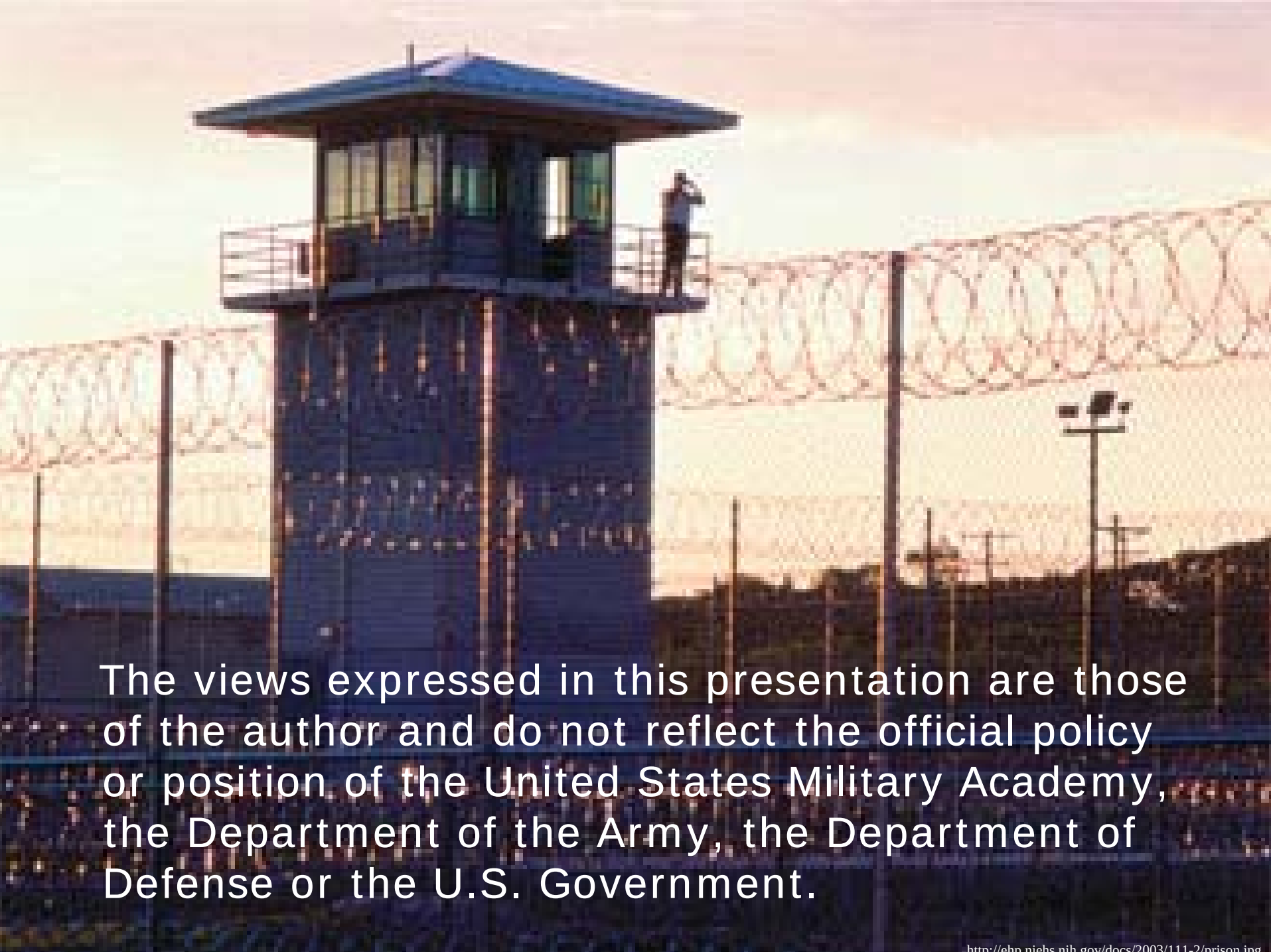
Googling

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.
http://ehp.niehs.nih.gov/docs/2003/111-2/prison.jpg

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."

http://www.google.com/press/zeitgeist.htm

Google Zeitgeist 2005

Google.com
Top Gainers of 2005

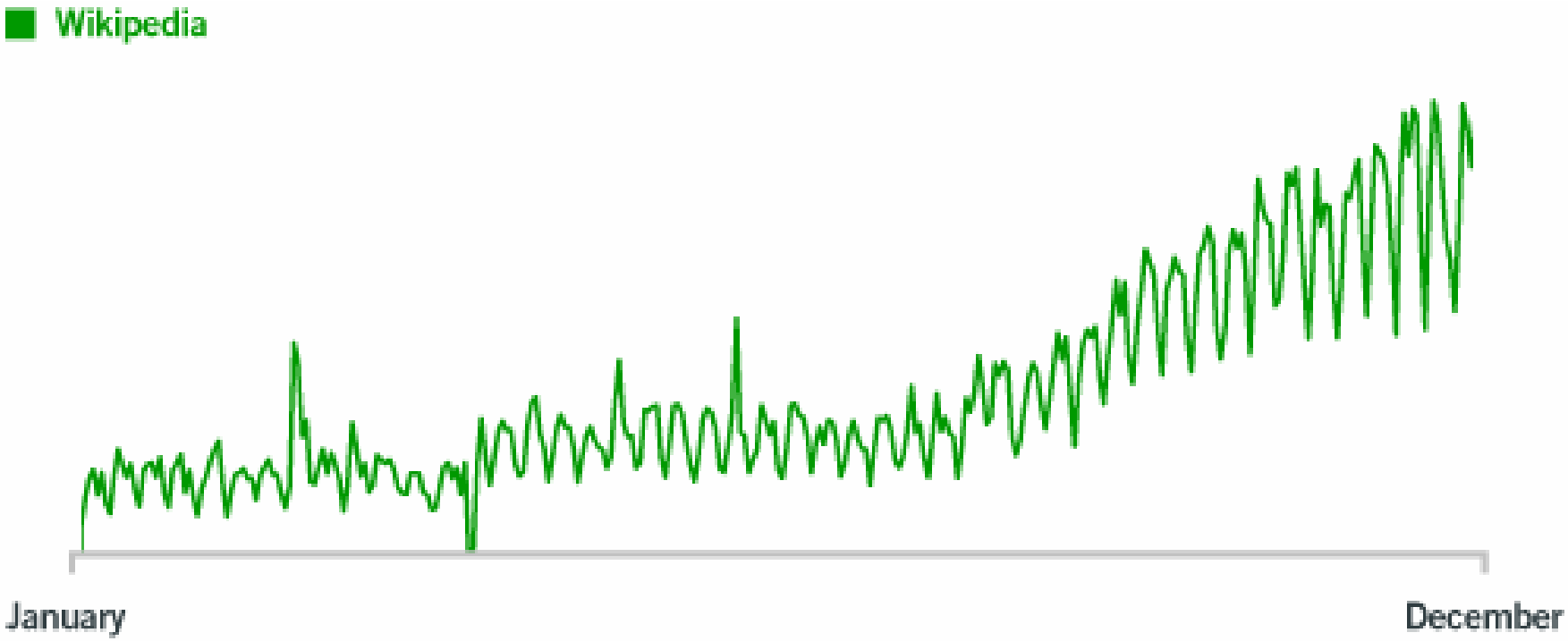
1. Myspace
2. Ares
3. Baidu
4. wikipedia
5. orkut
6. iTunes
7. Sky News
8. World of Warcraft
9. Green Day
10. Leonardo da Vinci

