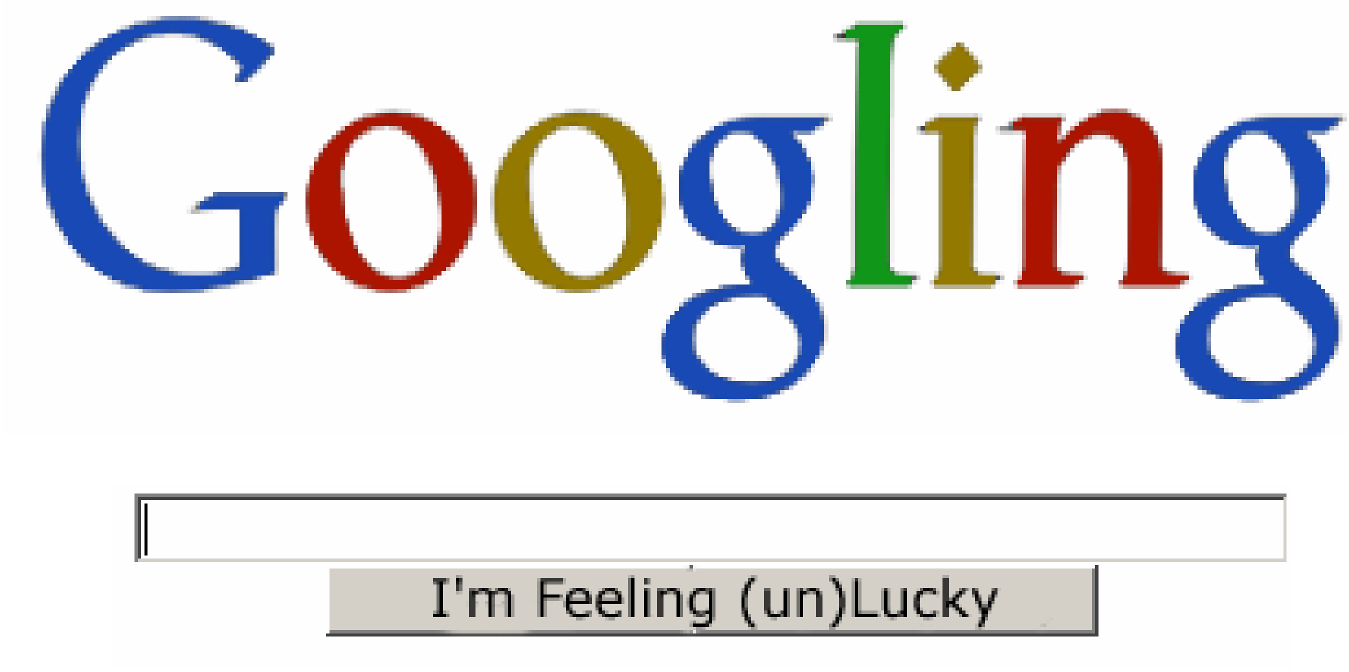
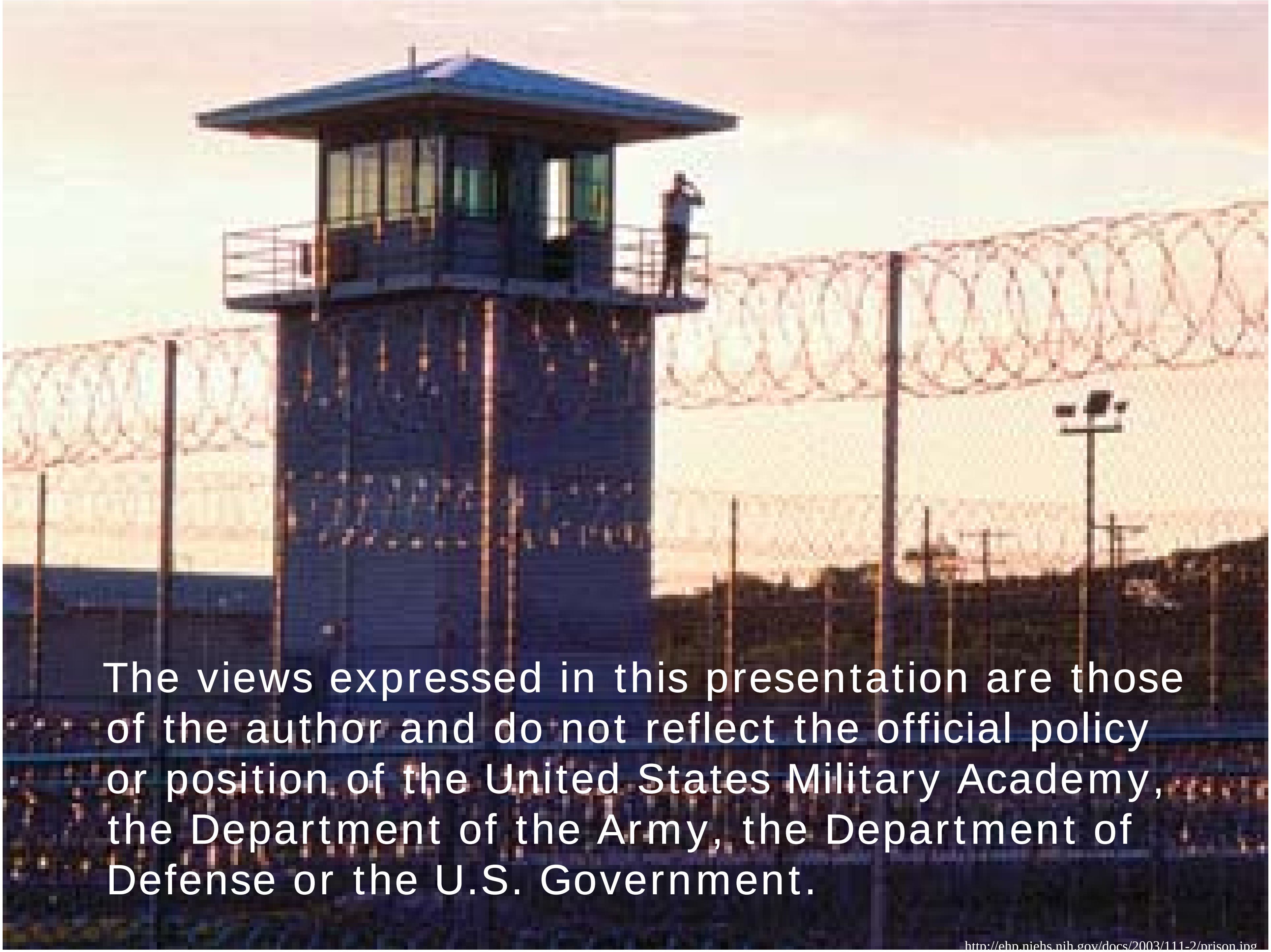


Googling

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

Scenarios

- Patent search
- Serious medical condition
- Failing company
- Anonymous blogger
- Death in the family

Usama Fayyad
Chief Data Officer
2005 Interview with ACM

YAHOO!

http://www.acm.org/sigs/sigkdd/explorations/issues/7-2-2005-12/fayyad.html

What are some of the biggest data mining challenges you face now at Yahoo?

“Yahoo!'s users, through their use of our network of products, generate over 10 terabytes of data per day. This is the equivalent of the entire text contents of the library of Congress. This is data that describes product usage, and does not include content, email, or images, etc.”

“The first and largest challenge is the ability to capture all of this data reliably, process it, reduce it, and use it to feed the many, many [reports and applications]”

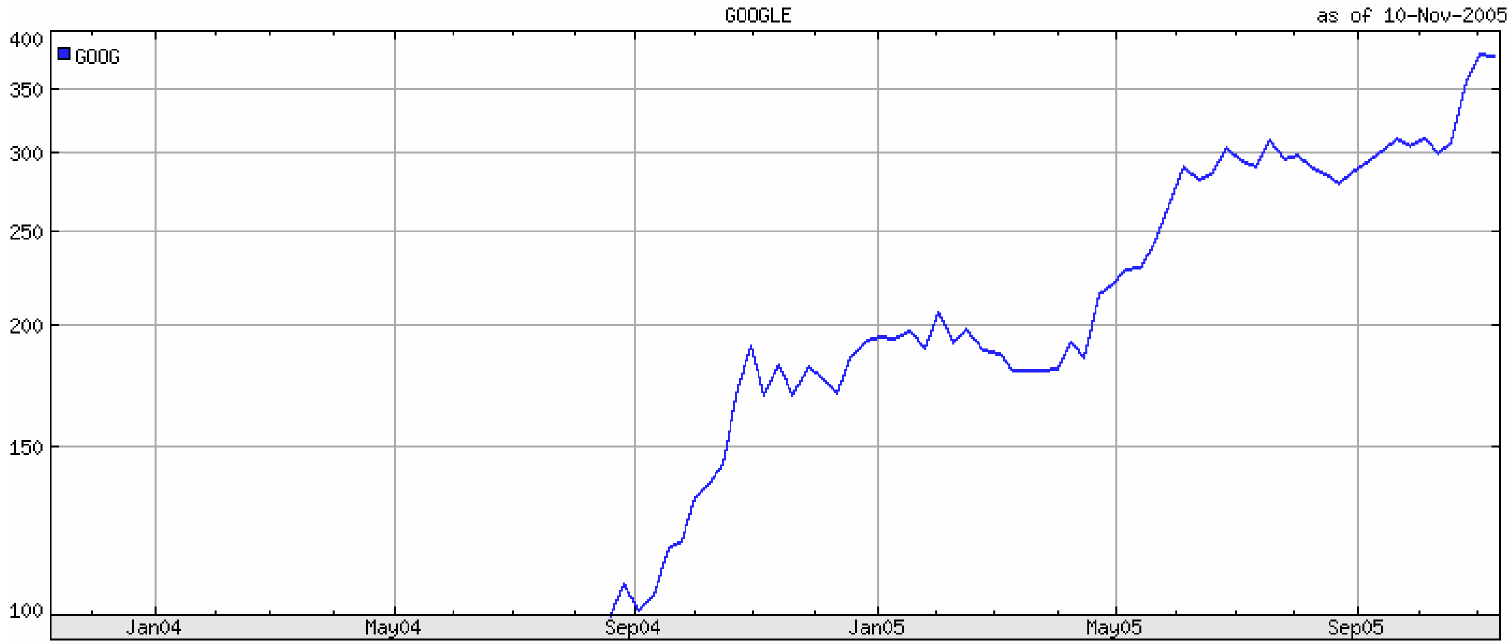
http://www.acm.org/sigs/sigkdd/explorations/issues/7-2-2005-12/fayyad.html

What about Privacy and Data Mining at Yahoo?

“Yahoo! was built around consumer trust and it has been our number one priority since day one. We would never do anything to compromise our users privacy”

“Yahoo!, perhaps more than any other online consumer company, understands that the long-term survival and health of the business is built on consumer trust.”

http://www.acm.org/sigs/sigkdd/explorations/issues/7-2-2005-12/fayyad.html



Copyright 2005 Yahoo! Inc. http://finance.yahoo.com/

Organize the world's information and
make it universally accessible and useful. →

Evil ———————————————— Good

← Generate 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

Related News...

"Google.cn granted license for operation in China"

censorship required to do "legal business" in China

"American Airlines subpoenas Google, YouTube"

AA demands Google reveal name of person who posted airline's training videos.

"Google Subpoena Woes Double"

DOJ, Child Pornography

"Yahoo helped Jail China writer"

"Phishing with Google Desktop"

(remote code execution via malicious website)

Sony Rootkit

Google balances privacy, reach

Elinor Mills (CNET)

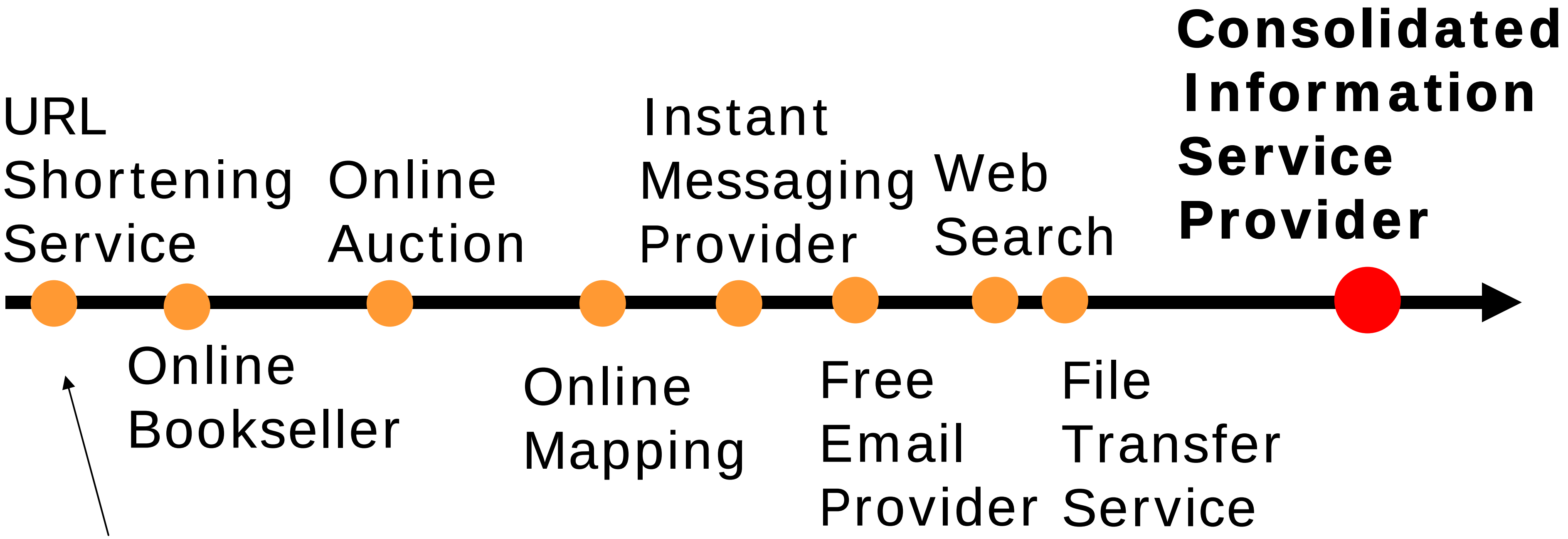
http://www.doxpara.com.nyud.net:8090/planetsony_usa.JPG

What if an entirely altruistic provider

- Insider threat
- Accidental information disclosure
- Merger / Sale of Company
- Change in leadership / corporate philosophy
- Impact of technical advances
- Legal compulsion

What if not?

