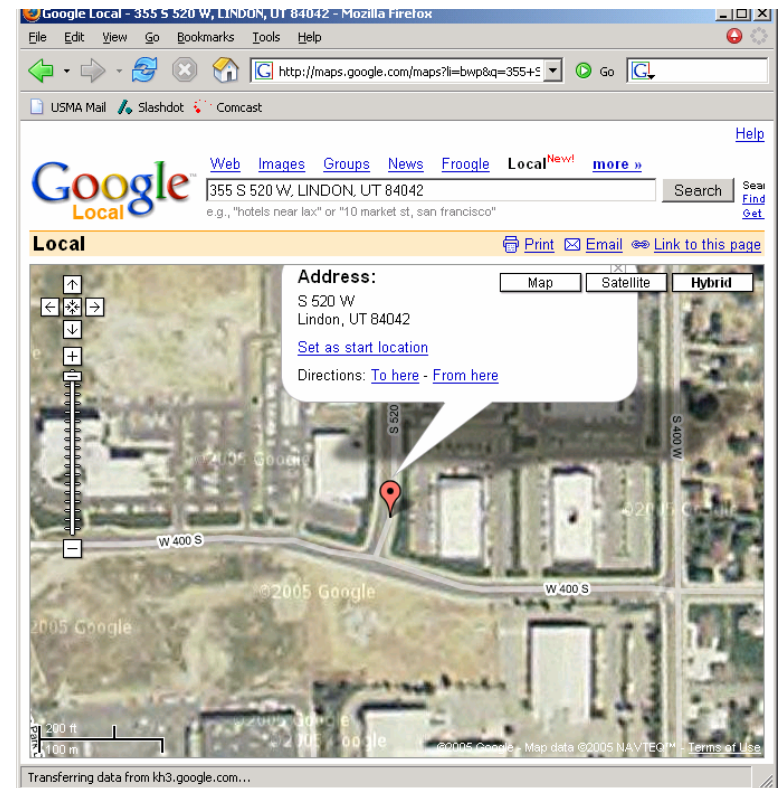
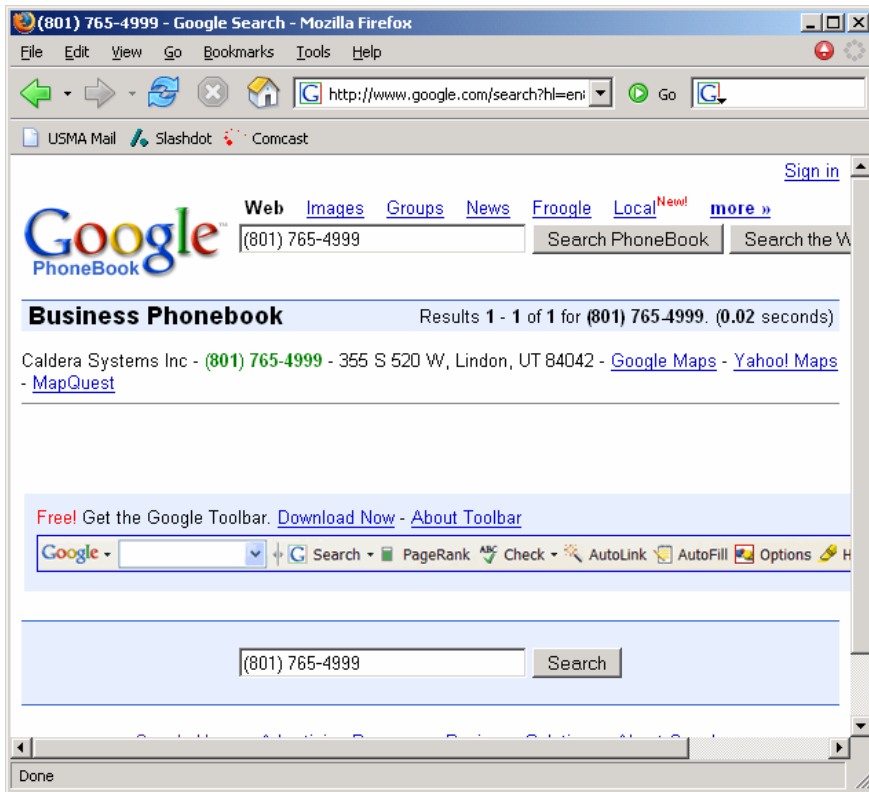


The Google Search Appliance makes the sea of lost data on your web servers, file systems and relational databases instantly available with one mouse click.