Google News
Top Searches in 2005

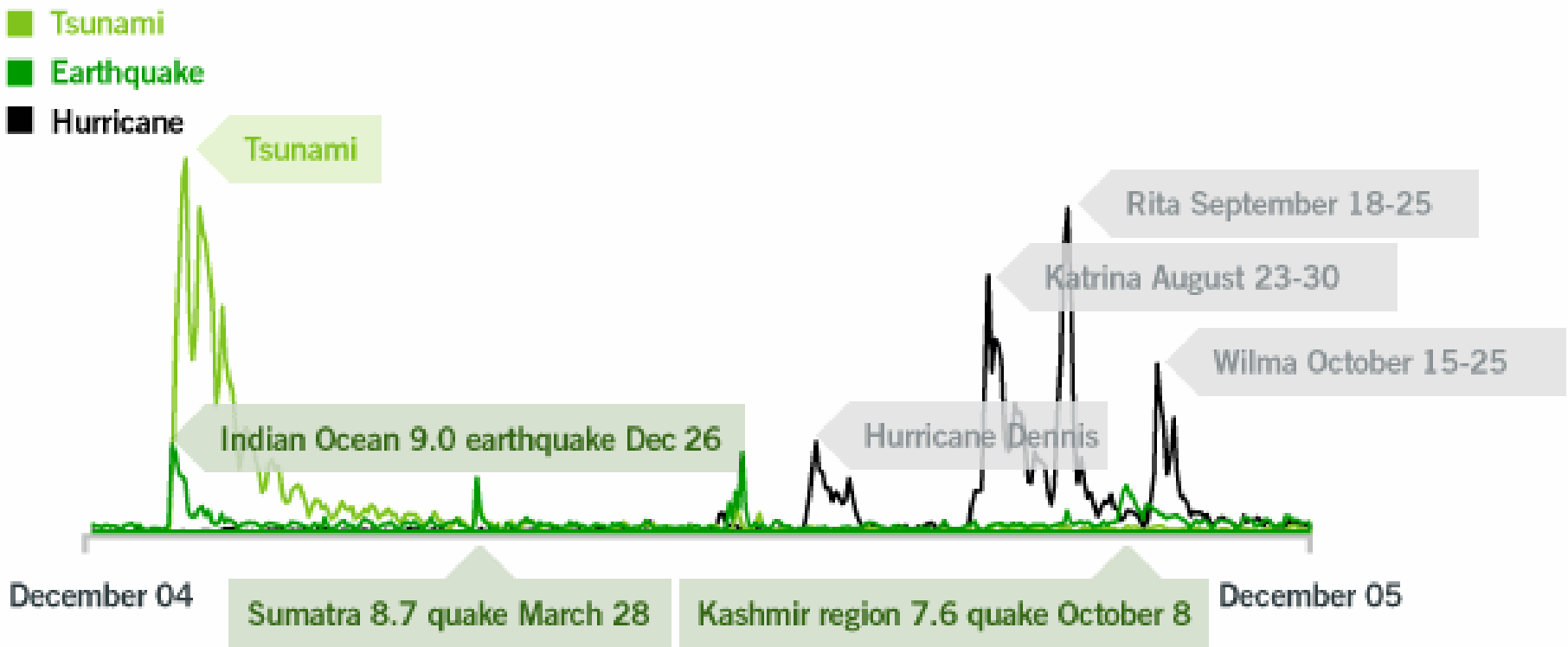
1. Janet Jackson
2. Hurricane Katrina
3. tsunami
4. xbox 360
5. Brad Pitt
6. Michael Jackson
7. American Idol
8. Britney Spears
9. Angelina Jolie
10. Harry Potter

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



http://www.google.com/press/zeitgeist2005/nature.html

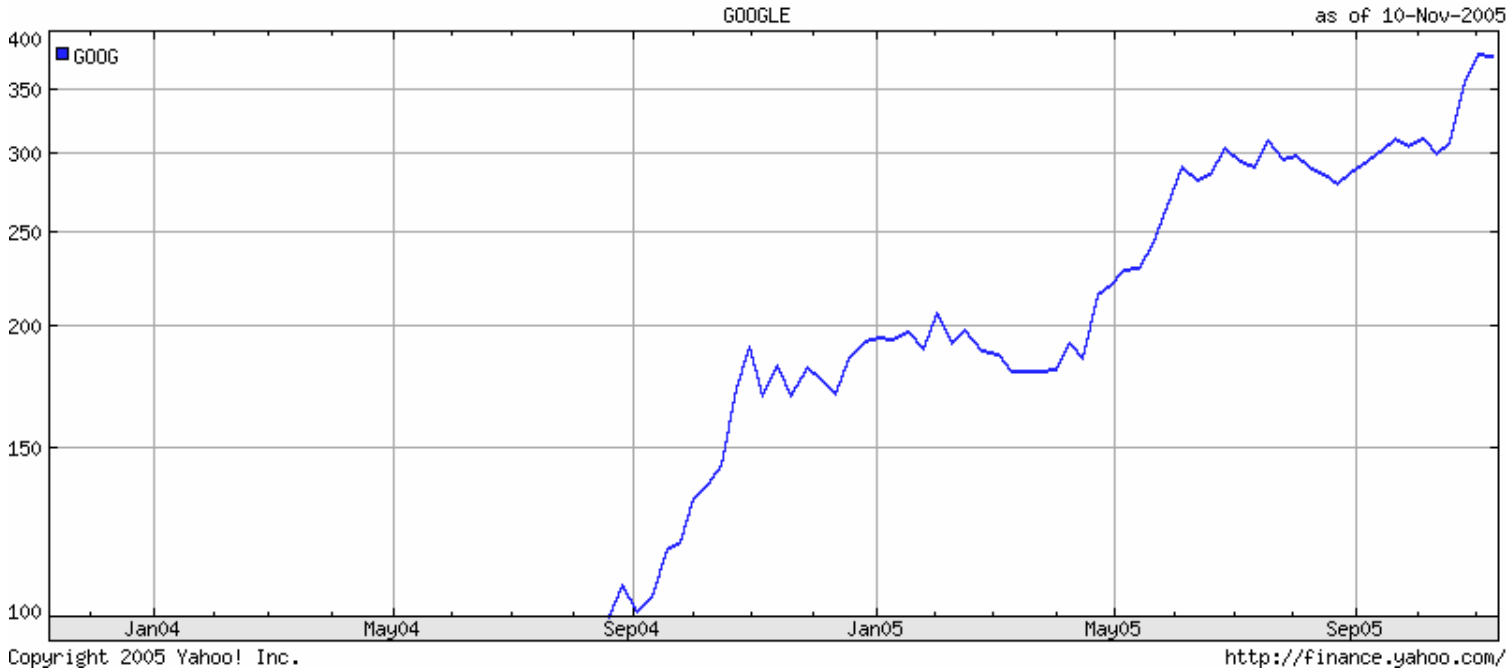


http://www.google.com/press/zeitgeist2005/nature.html

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."

This is entirely fictitious.



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

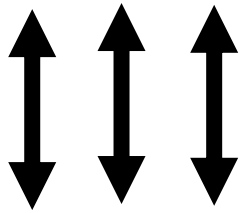
Evil

Good

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

http://www.nytimes.com/2005/10/30/business/yourmoney/30google.html?ex=1288324800&en=b0684c6ec54b2467&ei=5090&partner=rssuserland&emc=rss&pagewanted=all

Information others provide or retrieve

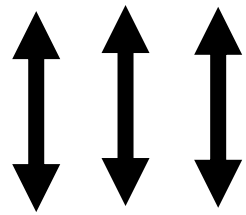


Information
you provide
or retrieve

Google™

Information Google retrieves
(think Googlebot)