Information Disclosed



URL Shortening Service

Online Bookseller

Online Auction

Online Mapping

Instant Messaging Provider

Free Email Provider

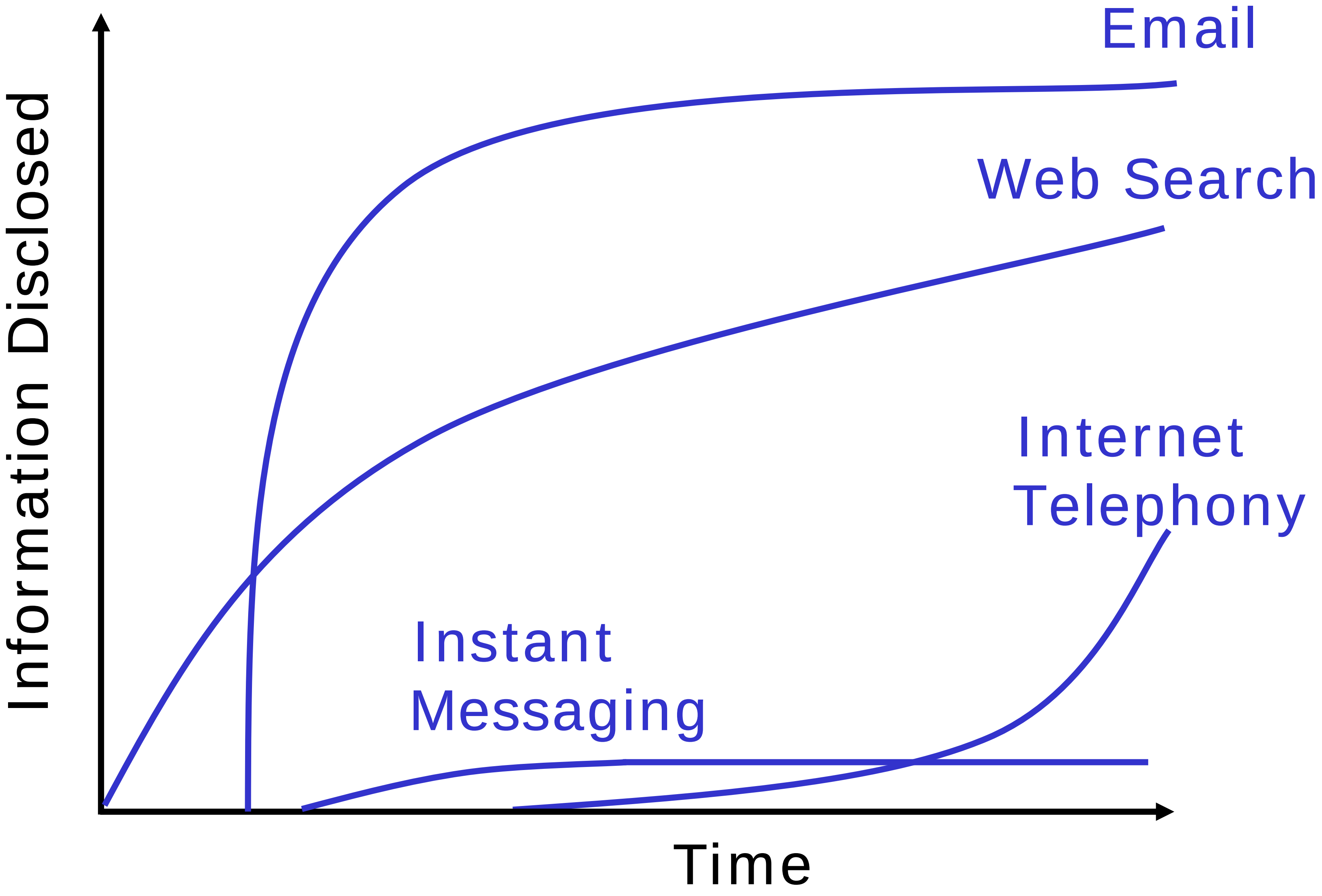
Web Search

File Transfer Service

Consolidated Information Service Provider

Tiny URL
16 Million URLs
325M hits/month

I have searched for things I wouldn’t want my grandmother* to know about.


Email
Web Search
Internet Telephony
Instant Messaging
Information Disclosed
Time

Case Study: Google

Google Zeitgeist 2005

Google.com
Top Gainers of 2005

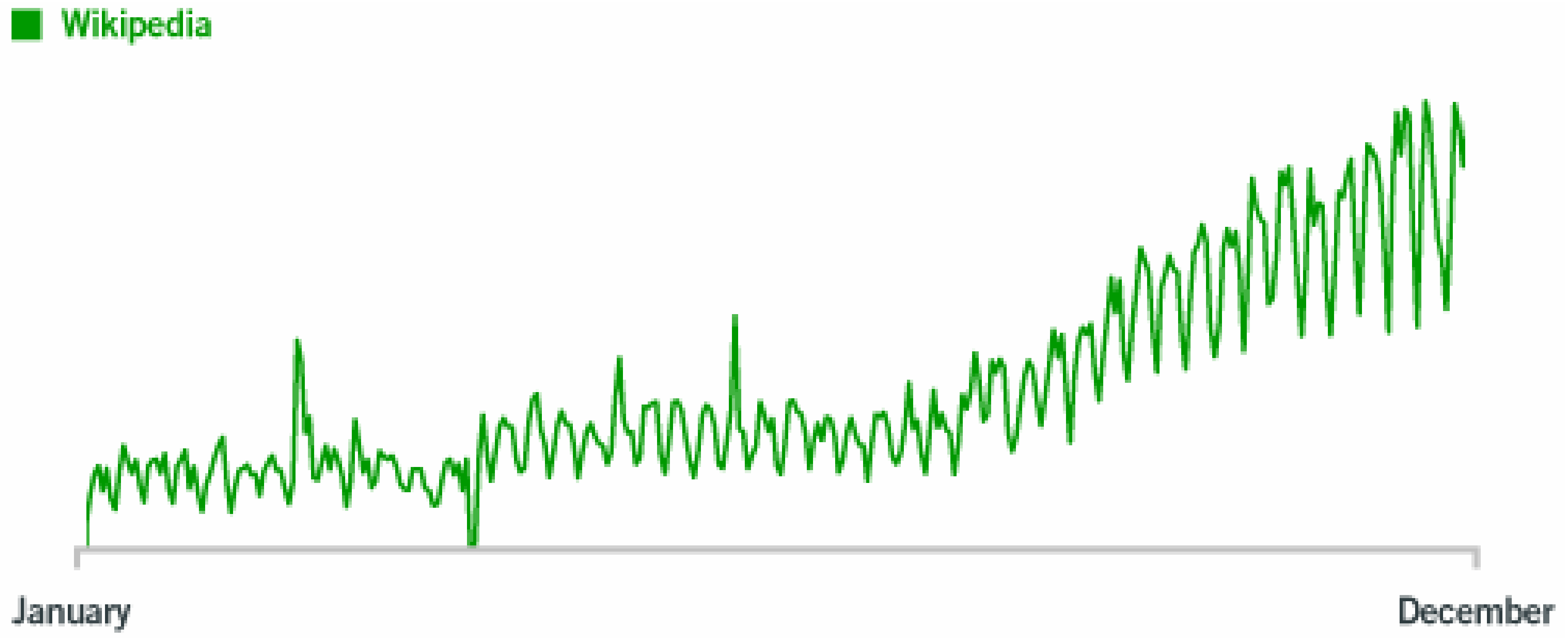
1. Myspace
2. Ares
3. Baidu
4. wikipedia
5. orkut
6. iTunes
7. Sky News
8. World of Warcraft
9. Green Day
10. Leonardo da Vinci

Google News
Top Searches in 2005

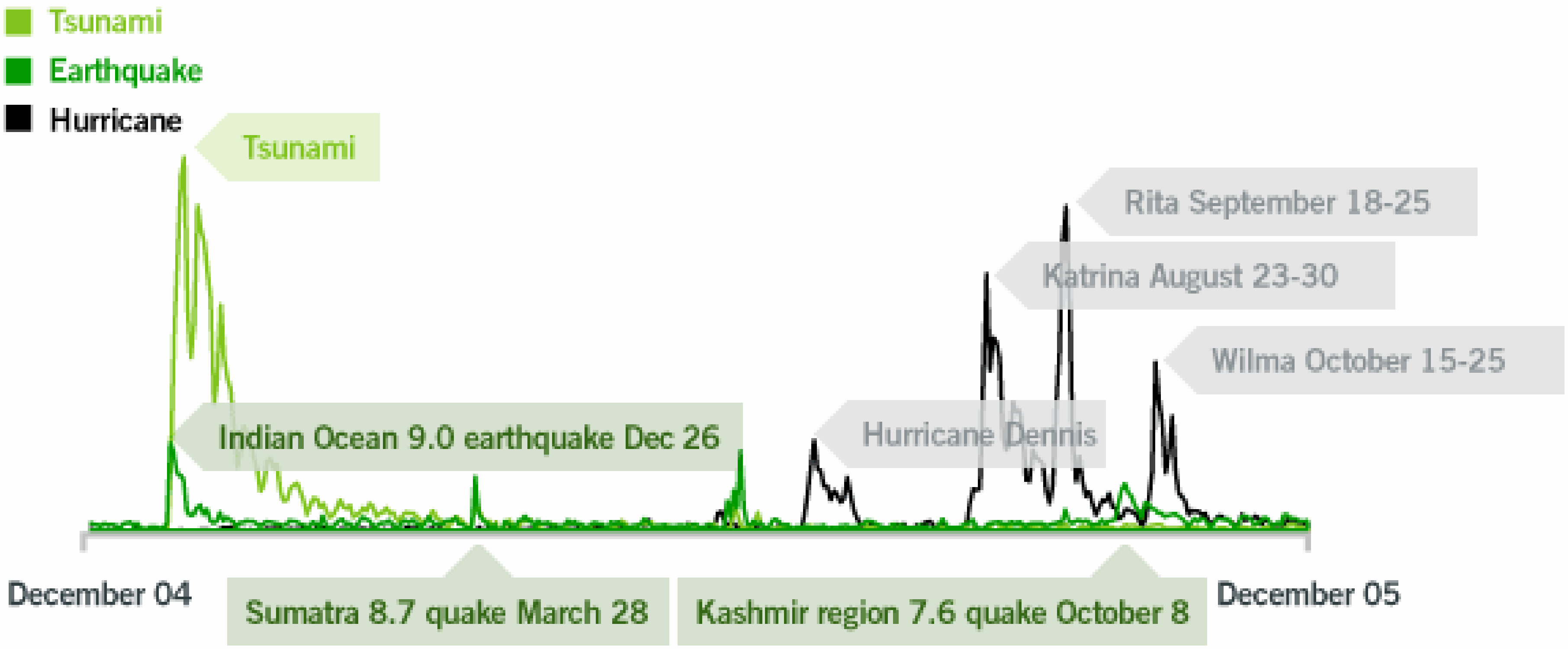
1. Janet Jackson
2. Hurricane Katrina
3. tsunami
4. xbox 360
5. Brad Pitt
6. Michael Jackson
7. American Idol
8. Britney Spears
9. Angelina Jolie
10. Harry Potter

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



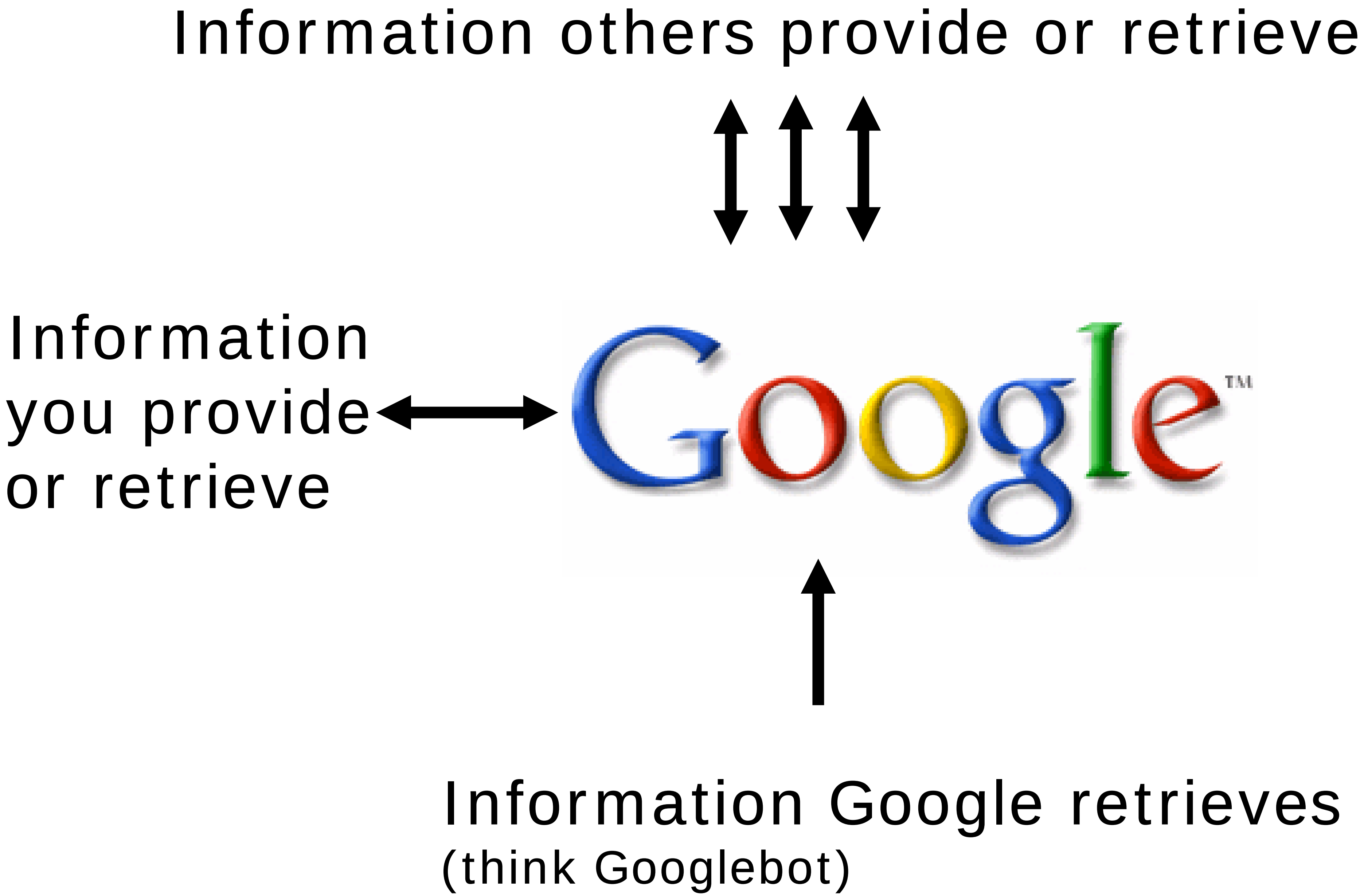
http://www.google.com/press/zeitgeist2005/nature.html