Just point it toward your content, add a search box to your site, and in a matter of hours, your users will be able to search through more than 220 different file formats in any language.

The Google Search Appliance indexes up to 15 million documents, and its security features ensure that users only see the documents to which they have proper access.

Residential/Commercial Phone Number Lookup



a search returns...

- name
- address
- phone number
- links to 3 mapping services

Travel Support

lax nyc - Google Search - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.google.com/search?q=lax+nyc

Slashdot Army Times GT Library Visual Basic S-2 Functions Visual Basic 6.0 Real Estate

Sign in

Google Web Images Groups News Froogle Local^{New!} more »


lax nyc Search Advanced Search Preferences

Web Results 1 - 10 of about 4,160,000 for **lax nyc**. (0.12 seconds)

Flights from Los Angeles, CA to New York, NY

Departing: Returning:
Search: [Expedia](#) [Hotwire](#) [Orbitz](#) [Priceline](#)

Local results for lax near New York, NY

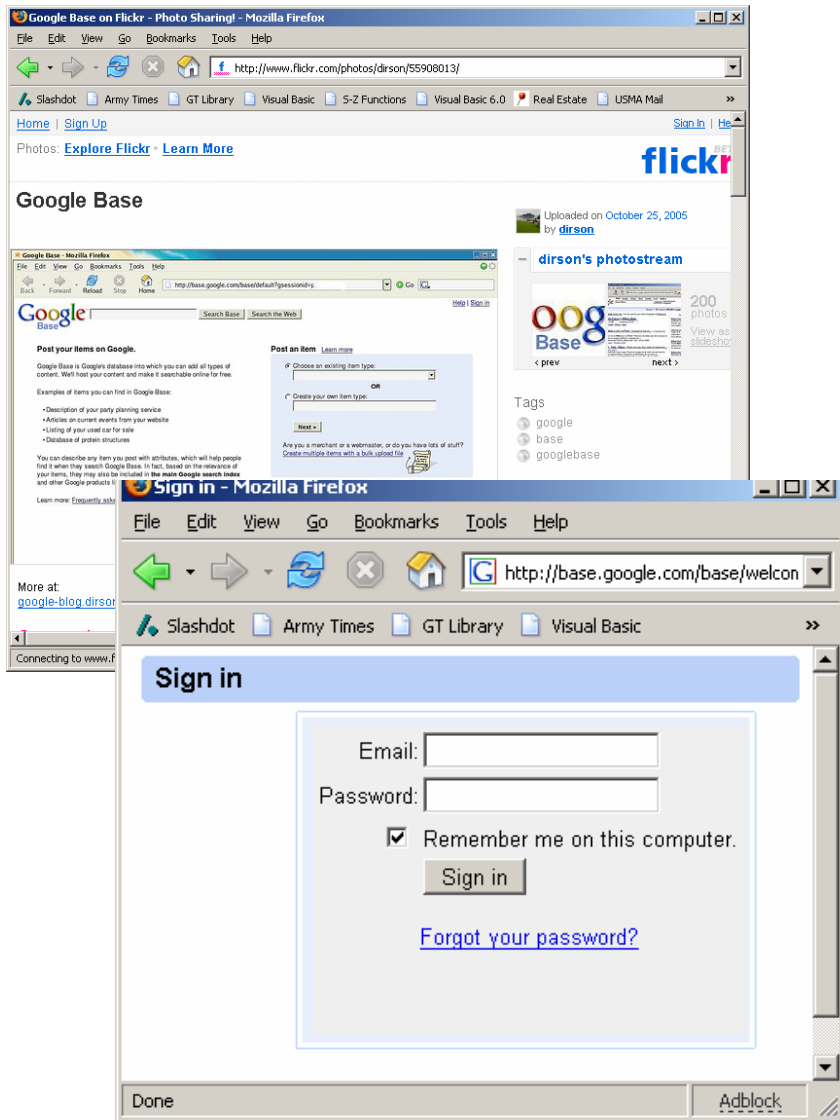
-  [Robert I Lax & Assoc](#) - 3.1 miles NE - 535 5th Ave # 21, New York, 10017 - (212) 818-9150
- [Lax Martin](#) - 3.2 miles N - 1185 Avenue of the Americ, New York, 10036 - (212) 372-1208
- [Lax Moshe](#) - 3.4 miles NE - 250 E 49th St, New York, 10017 - (212) 758-9840