Information others provide or retrieve

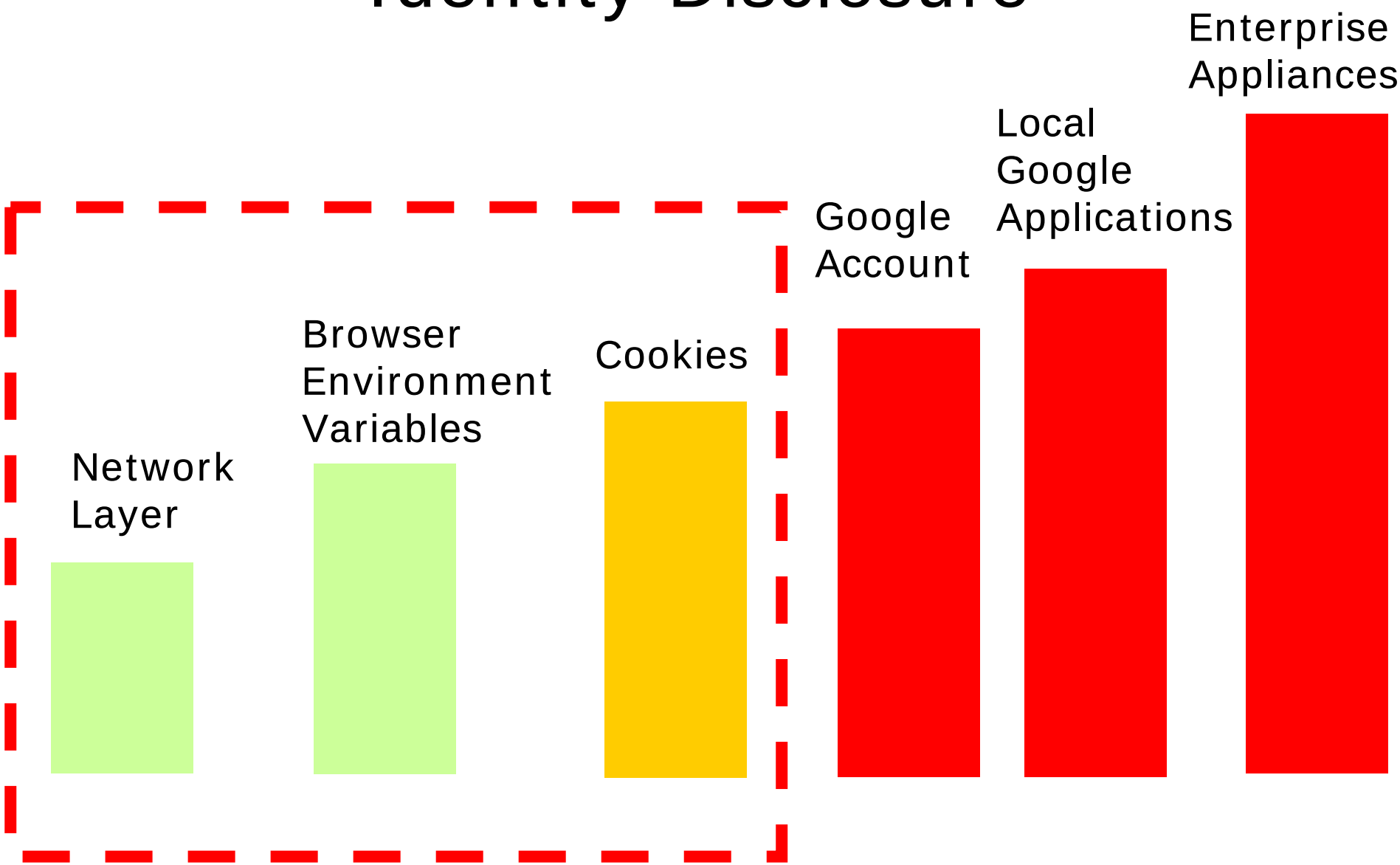


Information you provide or retrieve

Google™

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!

http://labs.google.com/

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?

http://www.factsfinder.com/Fingerprint.jpg

Google Search Statistics

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

Google Research...

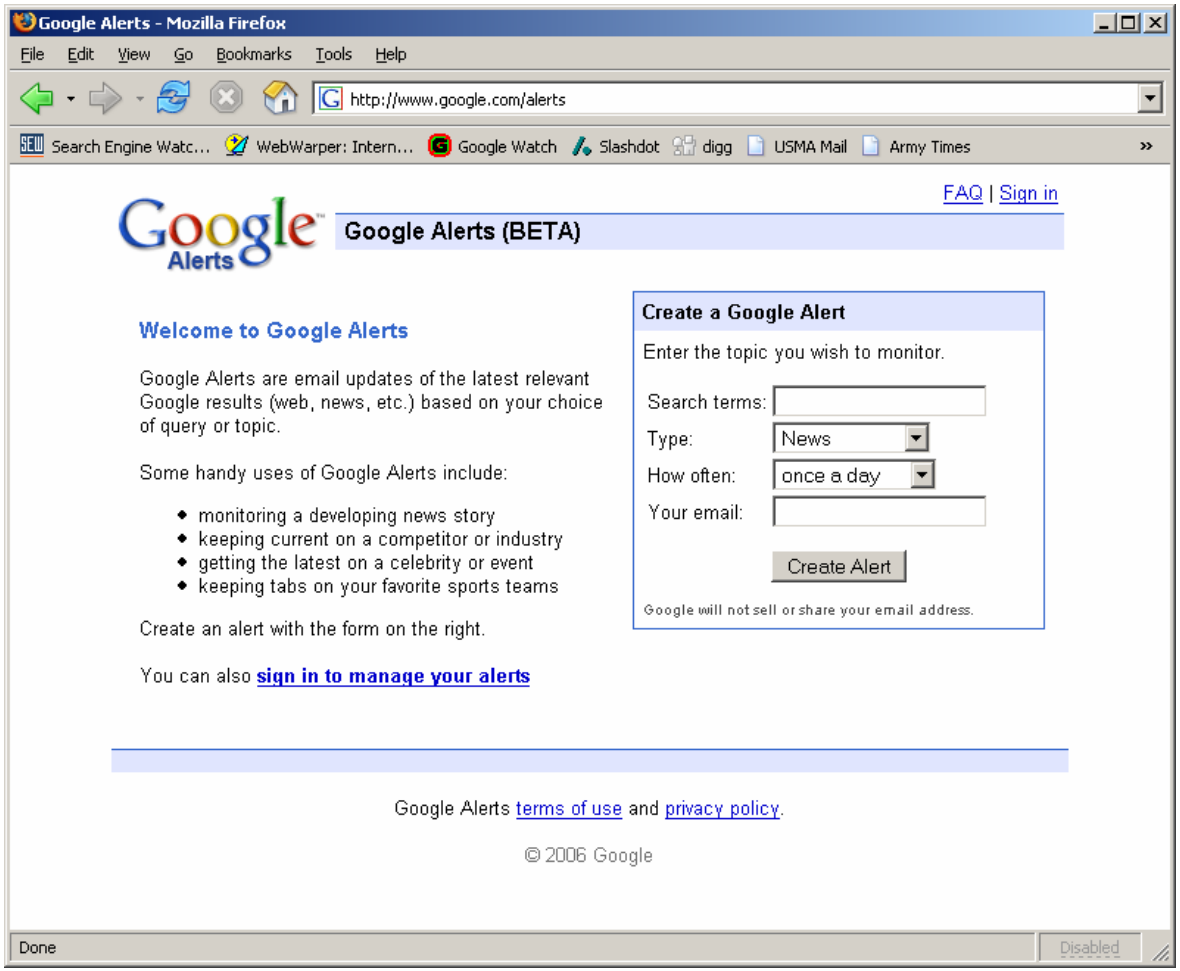
google
google api
google autolink
google base
google blacklist
google chat
google contest
google hacked
google intervened
google isp
google keyhole
google labs
google maps
google scholar
google screen
google sets
google toolbar
google wallet
google watch
google wireless
googlebot

Quoted Searches...

"informative art"
"installing pcap"
"installing pcap:
<name>
<name>
<name>
<name>
<name>
<name>
<name>
<name>
<name> usma

<name>
"kp ddk"
<name>
<name>
"luddite t-shirt"
"man elf"
"man elf" file
"man elf" linux
"man elf(\"
<name>
<name>
<name>