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/



Information others provide or retrieve

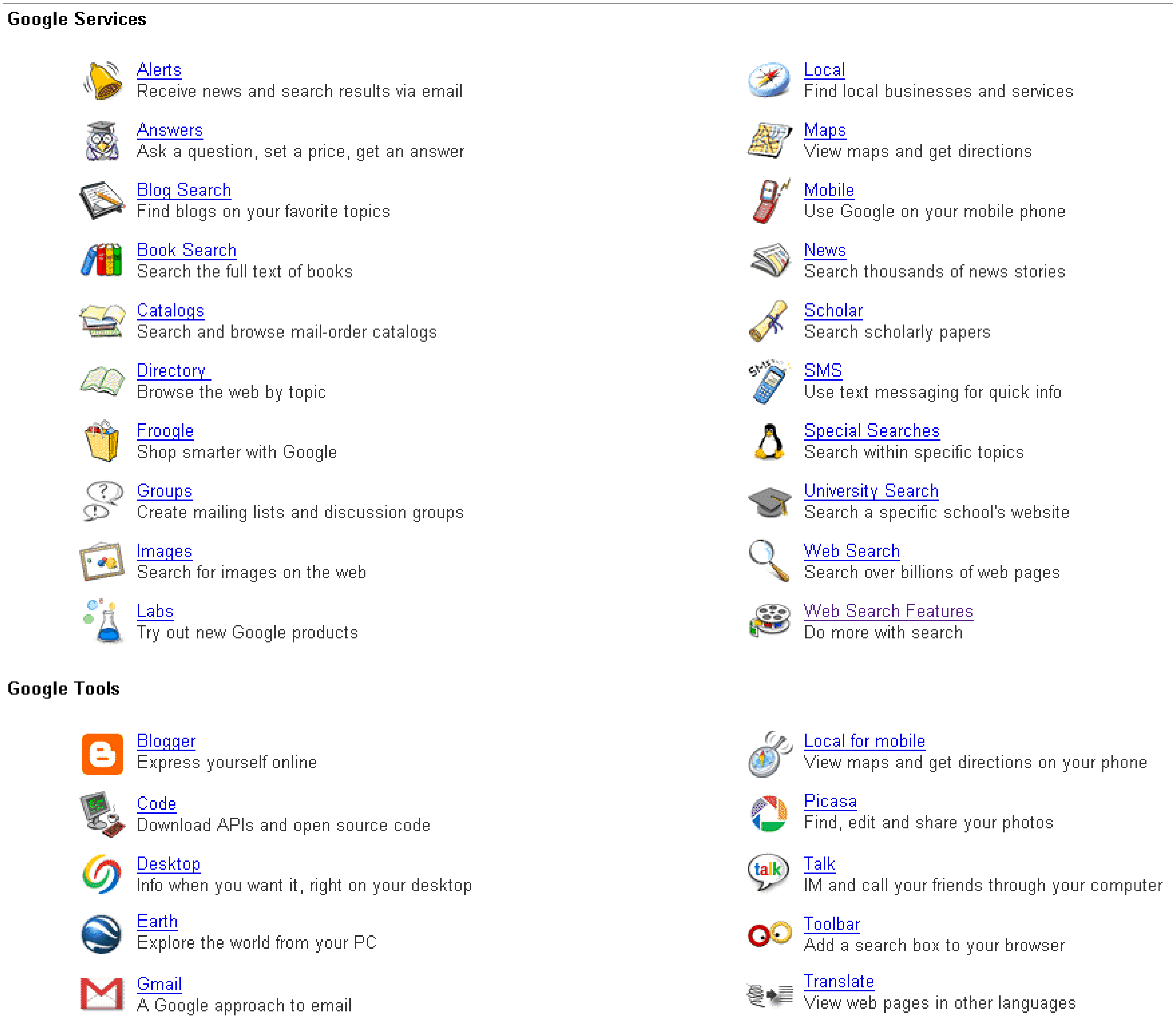
Information you provide or retrieve

Information Google retrieves
(think Googlebot)

Disclosure Vectors

“When Did You Lose Your Google Virginity?”

Quote by Scott Lemon, http://the.inevitable.org/anism/



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

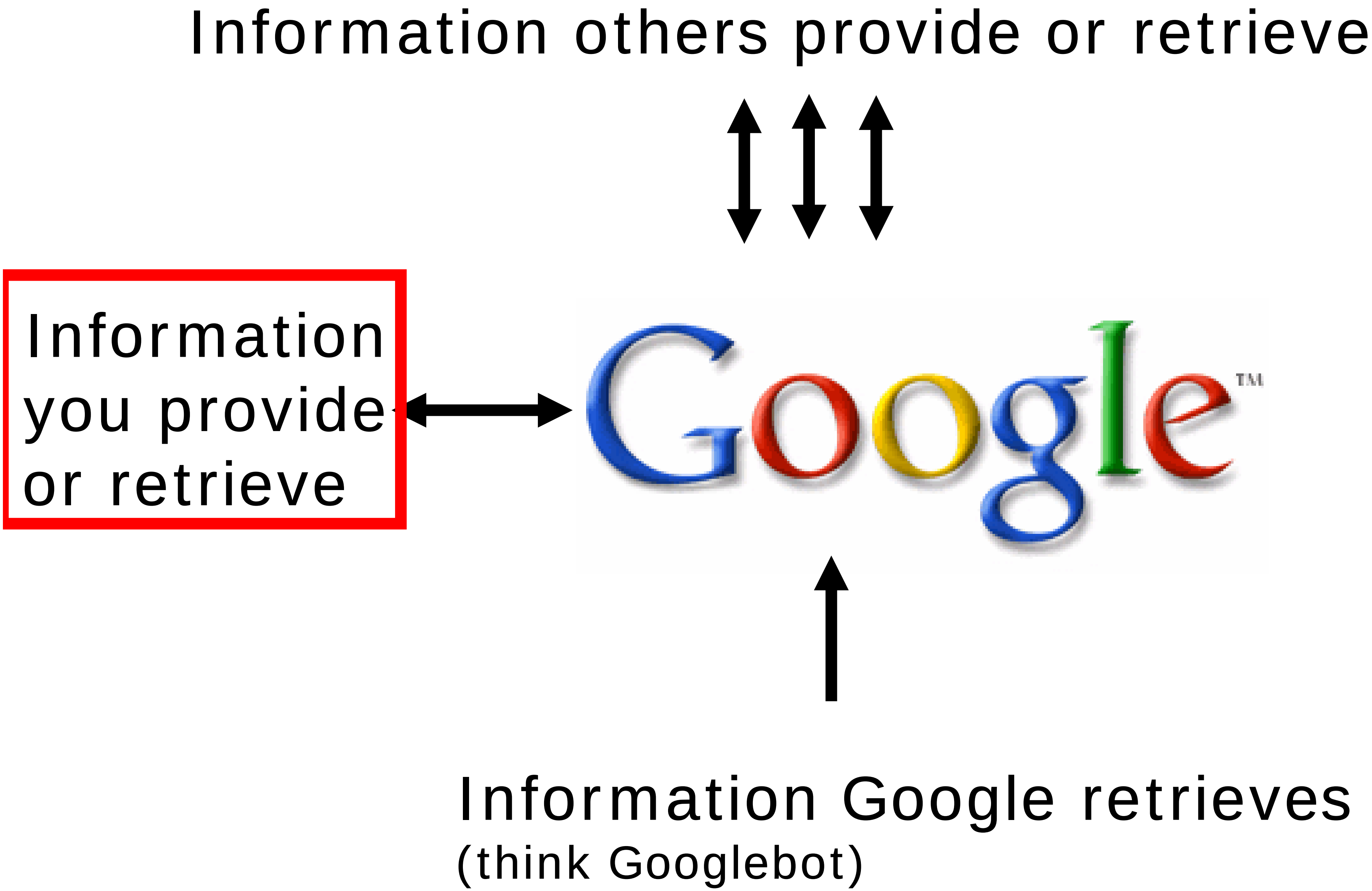
Local for mobile
View maps and get directions on your phone

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



Information others provide or retrieve

Information you provide or retrieve

Information Google retrieves

(think Googlebot)

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

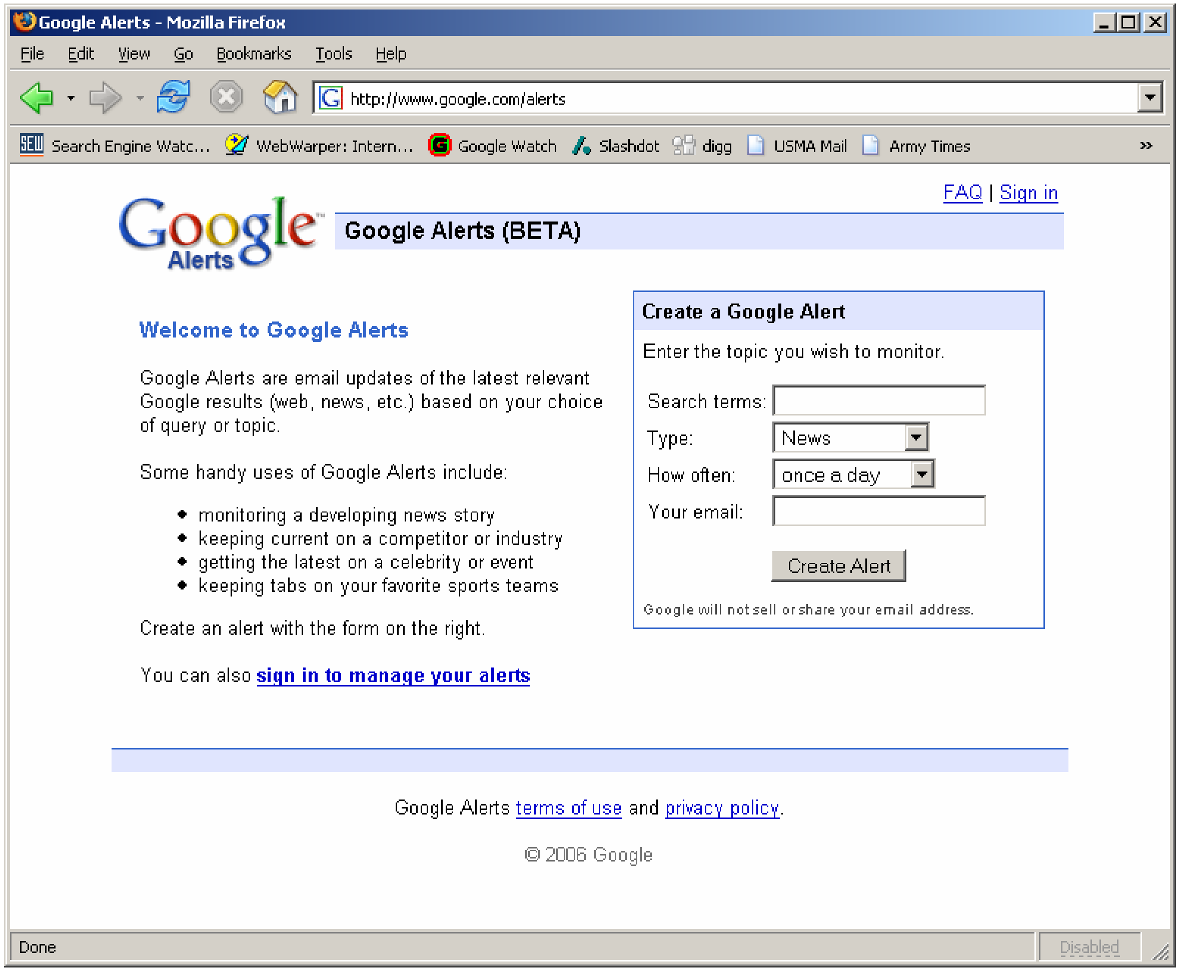
google watch

google wireless

googlebot

I use Google as an Address Book.

Google Alerts


Google Alerts - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/alerts
Search Engine Watc... WebWarper: Intern... Google Watch Slashdot digg USMA Mail Army Times
FAQ | Sign in
Google Alerts
Done
Disabled

Google Alerts (BETA)

Welcome to Google Alerts

Google Alerts are email updates of the latest relevant Google results (web, news, etc.) based on your choice of query or topic.

Some handy uses of Google Alerts include:

- monitoring a developing news story
- keeping current on a competitor or industry
- getting the latest on a celebrity or event
- keeping tabs on your favorite sports teams

Create an alert with the form on the right.

You can also **sign in to manage your alerts**

Create a Google Alert

Enter the topic you wish to monitor.

Search terms:

Type: News

How often: once a day

Your email:

Create Alert

Google will not sell or share your email address.

Google Alerts terms of use and privacy policy.