LAX NYC Cheap Airfare
cheap airfare to **NYC** from **LAX** to New York City from Los Angeles.
[playground.igougo.com/Cities/FareTrend_NYCLAX.asp - 14k - Cached - Similar pages](#)

WDPCC | Teams Contacts
Orca Lacrosse **NYC** New York, NY Philadelphia Premier Philadelphia, PA Philly Women's
Lax Philadelphia, PA Portland Women Portland, OR

Done Adblock

Google Base



- A “database into which you can add all types of content. We’ll host your content and make it searchable online for free”
- “A free classified advertising system” (NYT)
- Examples
 - Description of your party planning service
 - Articles on current events from you website
 - Listing of your used car for sale
 - Database of protein structures
- Hard to tell exactly what it is
- Going live date is TBD

Collaborative Word Processing



writely (beta)

The Web Word Processor

...that's [now part of Google](#).

- **Share documents instantly & collaborate real-time.**
Pick exactly who can access your documents.
- **Edit your documents from anywhere.**
Nothing to download -- your browser is all you need.
- **Store your documents securely online.**
Offsite storage plus data backup every 10 seconds.
- **Easy to use.**
Clean, uncluttered screens with a familiar, desktop feel.

Already registered with Writely?
[Sign in here.](#)

Let me know when I can sign up!

E-mail:

We have closed off new registrations until we move Writely to Google's systems.