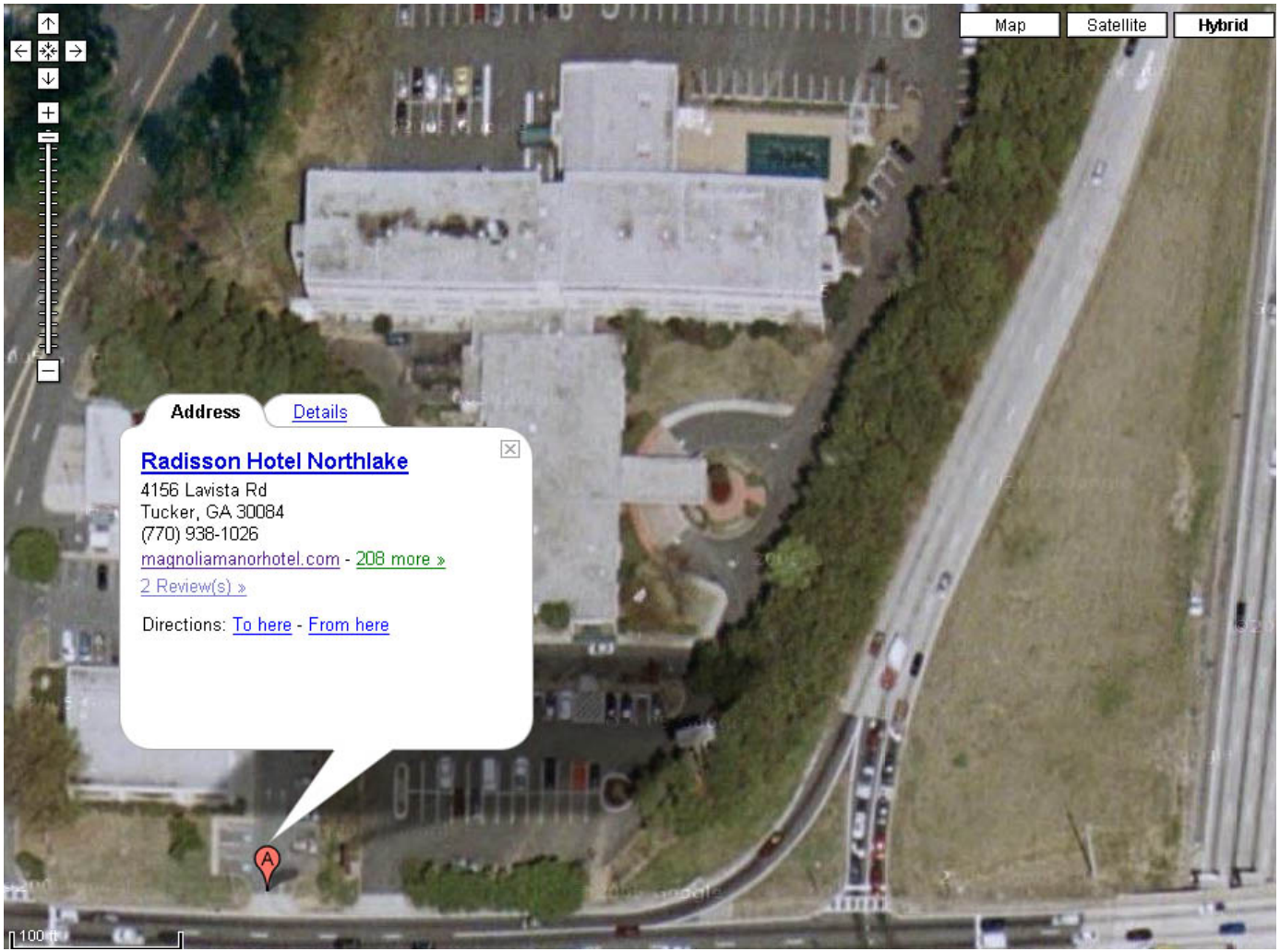
Google Alerts



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

http://www.google.com/alerts

Google Maps / Satellite Imagery



Locations of interest to you
(down to street level)

What have you looked at?

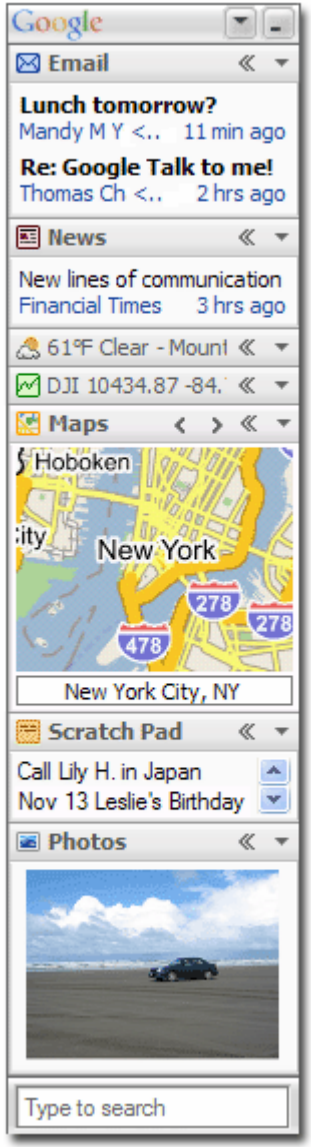
Google Earth
(desktop application)

- 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”