© 2006 Google

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

http://www.google.com/alerts

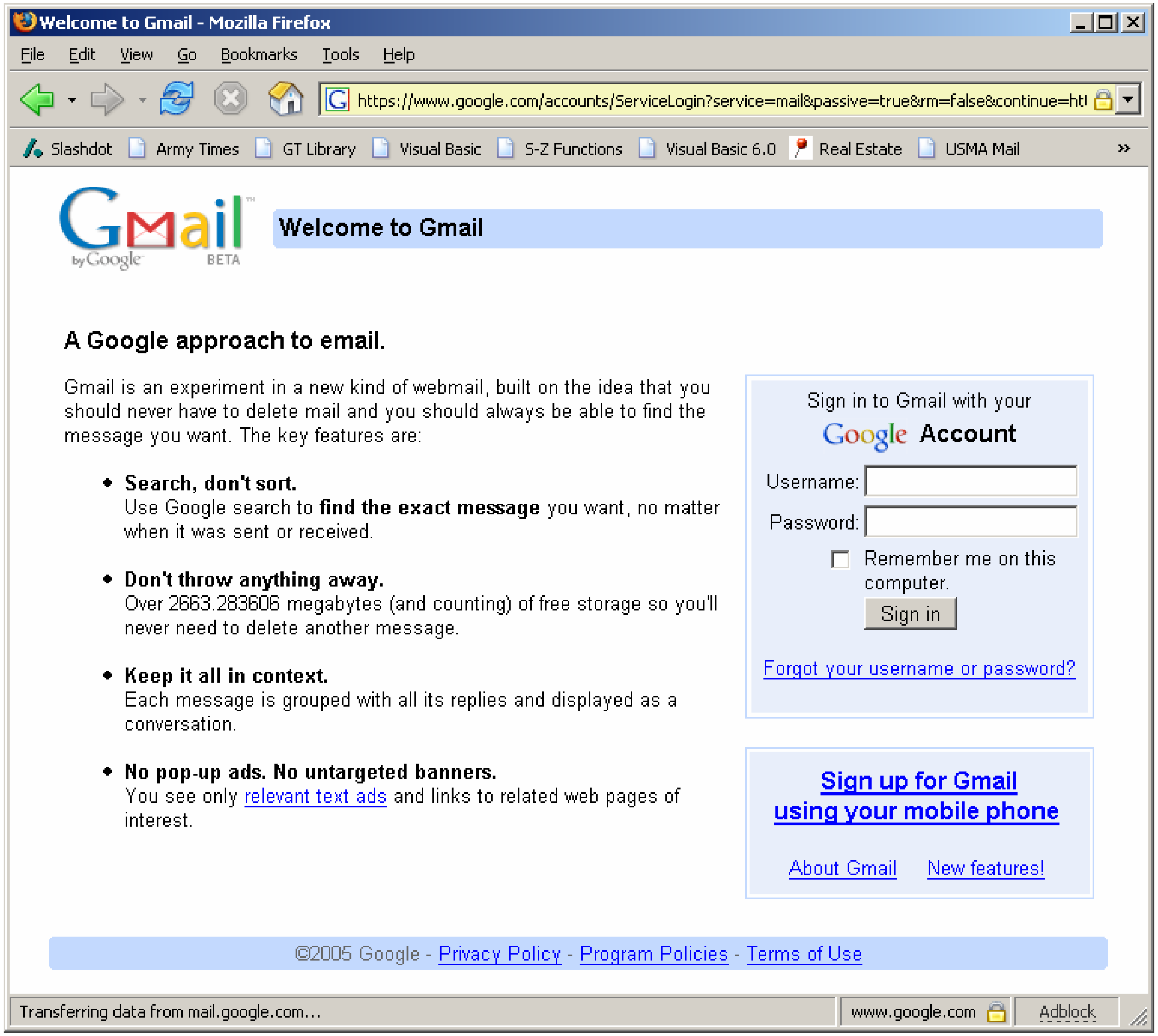
Google Maps / Satellite Imagery

Locations of interest to you (down to street level)

What have you looked at?

Image: maps.google.com

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=htt
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.

Sign in

Forgot your username or password?

Sign up for Gmail using your mobile phone

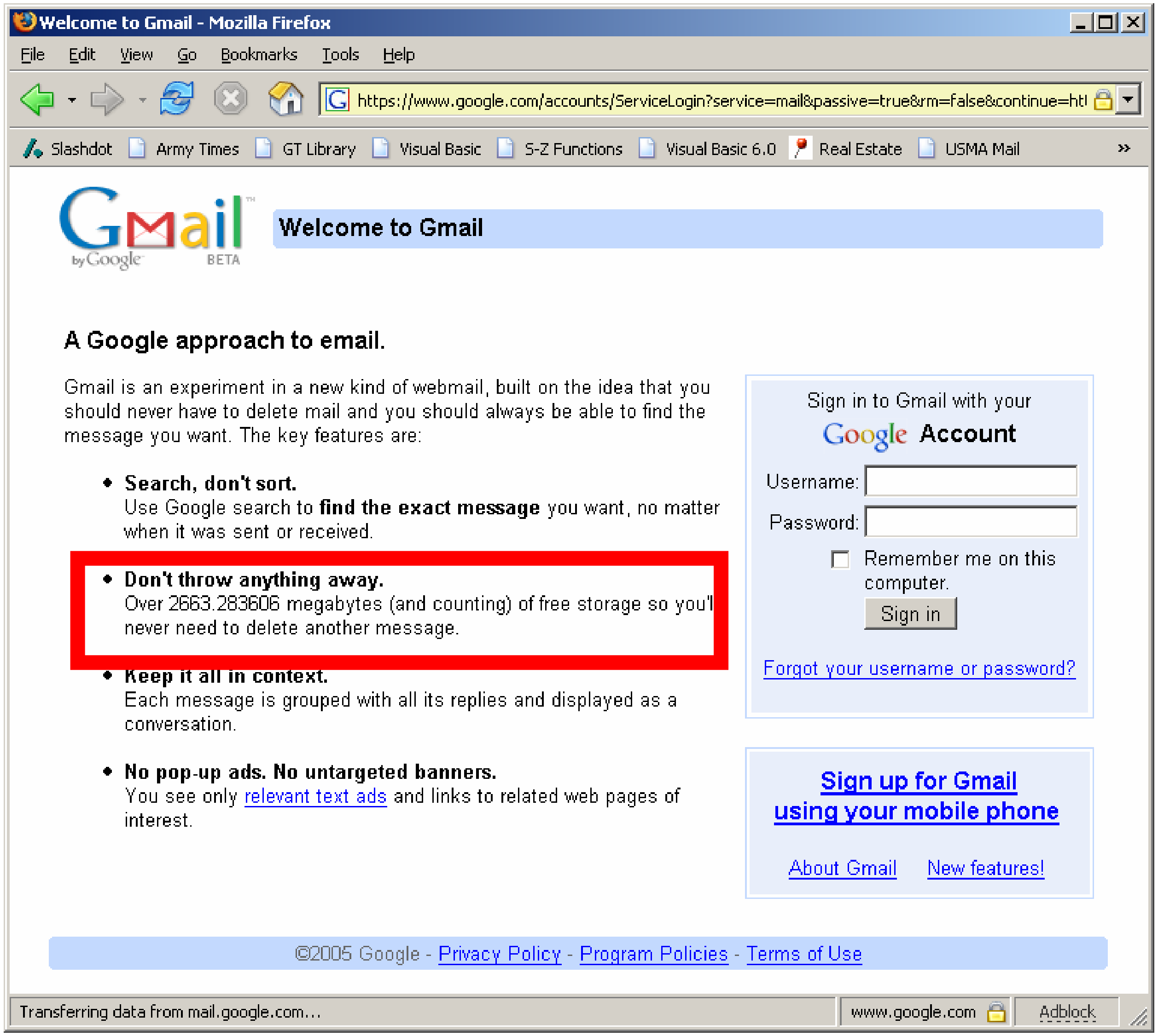
About Gmail New features!

©2005 Google - Privacy Policy - Program Policies - Terms of Use

Transferring data from mail.google.com...

- 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=htt
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.

Sign in

Forgot your username or password?

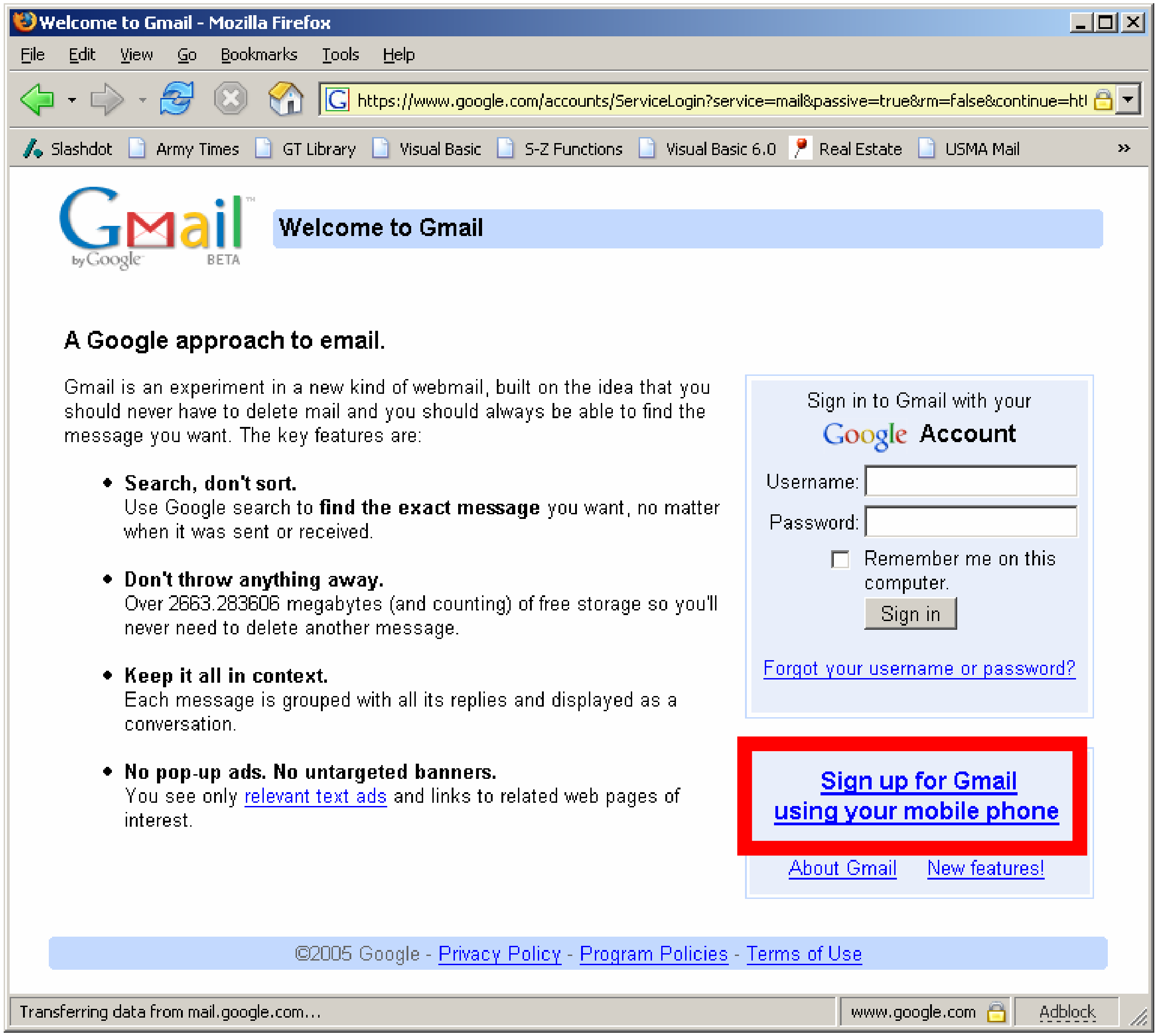
Sign up for Gmail using your mobile phone

About Gmail New features!

©2005 Google - Privacy Policy - Program Policies - Terms of Use

- 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=htt
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.

Sign in
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...

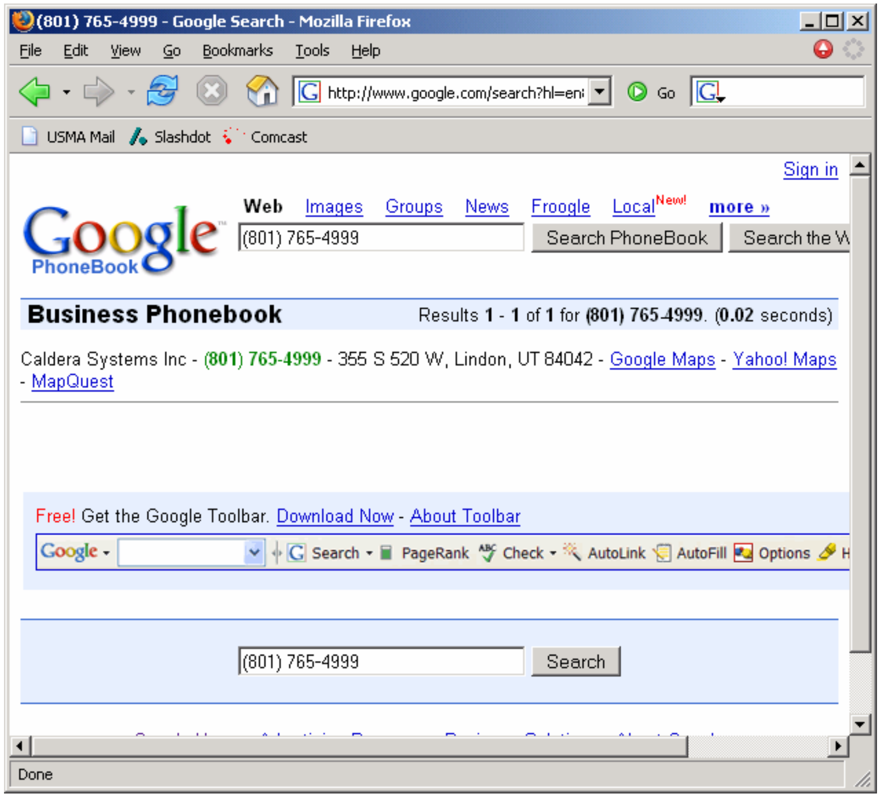
- email recipients
- email sources
- content
- N-order contact

I use a Gmail account.

I have sent an email to a Gmail account.

My Employees are registered Gmail users.