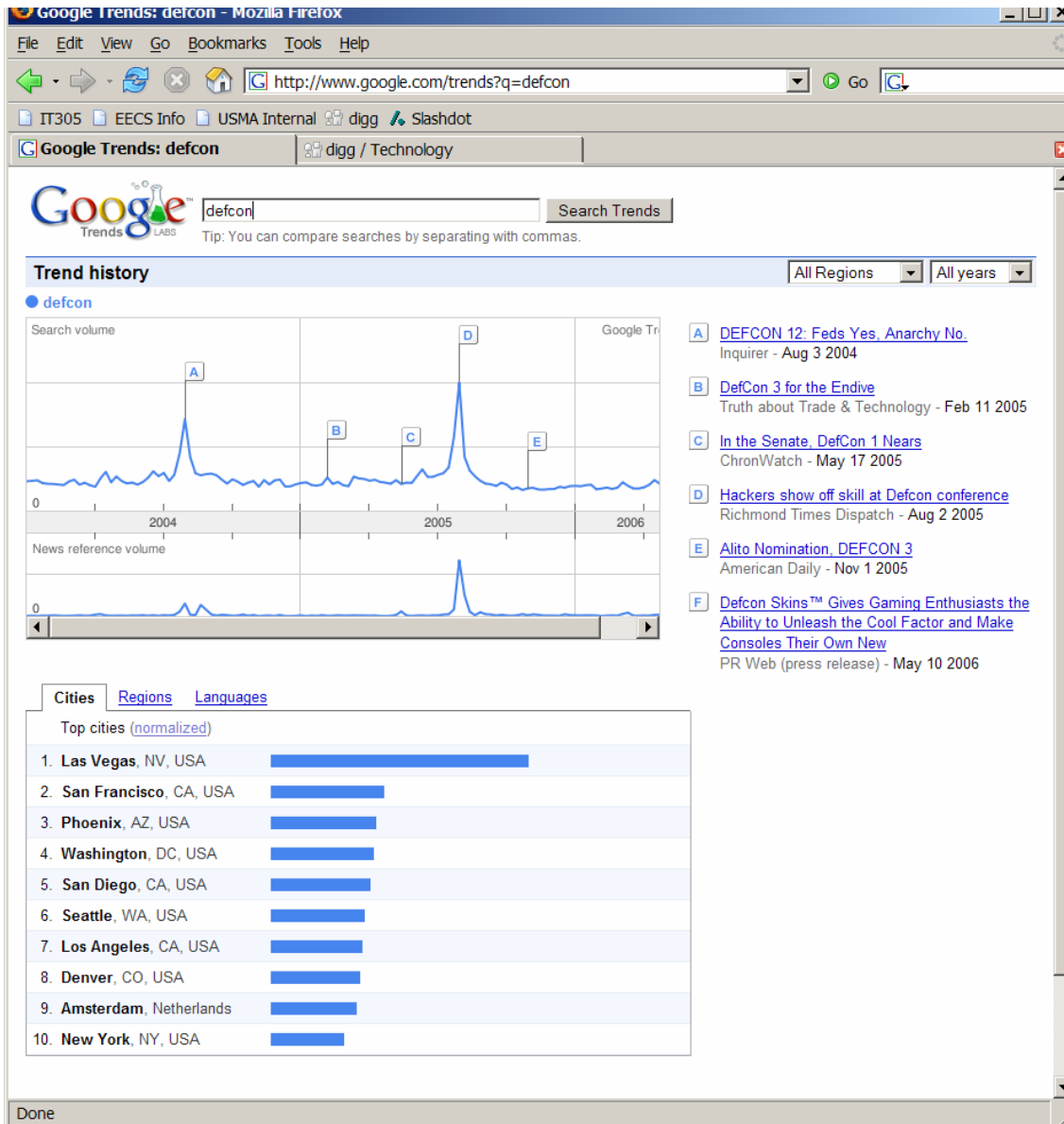
Please submit your e-mail address and we'll send you an email when registration opens back up!

➔ [Take a tour](#) ➔ [Find out more](#)

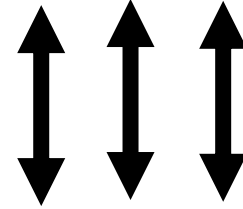
Can I ... upload from Word? **YES!**
Can I ... save to my desktop? **YES!**
Can I ... publish on the web? **YES!**
Can I ... post on my blog? **YES!**

Done Disabled

Google Trends



Information others provide or retrieve

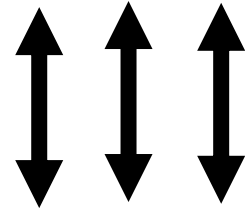


Information
you provide
or retrieve



Information Google retrieves
(think Googlebot)

Information others provide or retrieve



Information
you provide
or retrieve



Information Google retrieves
(think Googlebot)

Key Phrases

(Google.com 90 days)

<name>

<school>

<state>

<system>

<article> .. <article N>

<collaborator> .. <collaborator N>

<organization> .. <organization N>

<research area> .. <research area N>

<magazine> .. <magazine N>

<media type> .. <media type N>

Key Phrases

(Google.com 90 days)