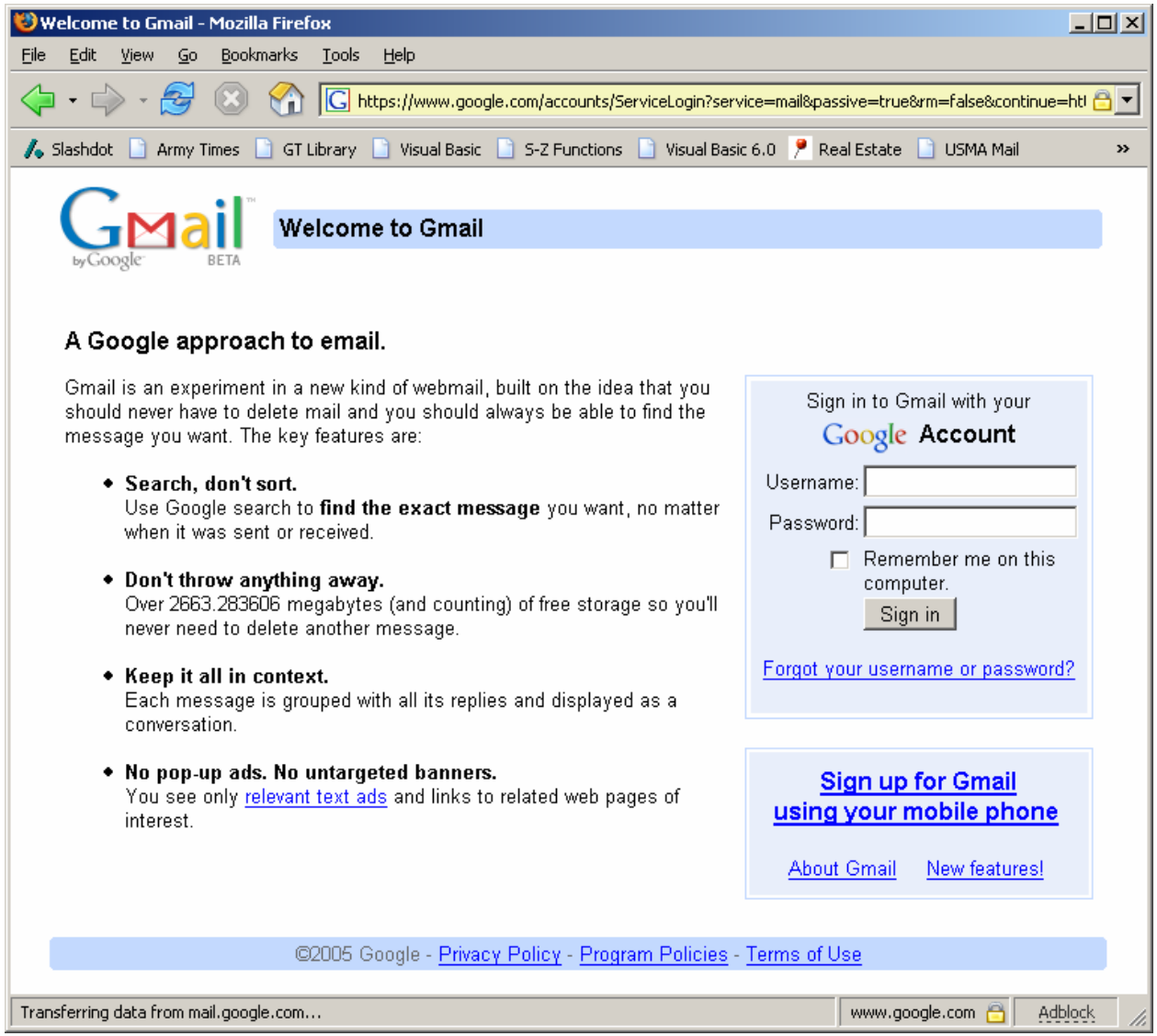


Plug-ins

- **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

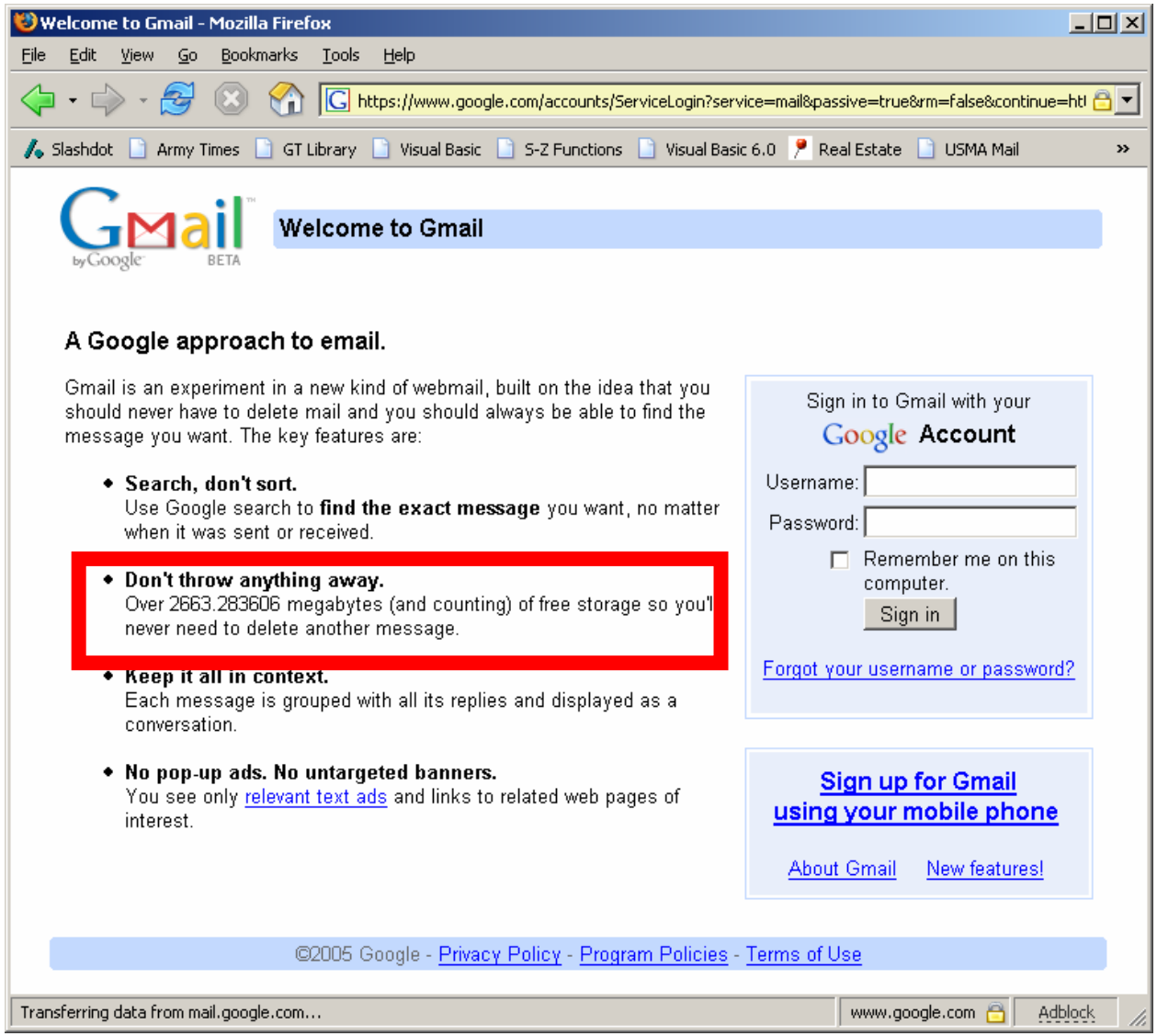
http://desktop.google.com/

Gmail



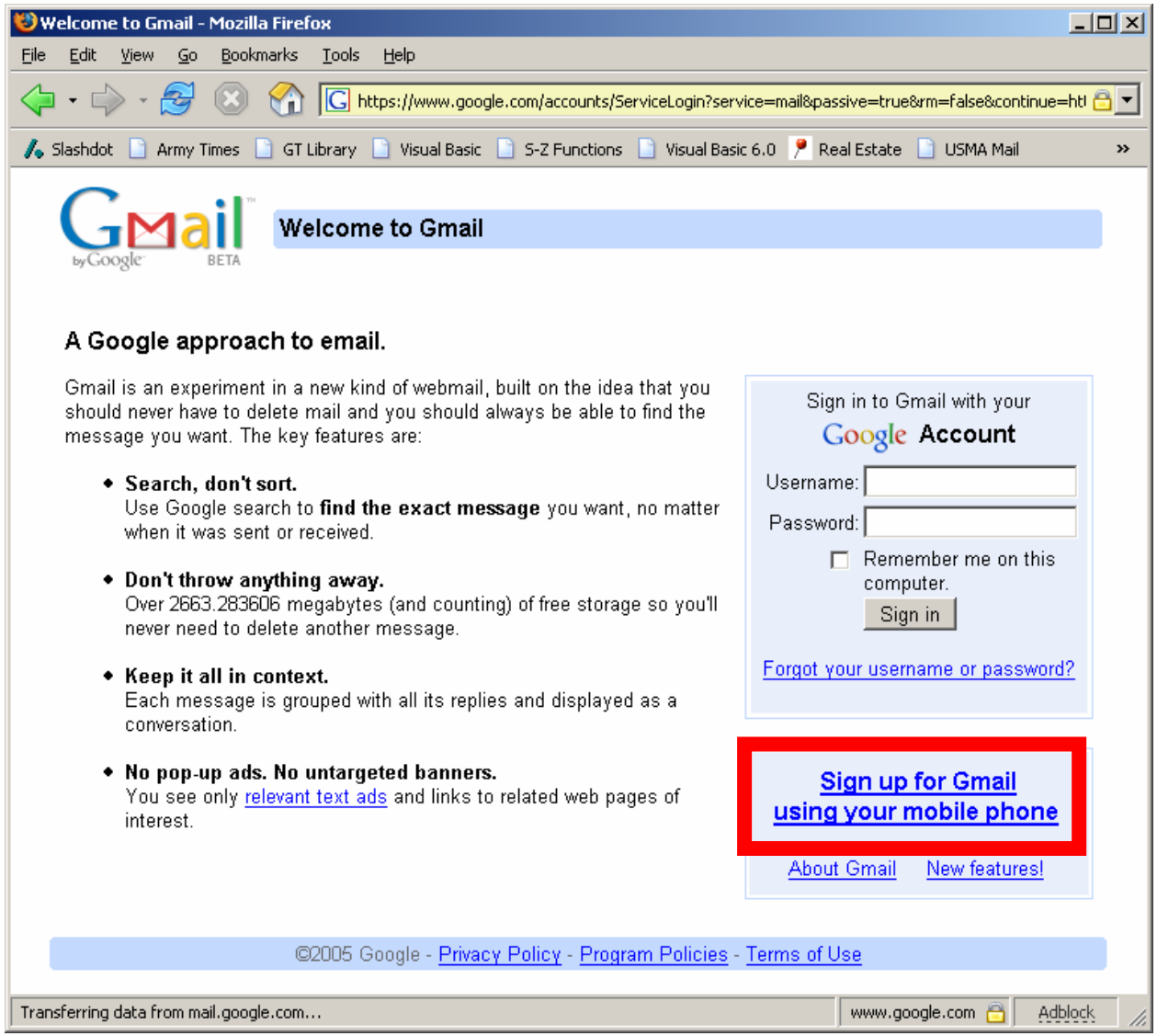
- email recipients
- email sources
- content
- N-order contact