Residential/Commercial Phone Number Lookup


(801) 765-4999 - Google Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/search?hl=en
Go
USMA Mail Slashdot Comcast
Sign in
Google PhoneBook
Web Images Groups News Froogle Local New! more »
Search PhoneBook
Search the W
Google Search PageRank Check AutoLink AutoFill Options
Search
Done

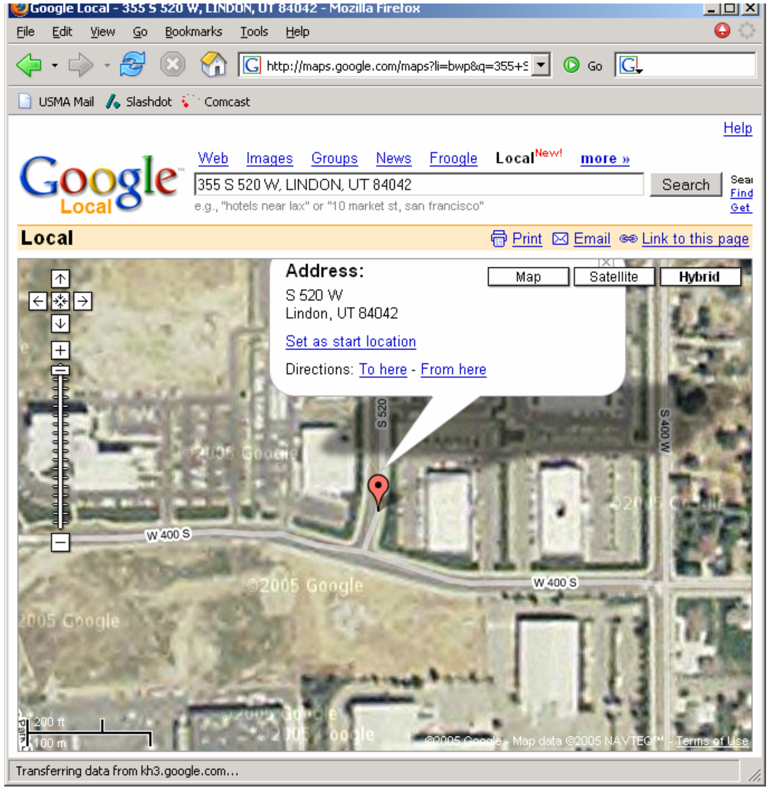
(801) 765-4999

Business Phonebook Results **1** - **1** of **1** for **(801) 765-4999**. (**0.02** seconds)

Caldera Systems Inc - (801) 765-4999 - 355 S 520 W, Lindon, UT 84042 - Google Maps - Yahoo! Maps - MapQuest

Free! Get the Google Toolbar. Download Now - About Toolbar

(801) 765-4999


Google Local - 355 S 520 W, LINDON, UT 84042 - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://maps.google.com/maps?li=bwp&q=355+S
Go
USMA Mail Slashdot Comcast
Help
Google Local
Web Images Groups News Froogle Local New! more »
Search
Local
Print Email Link to this page
Map Satellite Hybrid
S 520
S 400 W
W 400 S
W 400 S
©2005 Google
200 ft
100 m
©2005 Google - Map data ©2005 NAVTEQ™ - Terms of Use
Transferring data from kh3.google.com...

355 S 520 W, LINDON, UT 84042

e.g., "hotels near lax" or "10 market st, san francisco"

Address:
S 520 W
Lindon, UT 84042

Set as start location

Directions: To here - From here

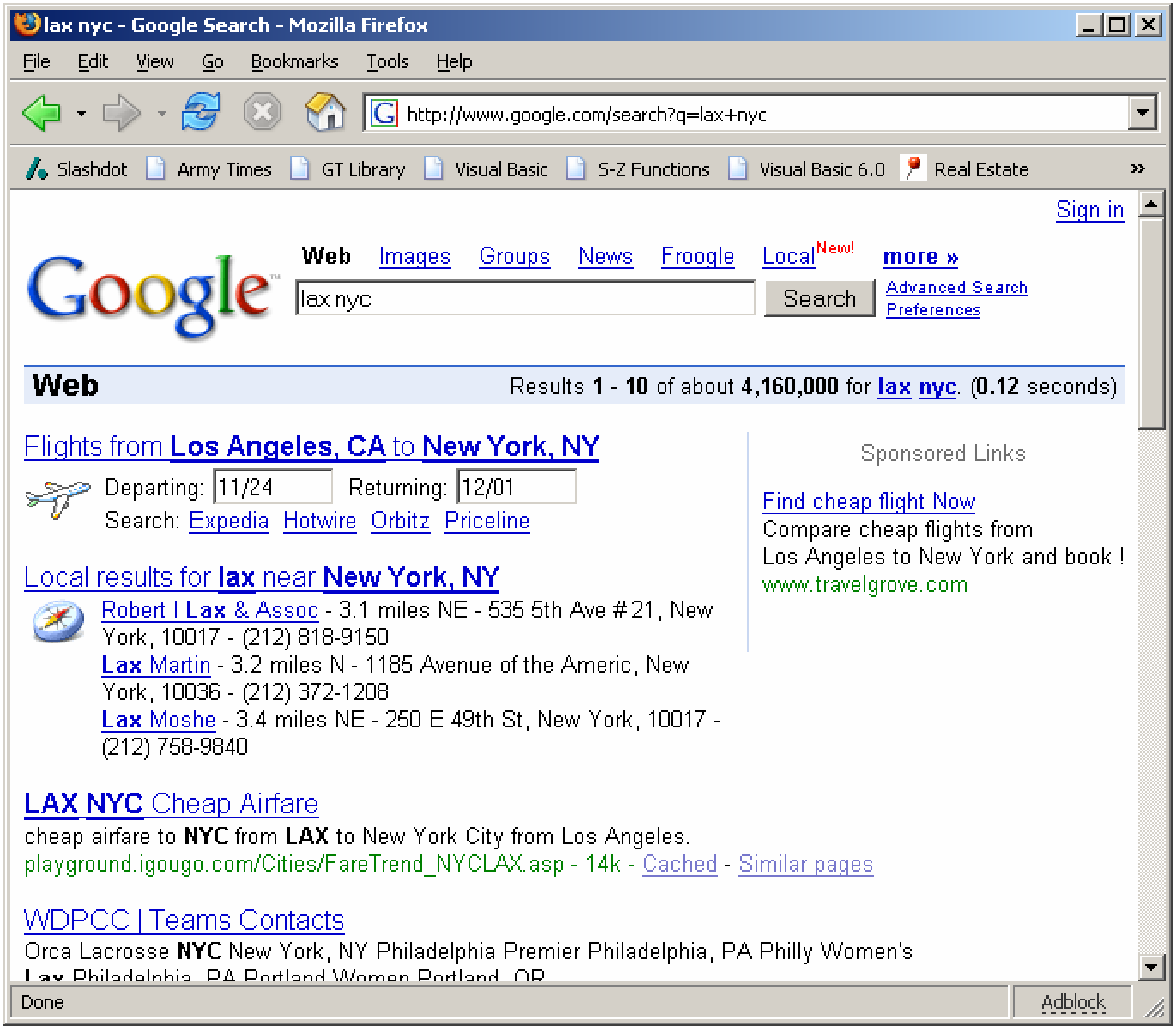
a search returns…

- name
- phone number

-address

-links to 3 mapping services

Travel Support



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: 11/24 Returning: 12/01

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

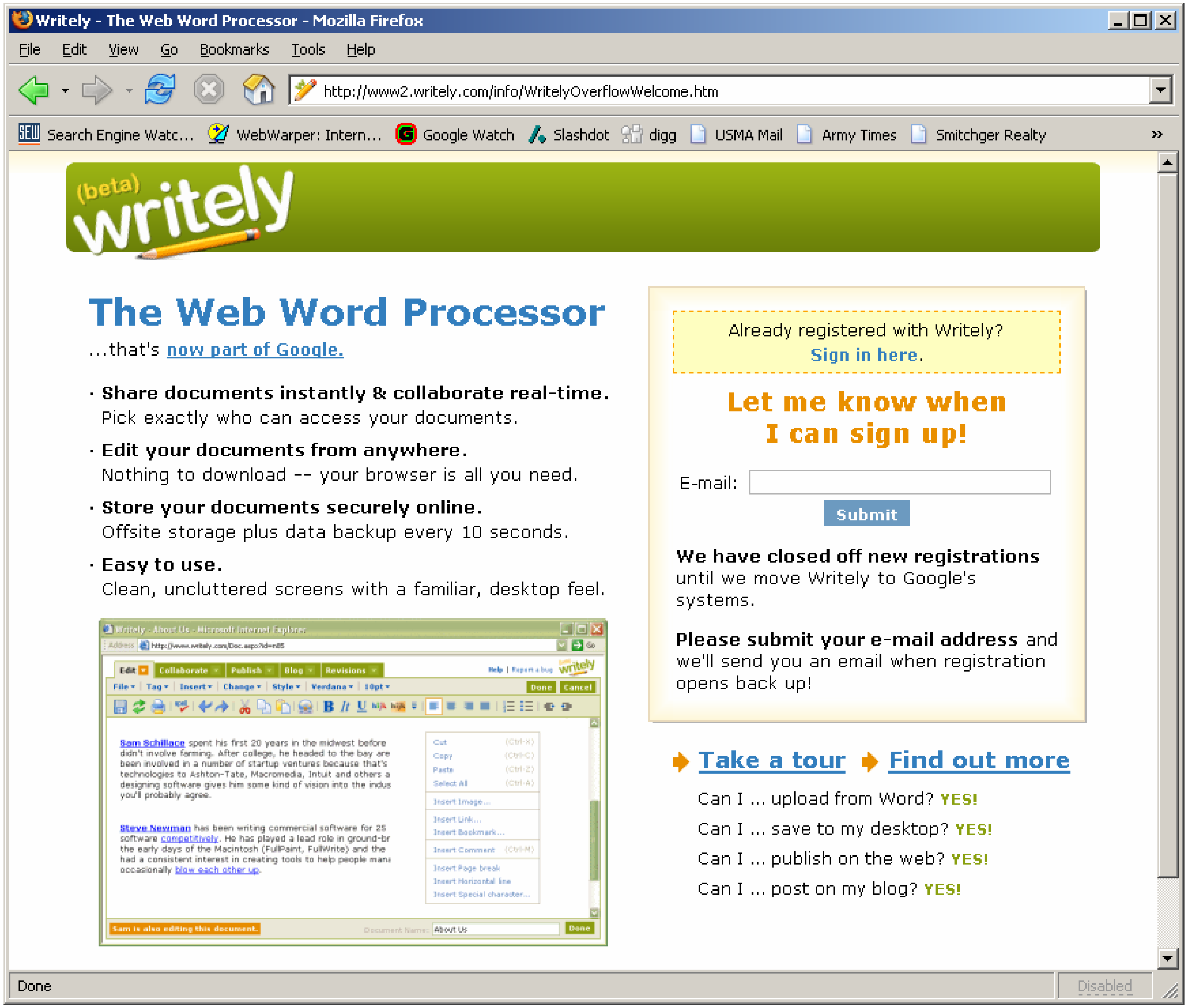
Sponsored Links

Find cheap flight Now

Compare cheap flights from Los Angeles to New York and book !

www.travelgrove.com

Collaborative Word Processing


Writely - The Web Word Processor - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www2.writely.com/info/WritelyOverflowWelcome.htm
Search Engine Watc... WebWarper: Intern... Google Watch Slashdot digg USMA Mail Army Times Smitchger Realty
(beta) writely

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.

Sam Schillace spent his first 20 years in the midwest before didn't involve farming. After college, he headed to the bay are been involved in a number of startup ventures because that's technologies to Ashton-Tate, Macromedia, Intuit and others a designing software gives him some kind of vision into the indus you'll probably agree.

Steve Newman has been writing commercial software for 25 software competitively. He has played a lead role in ground-br the early days of the Macintosh (FullPaint, FullWrite) and the had a consistent interest in creating tools to help people manu occasionally blow each other up.

Sam is also editing this document.

Document Name: About Us

Already registered with Writely?
Sign in here.

Let me know when I can sign up!

E-mail:

Submit

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!**

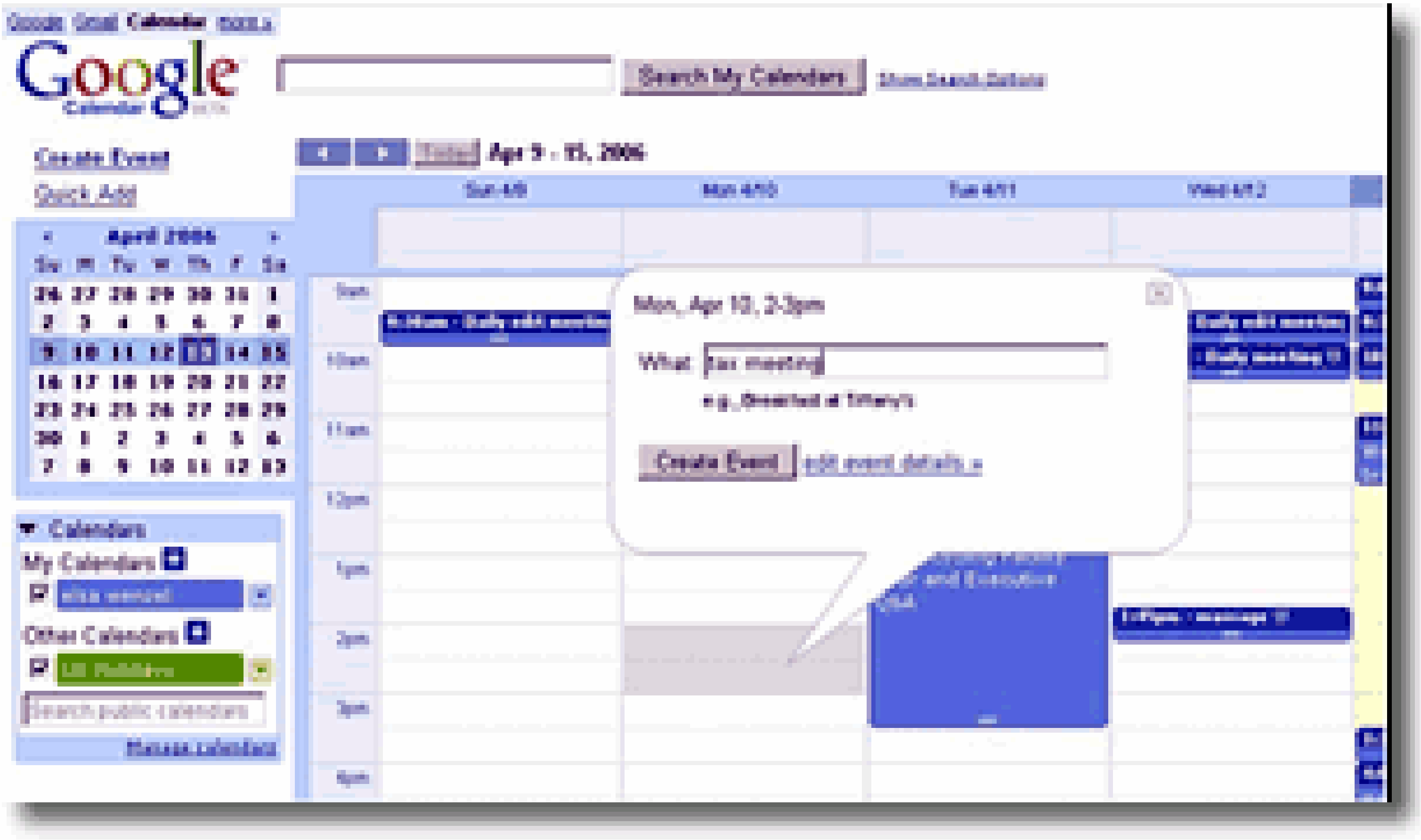
Can I ... publish on the web? **YES!**

Can I ... post on my blog? **YES!**

Done

Disabled

Google Calendar



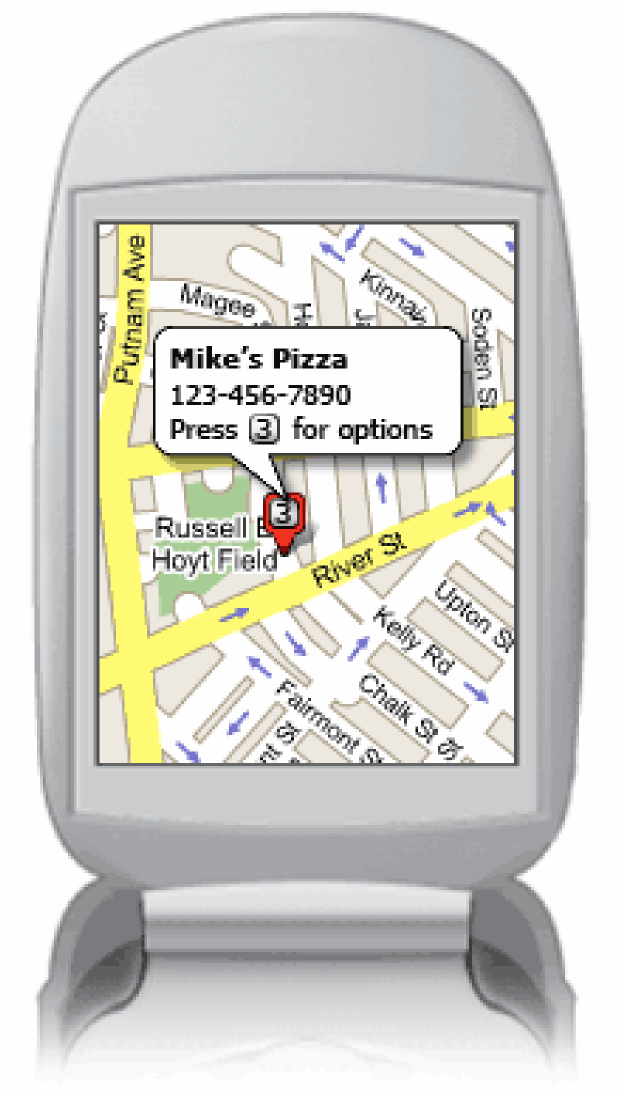
Mon, Apr 10, 2-3pm

What tax meeting

e.g., Breakfast at Tiffany's

http://www.zdnet.co.uk/i/z/rv/2006/04/google-calendar-330x225.gif

Local Information for Local Devices

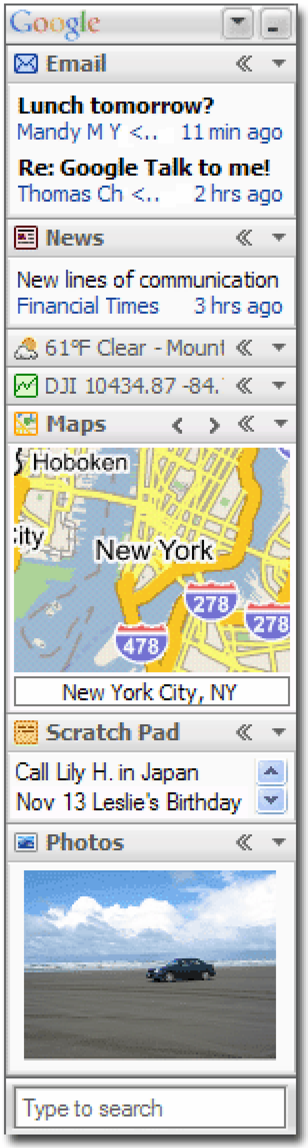


- Local for Mobile
  - detailed directions
  - search results integrated with map
  - zoom in/out, drag maps
  - satellite imagery
- free download

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

Google Desktop

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



Lunch tomorrow?
Mandy M Y <.. 11 min ago

Re: Google Talk to me!
Thomas Ch <.. 2 hrs ago

New lines of communication
Financial Times 3 hrs ago

61°F Clear - Mount

DJI 10434.87 -84.

New York City, NY

Call Lily H. in Japan
Nov 13 Leslie's Birthday

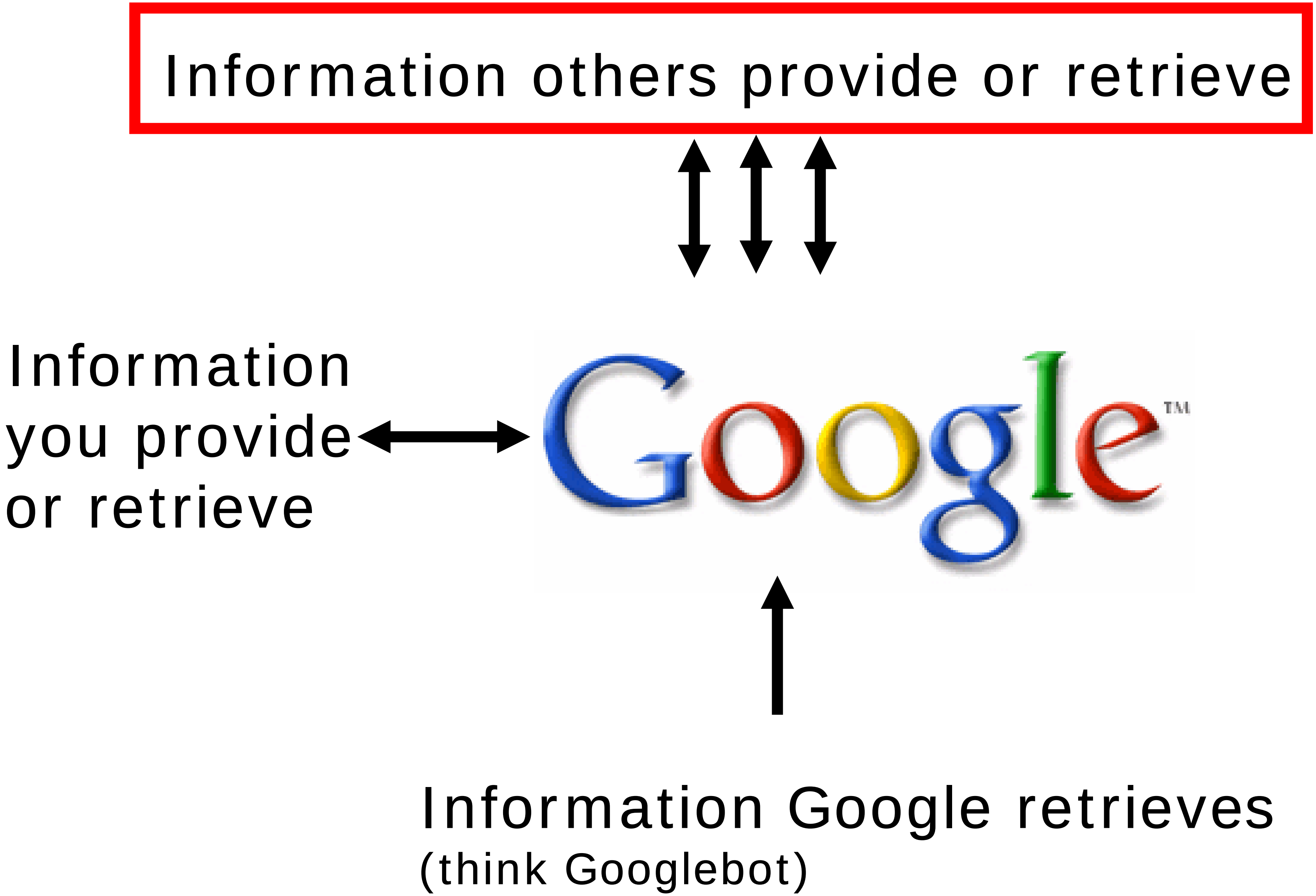
Type to search

http://desktop.google.com/

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.

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



Information others provide or retrieve

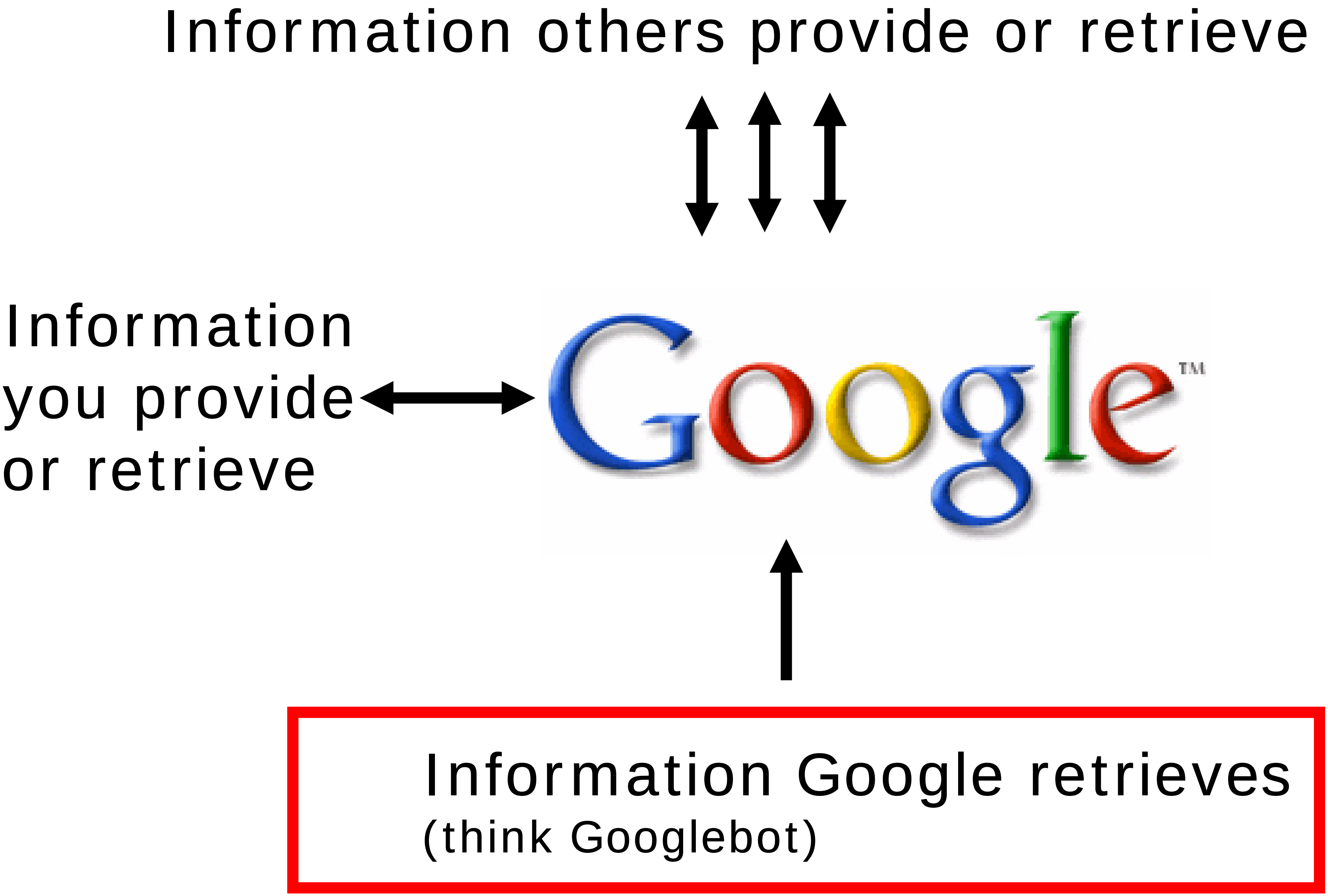
Information you provide or retrieve

Information Google retrieves
(think Googlebot)

Keyphrases

(from Google.com)

- greg conti
- gregory conti
- a framework for countering denial-of-information attacks
- ids rainstorm
- network security visualization
- 124th mi bn
- awareness teach information security filetype:pdf
- passive visual fingerprinting of network attack tools
- usma ia
- 92d information warfare aggressor squadron
- conti gatech
- greg conti rumint
- hacking and innovation cacm