<article>+<collaborator>

<name> + <school>

<name> + <state>

<name> + <system>

<name> + <collaborator>

<school> + <research area>

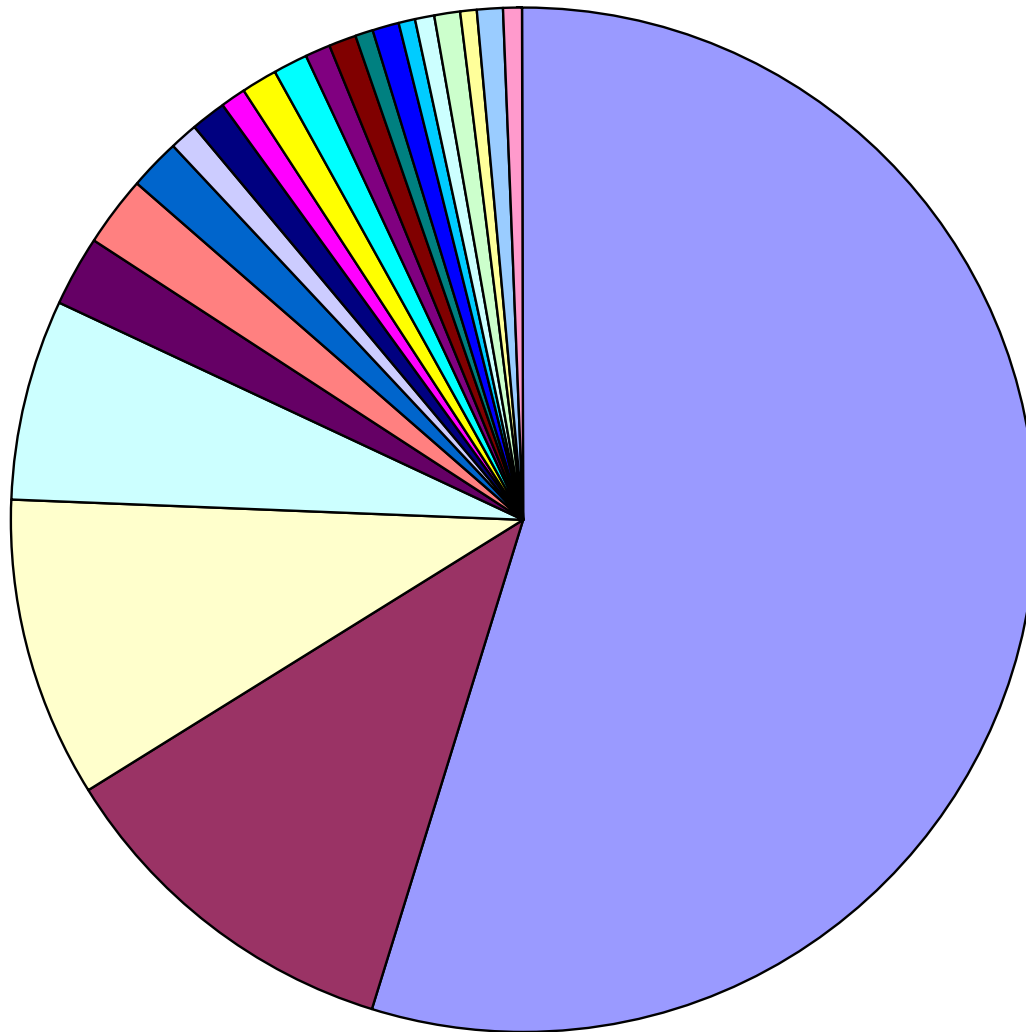
<school> + <media type>

<research area> + <magazine>

<research area> + <media type>

<research area> + <collaborator> + <magazine>

Referring Domains...



- www.astalavista.com
- www.eviloctal.com
- www.google.com
- astalavista.com
- www.honeynet.org
- www.jemput.com
- www.blackhat.com
- 64.233.179.104
- images.google.com
- search.yahoo.com
- userpages.umbc.edu
- www.google.co.in
- www.gregconti.com
- project.honeynet.org
- www-static.cc.gatech.edu
- www.findone.net
- www.google.de
- www.google.fr
- www.google.nl
- www.pricelesswarehome.org
- www.ringolab.com

I get a feeling
Microsoft's
monopoly will look
like a minor
bother compared
to Google's
omnipresence one
day.

-from Slashdot



Countermeasures

- Anonymous browsing (Tor, anonymizer...)
- Diversity
- Policy/Law
- Go directly to the website you want
- Anonymizer-ish application for individuals and organizations but operates higher up the semantic scale
- Aggregators
- Encrypt content
- Switching Proxy Plug-ins
- Cookie rewriters

Summary

- Information portals possess the capability to create massive strategic, operational and tactical intelligence stockpiles
- A **tremendous** amount of trust is being placed in Google
- Will this trust come back to haunt us?

Discussion

Greg Conti

gregory-conti@usma.edu

<http://www.rumint.org/gregconti/index.html>