Gmail



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

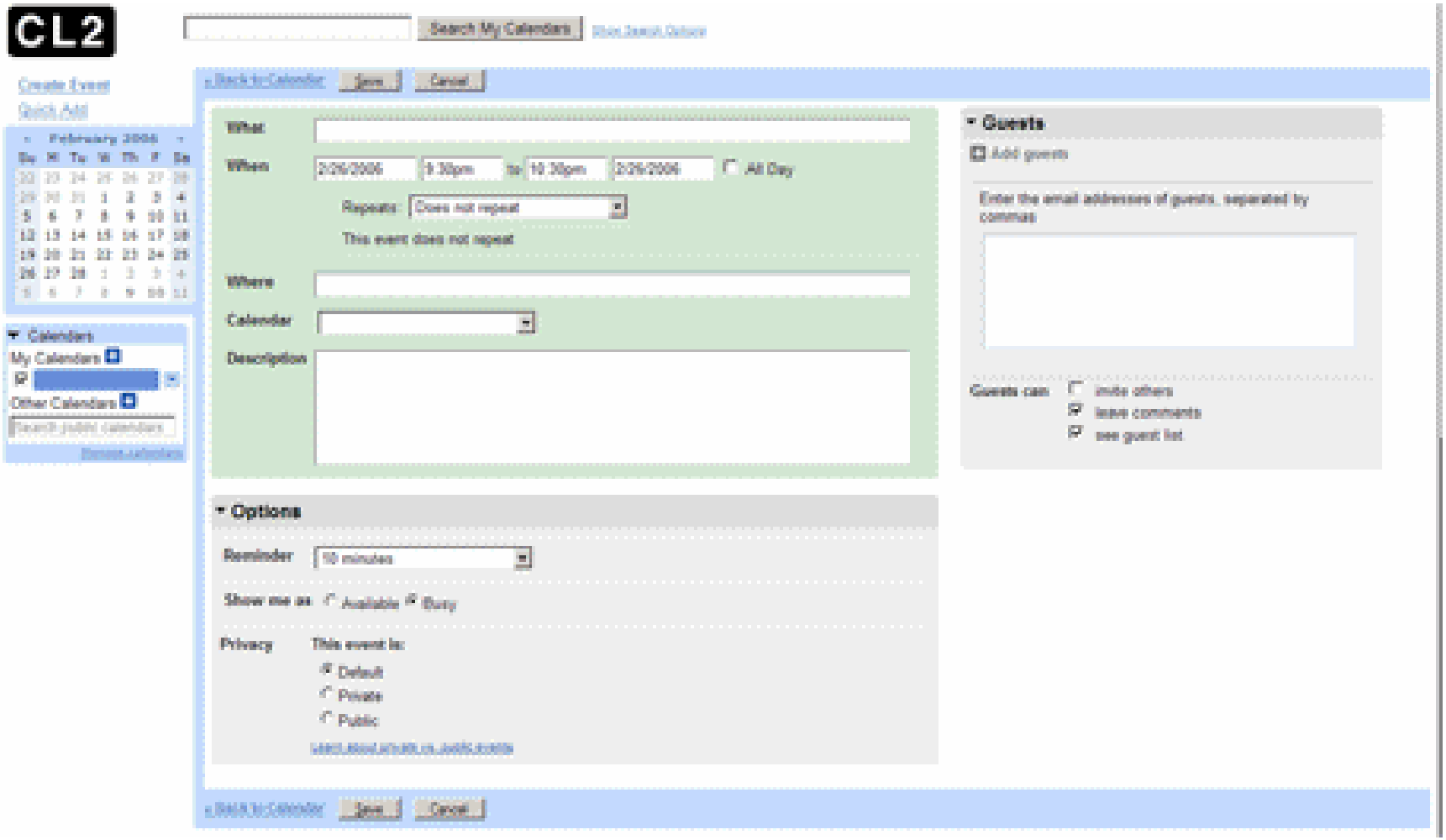
Gmail



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

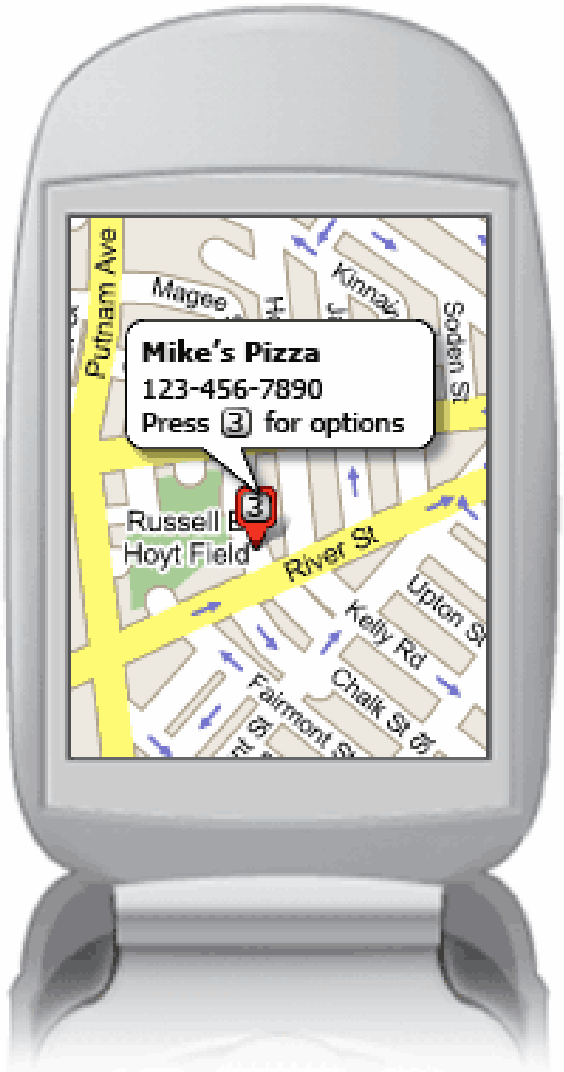
Google Calendar (CL2)

(unconfirmed)



http://slashdot.org/articles/06/03/10/0522252.shtml

Local Information for Local Devices



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

http://www.google.com/glm/index.html

Search Appliances

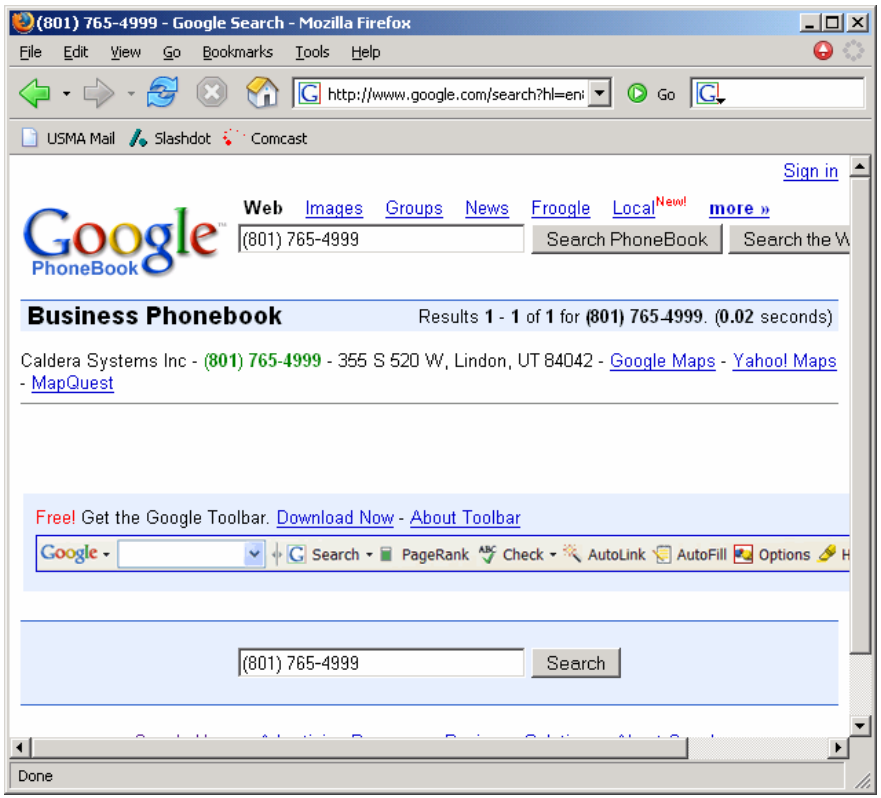
The Google Search Appliance makes the sea of lost data on your web servers, file systems and relational databases instantly available with one mouse click.

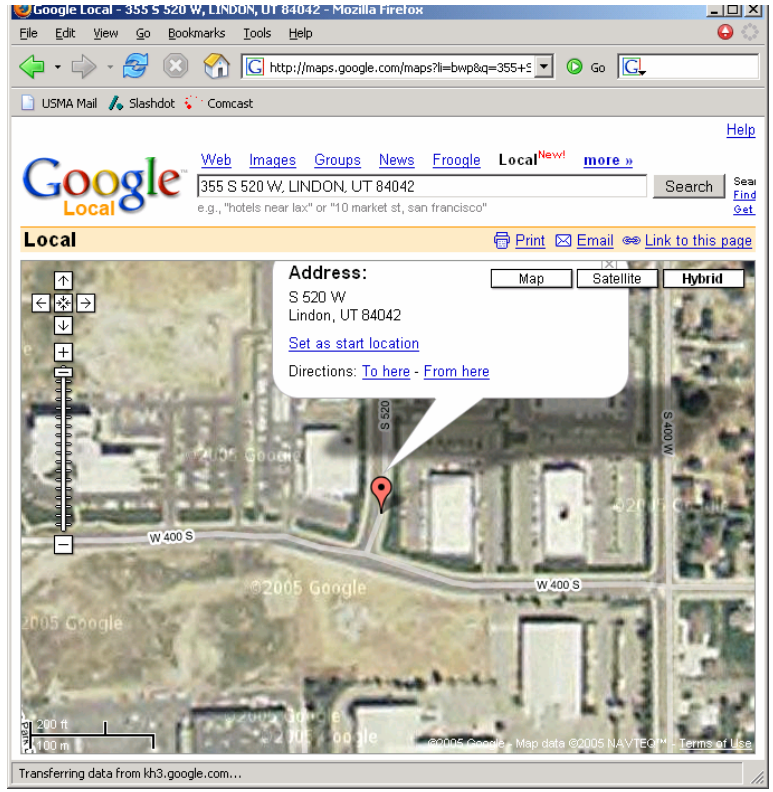
Just point it toward your content, add a search box to your site, and in a matter of hours, your users will be able to search through more than 220 different file formats in any language.