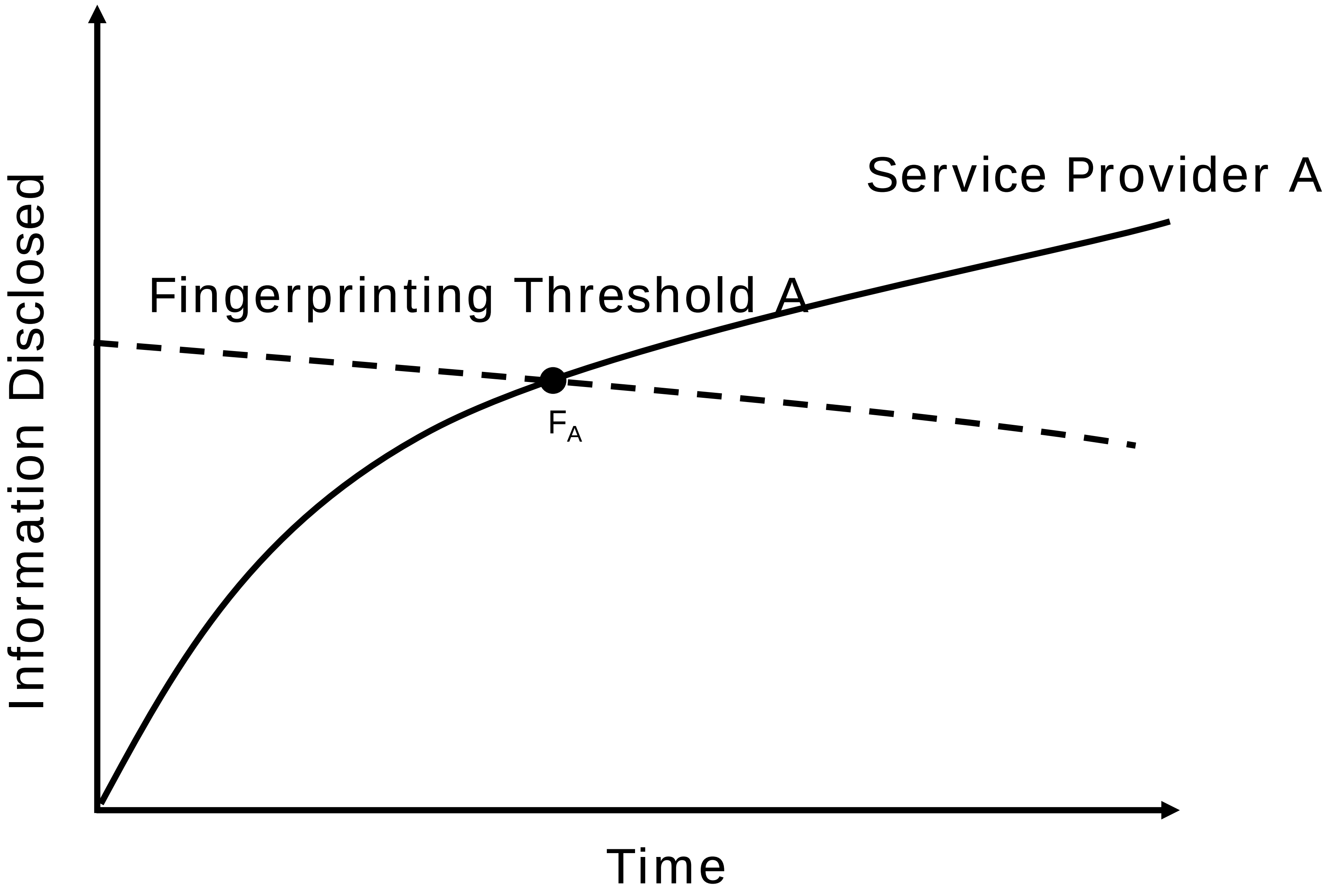


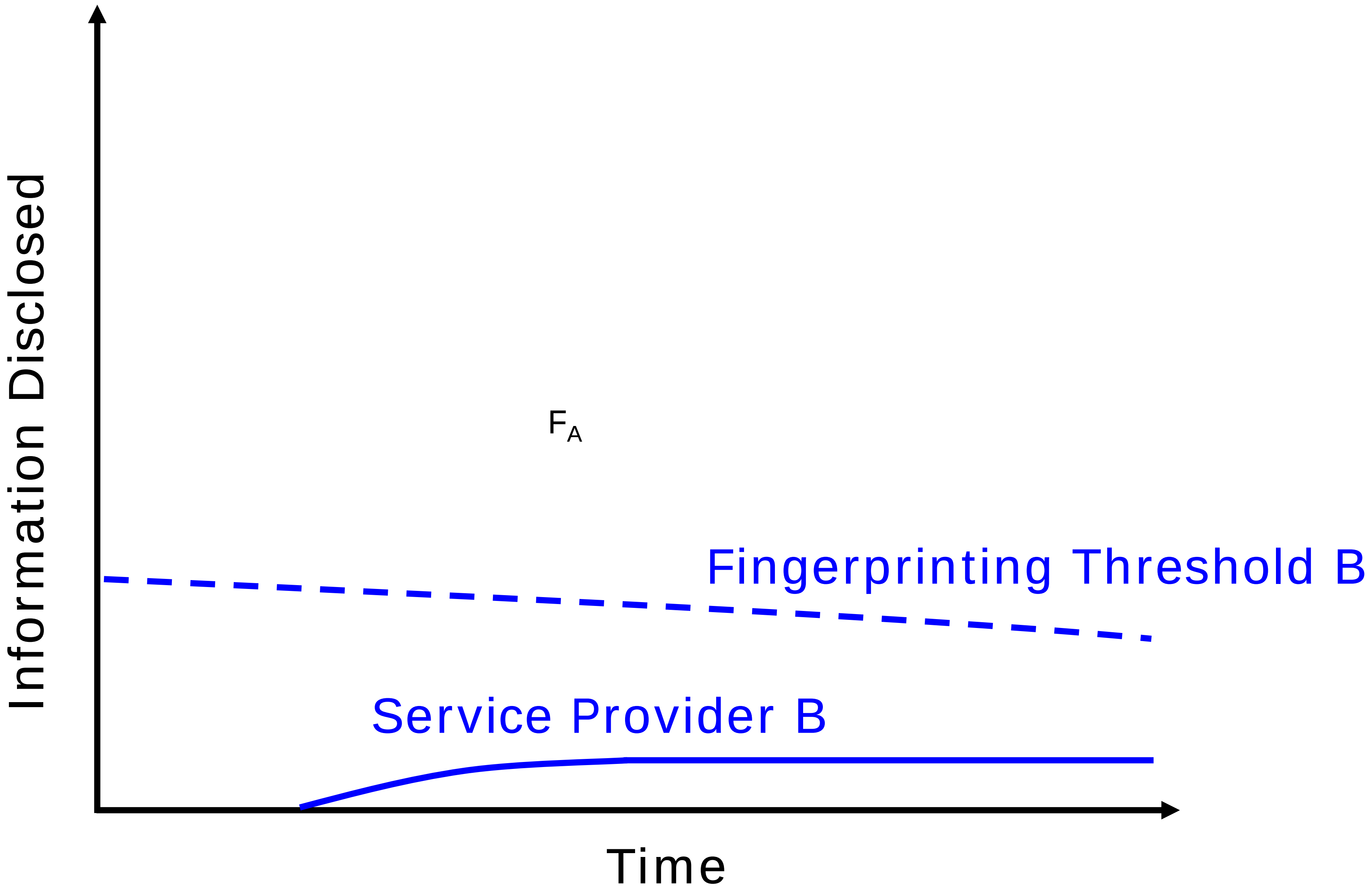
Information others provide or retrieve

Information
you provide
or retrieve

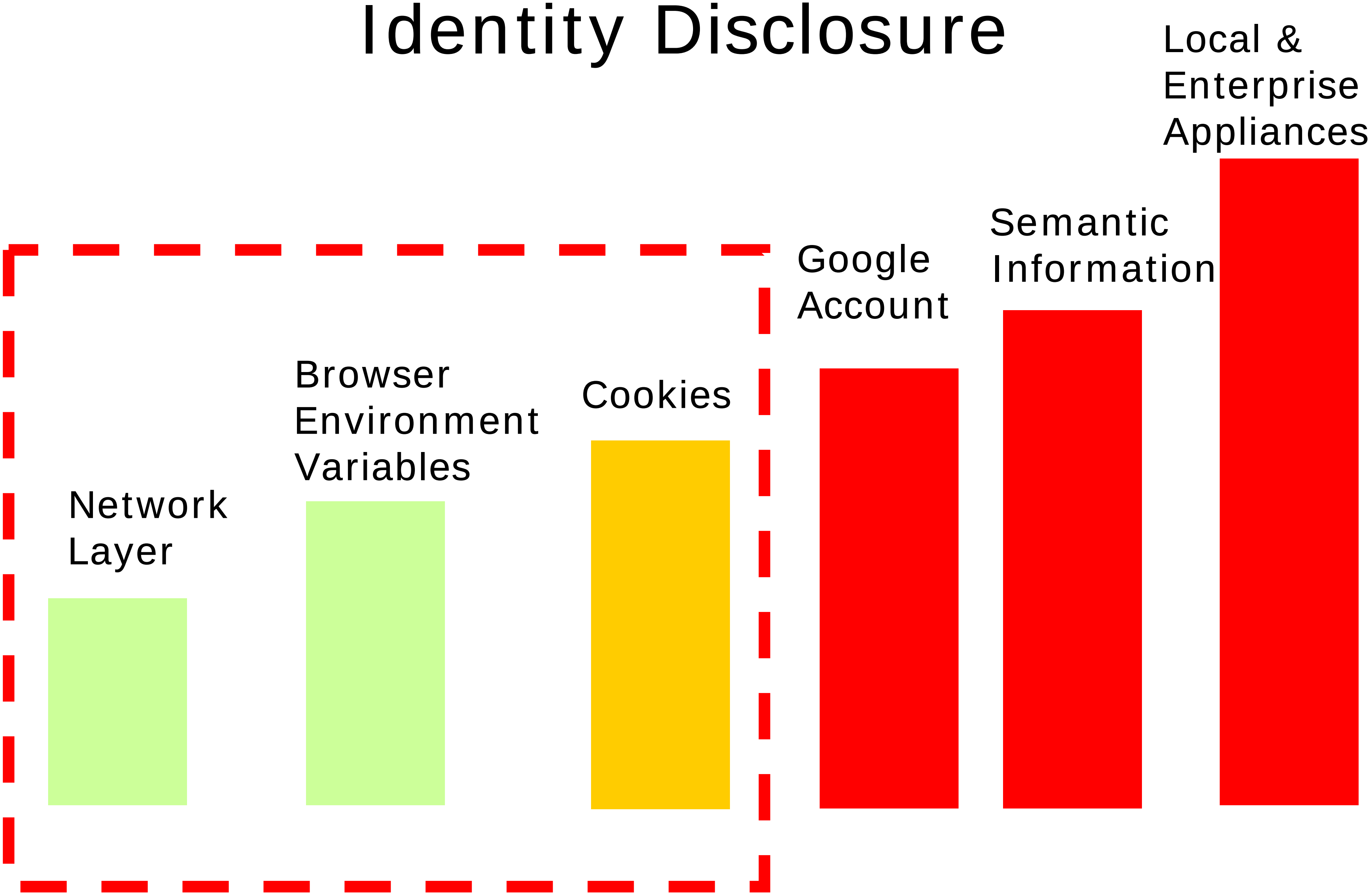
Information Google retrieves
(think Googlebot)