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.

http://www.google.com/enterprise/gsa/index.html

Residential/Commercial Phone Number Lookup

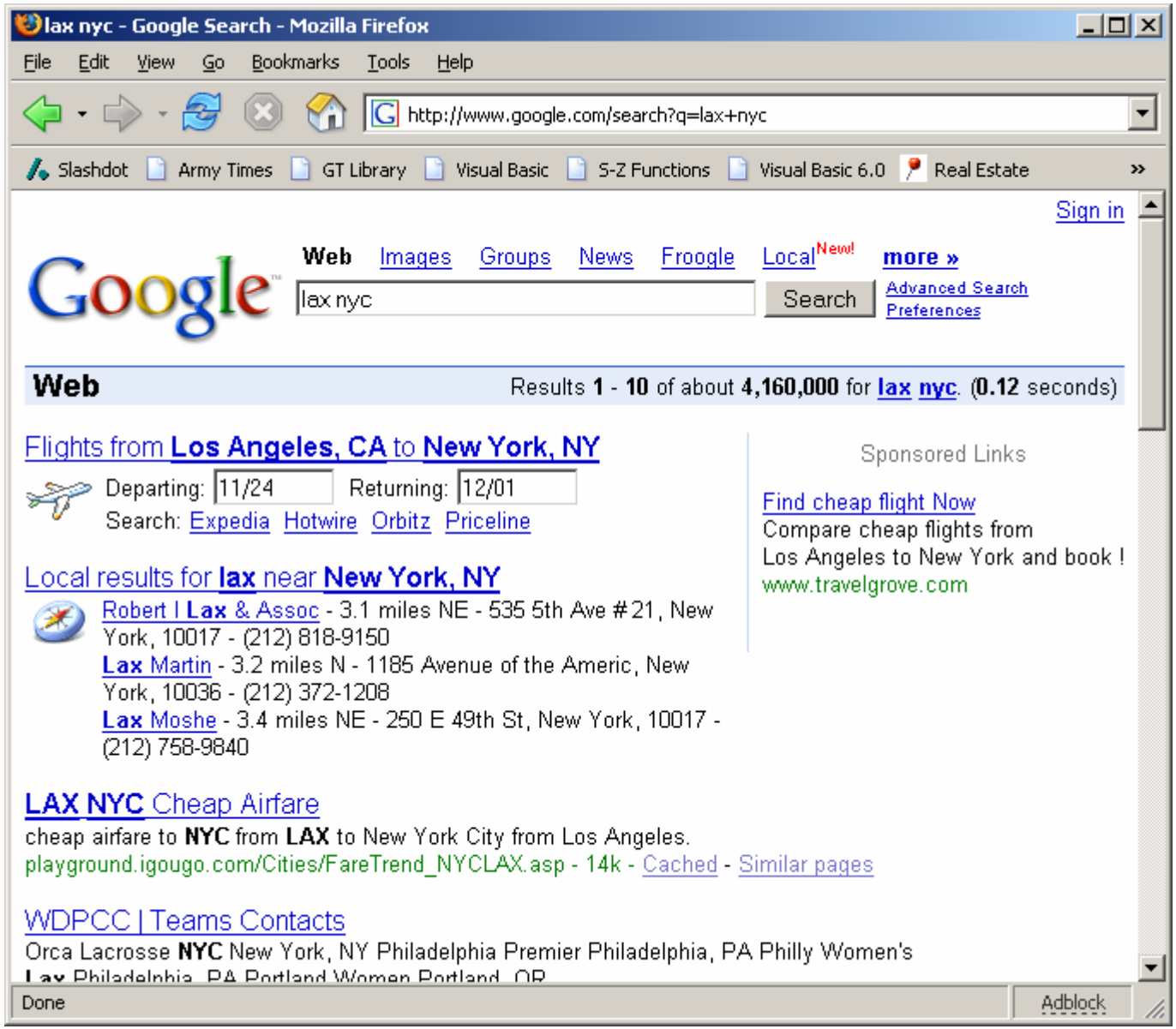




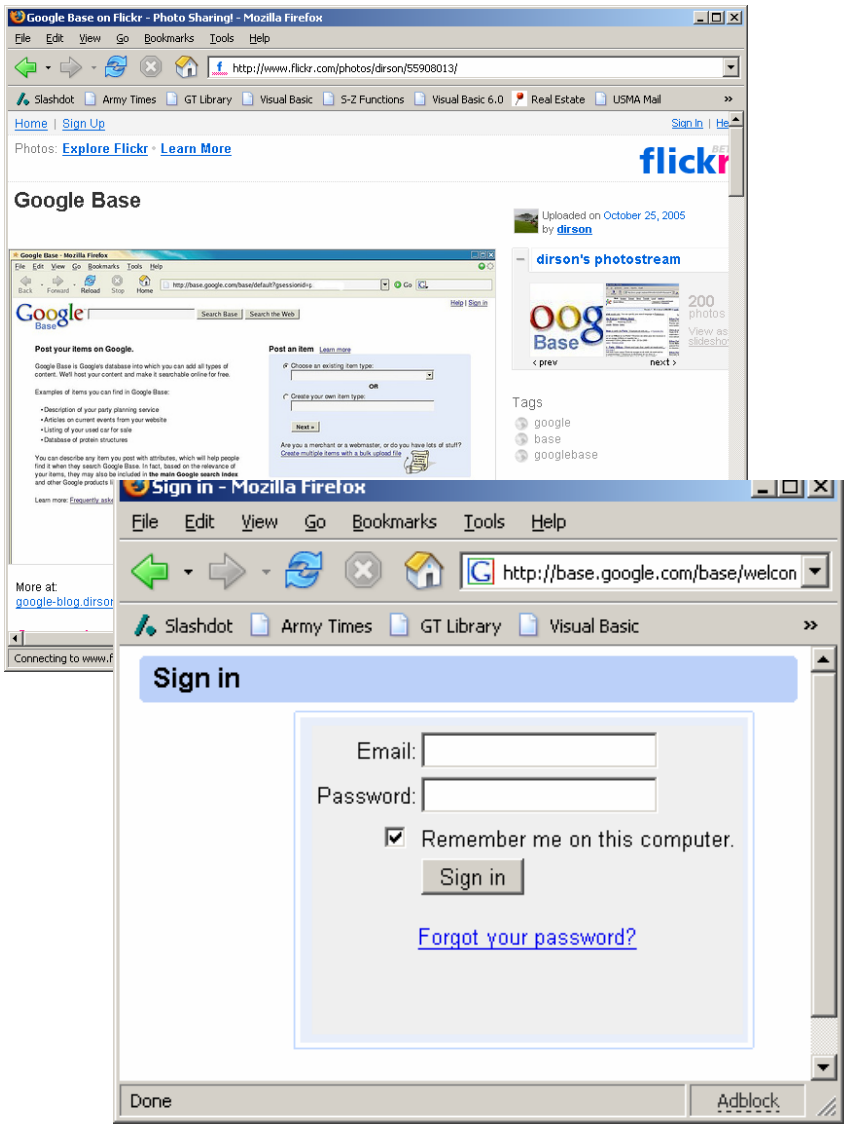
a search returns...

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

Travel Support



Google Base

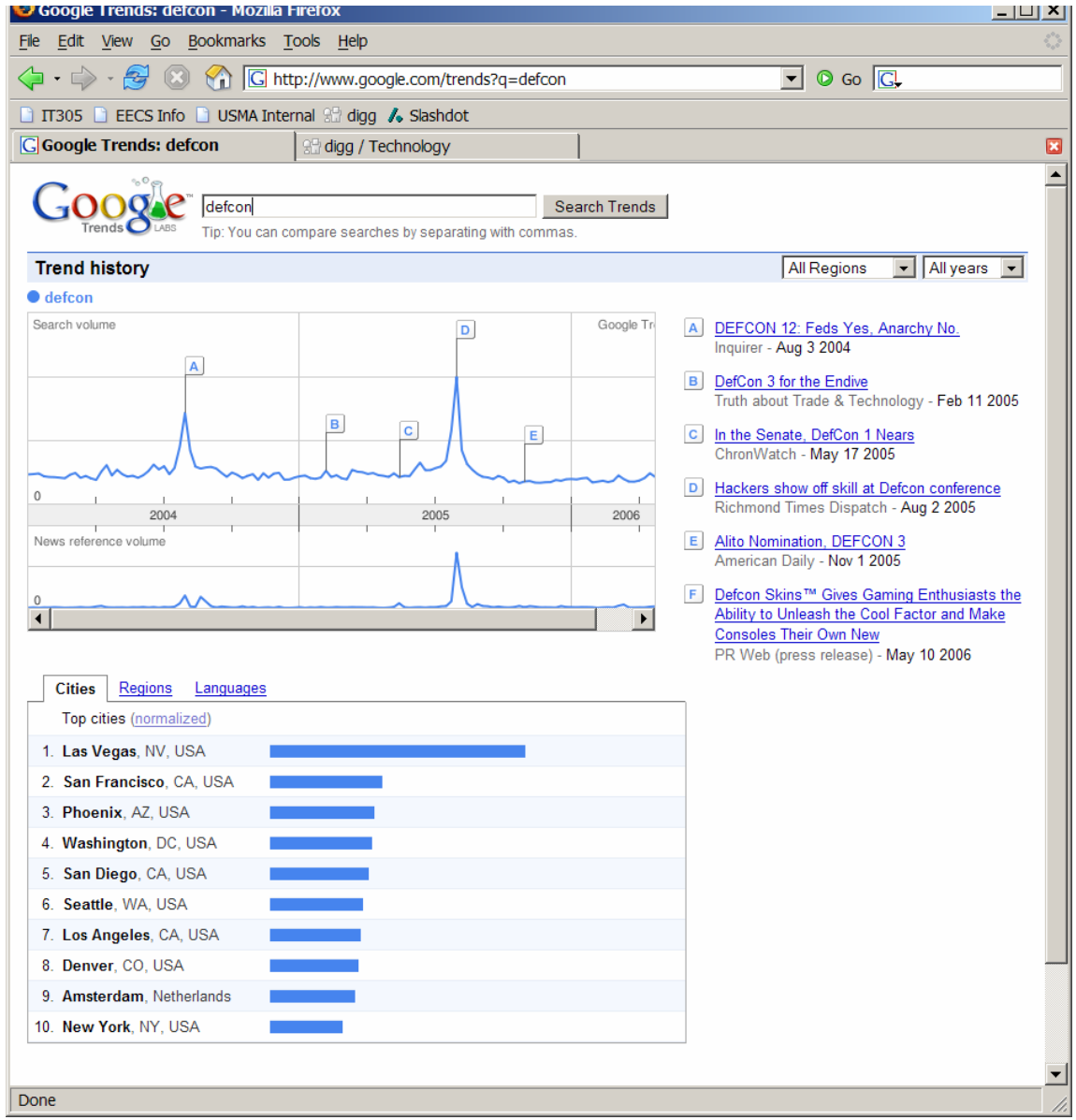


- A “database into which you can add all types of content. We’ll host your content and make it searchable online for free”
- “A free classified advertising system” (NYT)
- Examples
  - Description of your party planning service
  - Articles on current events from you website
  - Listing of your used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

http://www.flickr.com/photos/dirson/55908013/ http://blog.outer-court.com/archive/2005-10-25-n57.html http://base.google.com/ http://www.googlebase.com/

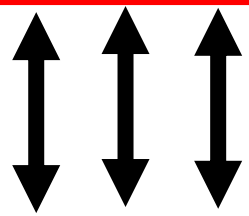
Collaborative Word Processing

Google Trends



http://www.google.com/trends?q=defcon

Information others provide or retrieve

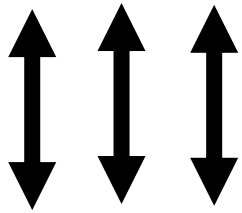


Information you provide or retrieve

Google™

Information Google retrieves
(think Googlebot)

Information others provide or retrieve



Information you provide or retrieve

Google™

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>

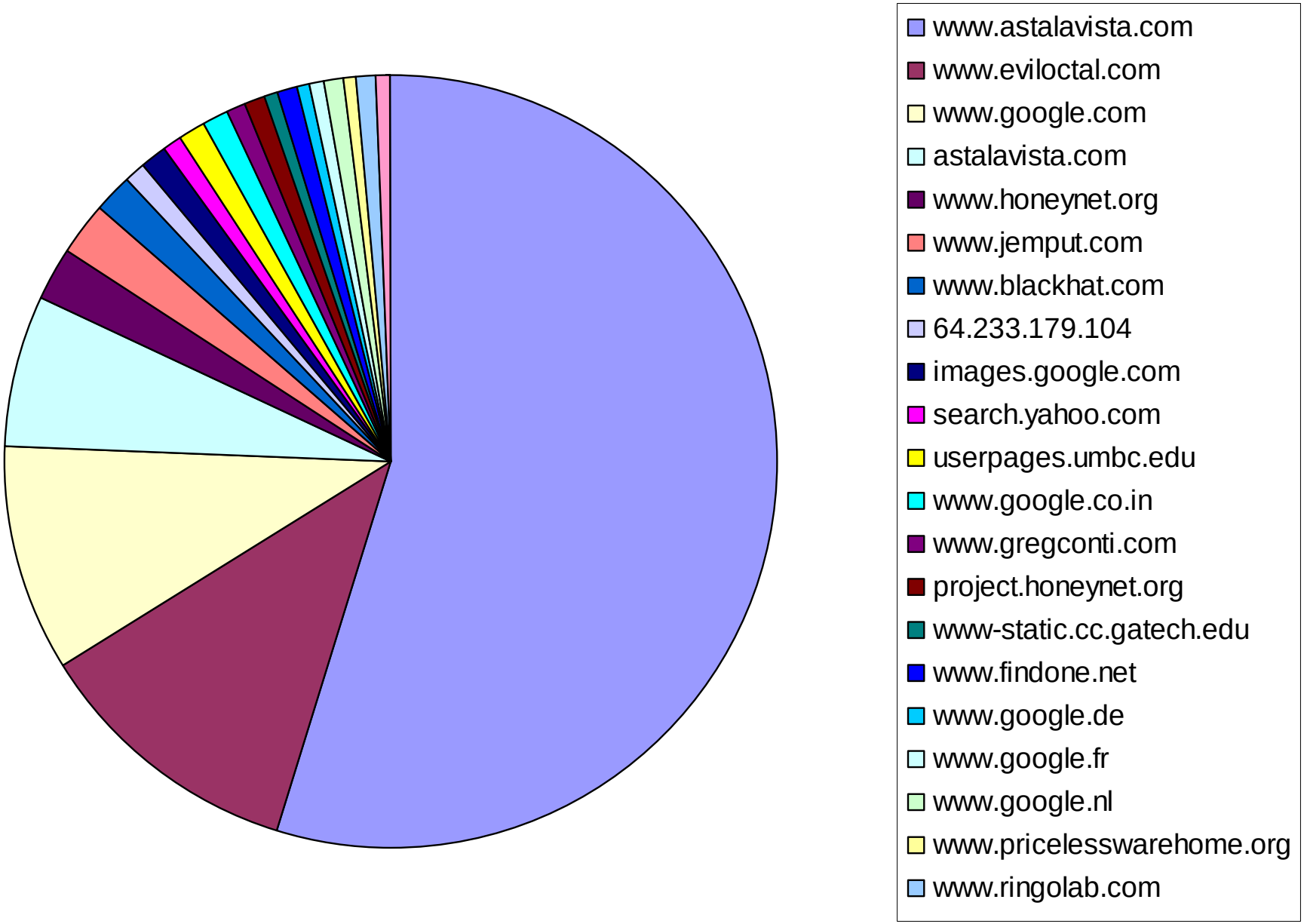
<school> + <media type>

<research area> + <magazine>

<research area> + <media type>

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

Referring Domains...



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

-from Slashdot

http://slashdot.org/comments.pl?sid=168861&cid=14076559

photo: http://deseretnews.com/

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

http://usuarios.lycos.es/bbrp/matematicas/mat_2ACV06_1.jpg
Futurama, Fox Broadcasting Company