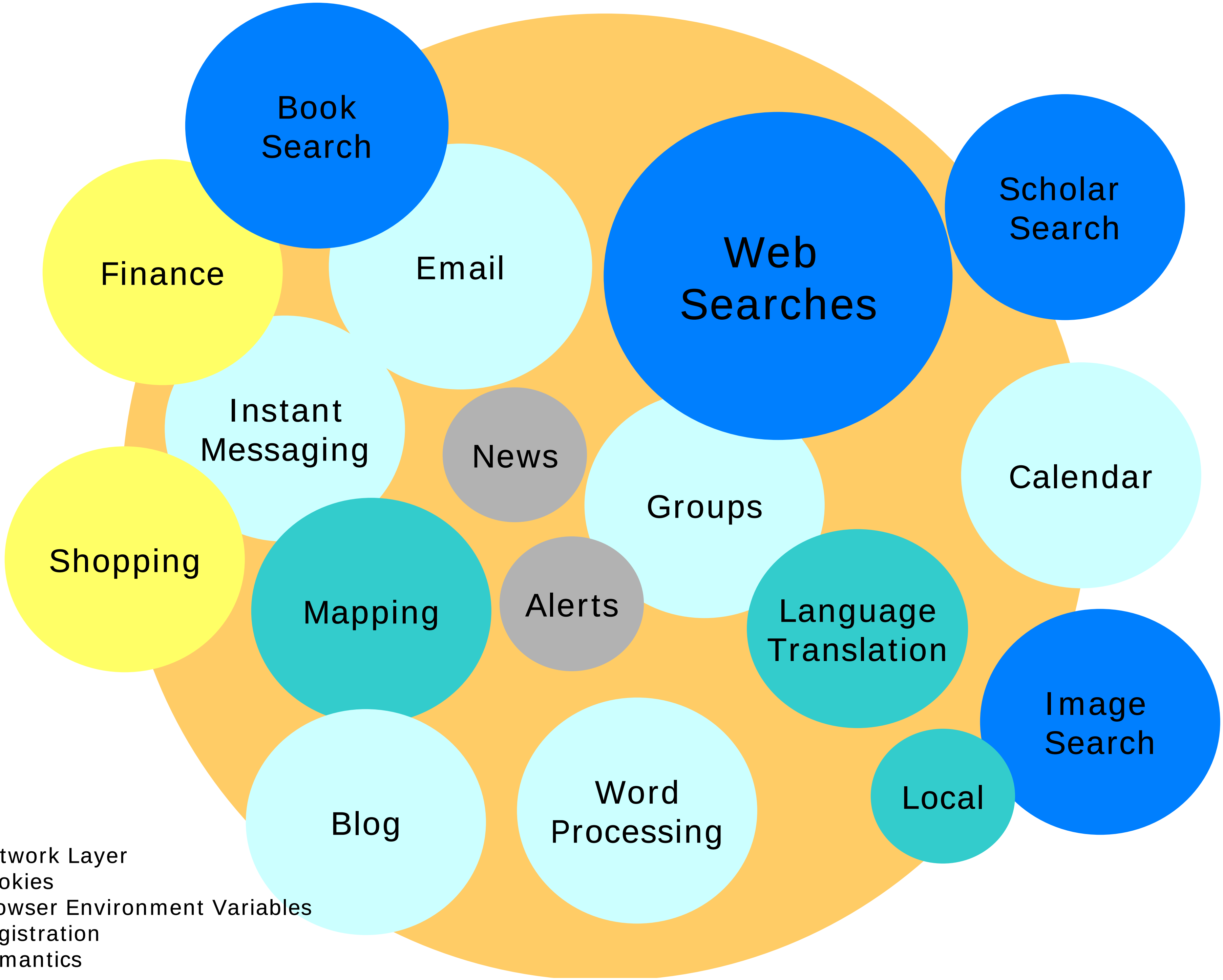
Fingerprinting


Service Provider A
Fingerprinting Threshold A
F_A
Information Disclosed
Time


Information Disclosed
F_A
Fingerprinting Threshold B
Service Provider B
Time

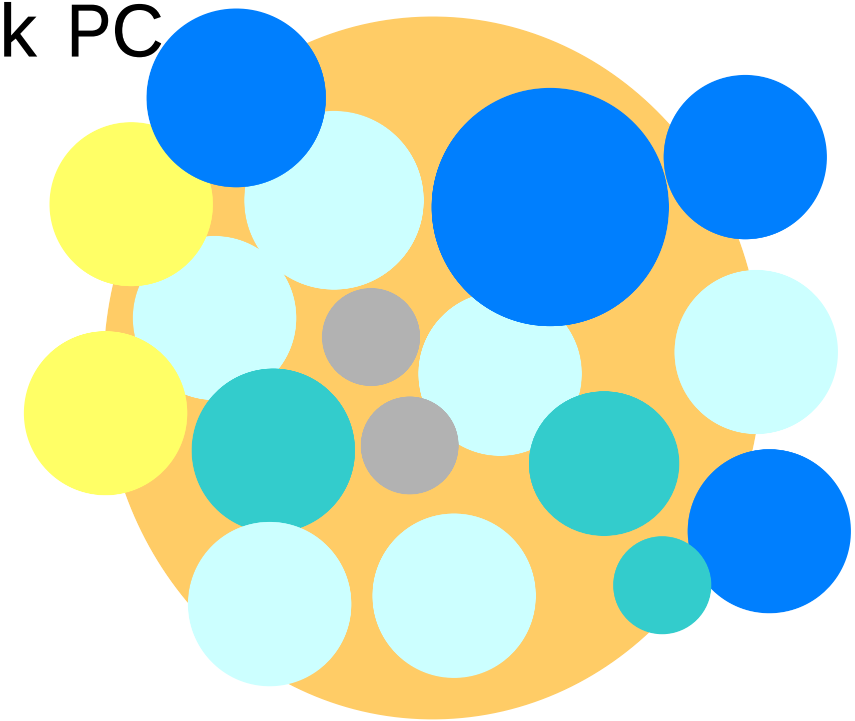
Identity Disclosure



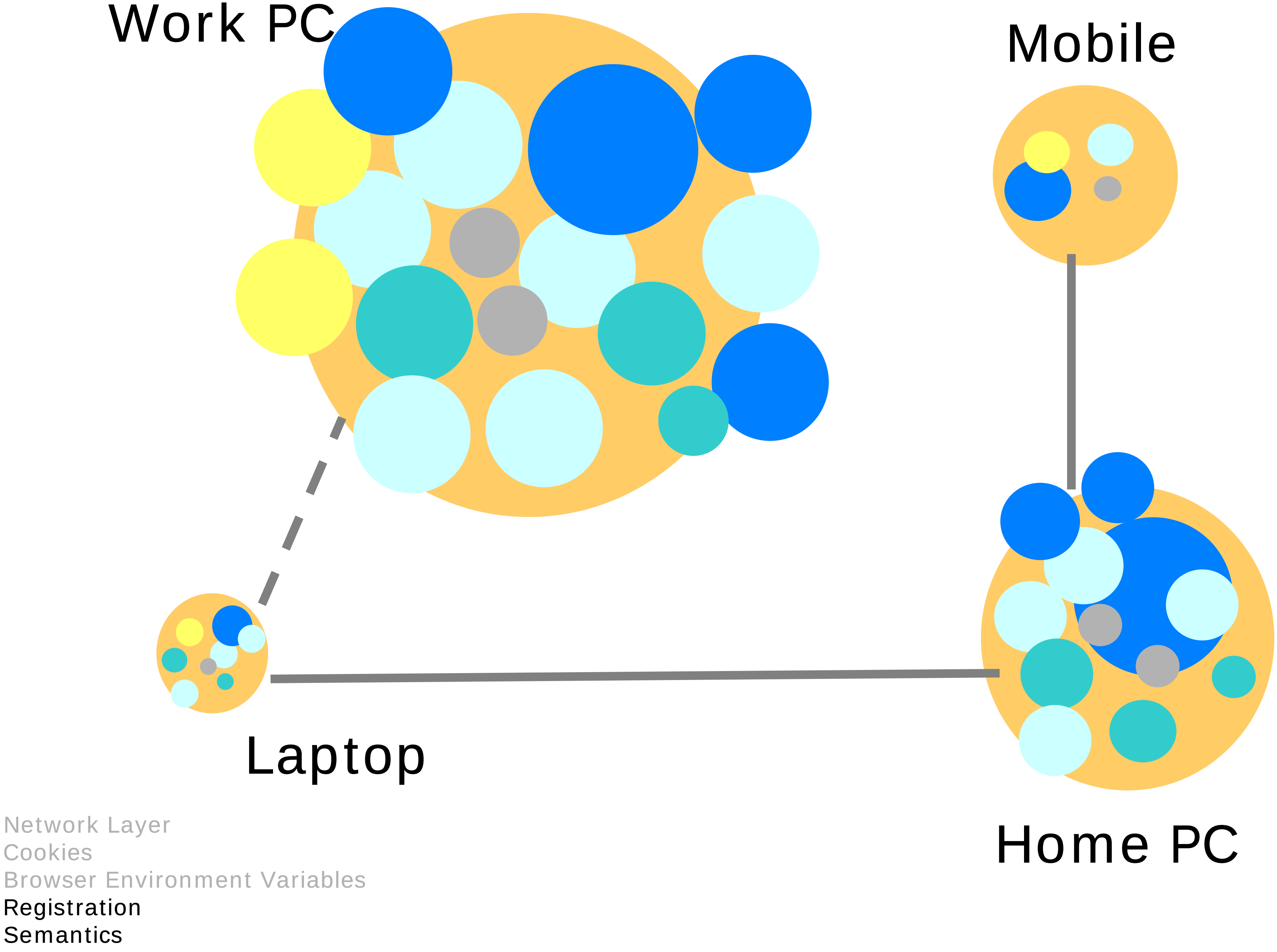


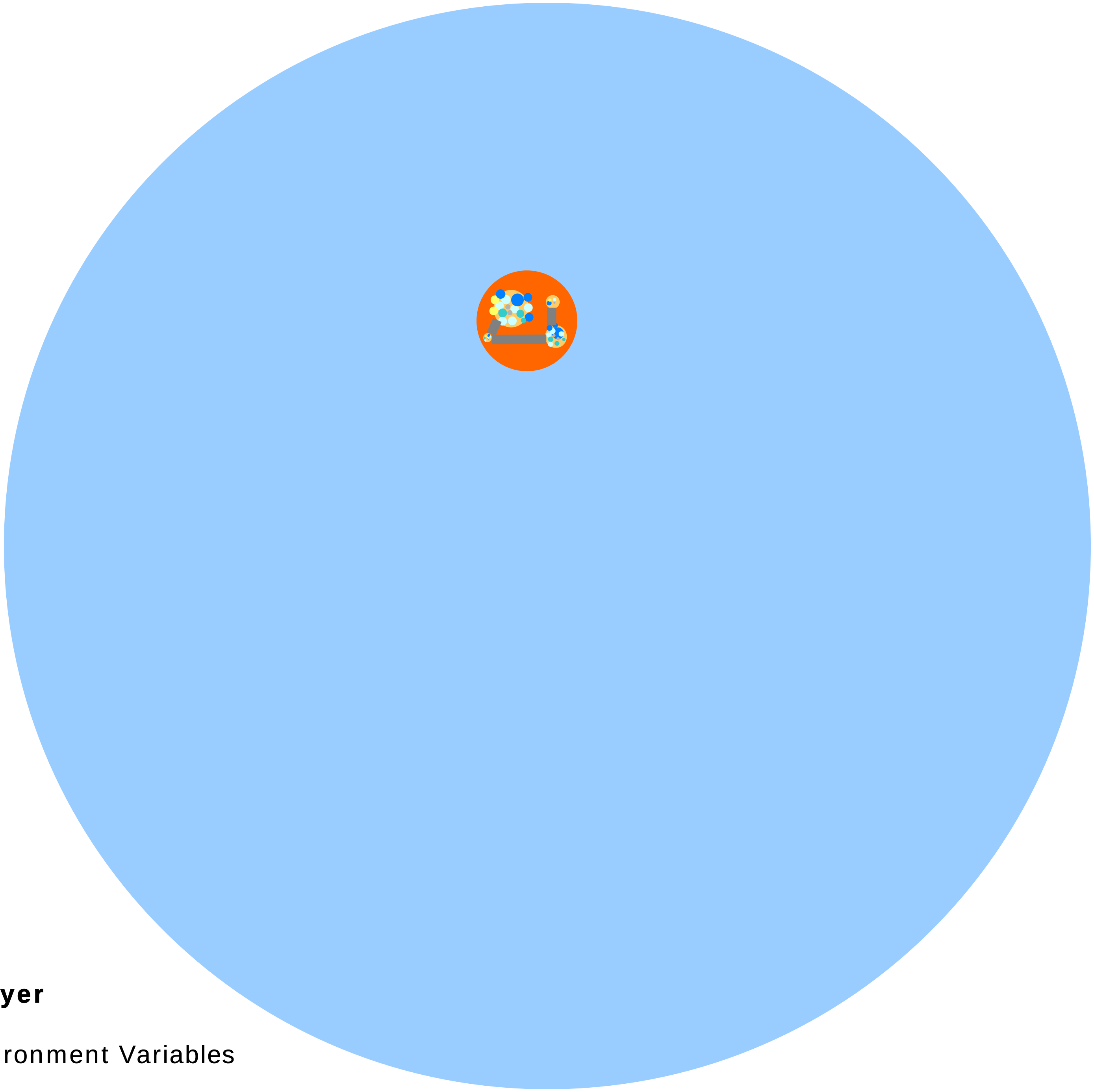
Network Layer
Cookies
Browser Environment Variables
Registration
Semantics

Work PC



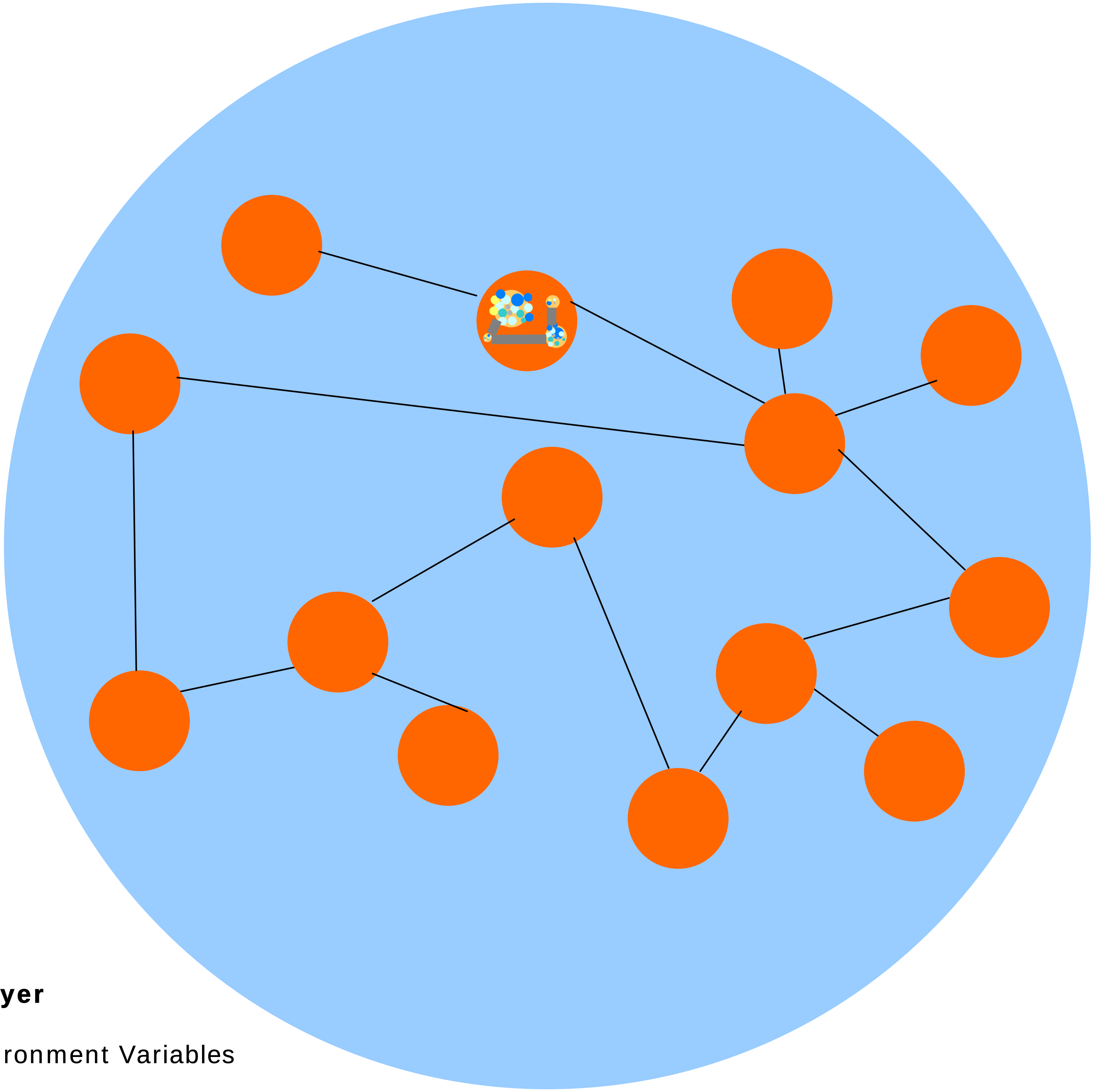
Network Layer
Cookies
Browser Environment Variables
Registration
Semantics


Work PC
Mobile
Laptop
Home PC
Network Layer
Cookies
Browser Environment Variables
Registration
Semantics



Network Layer
Cookies
Browser Environment Variables
Registration
Semantics

Company
Organization
Social Network



Network Layer
Cookies
Browser Environment Variables
Registration
Semantics

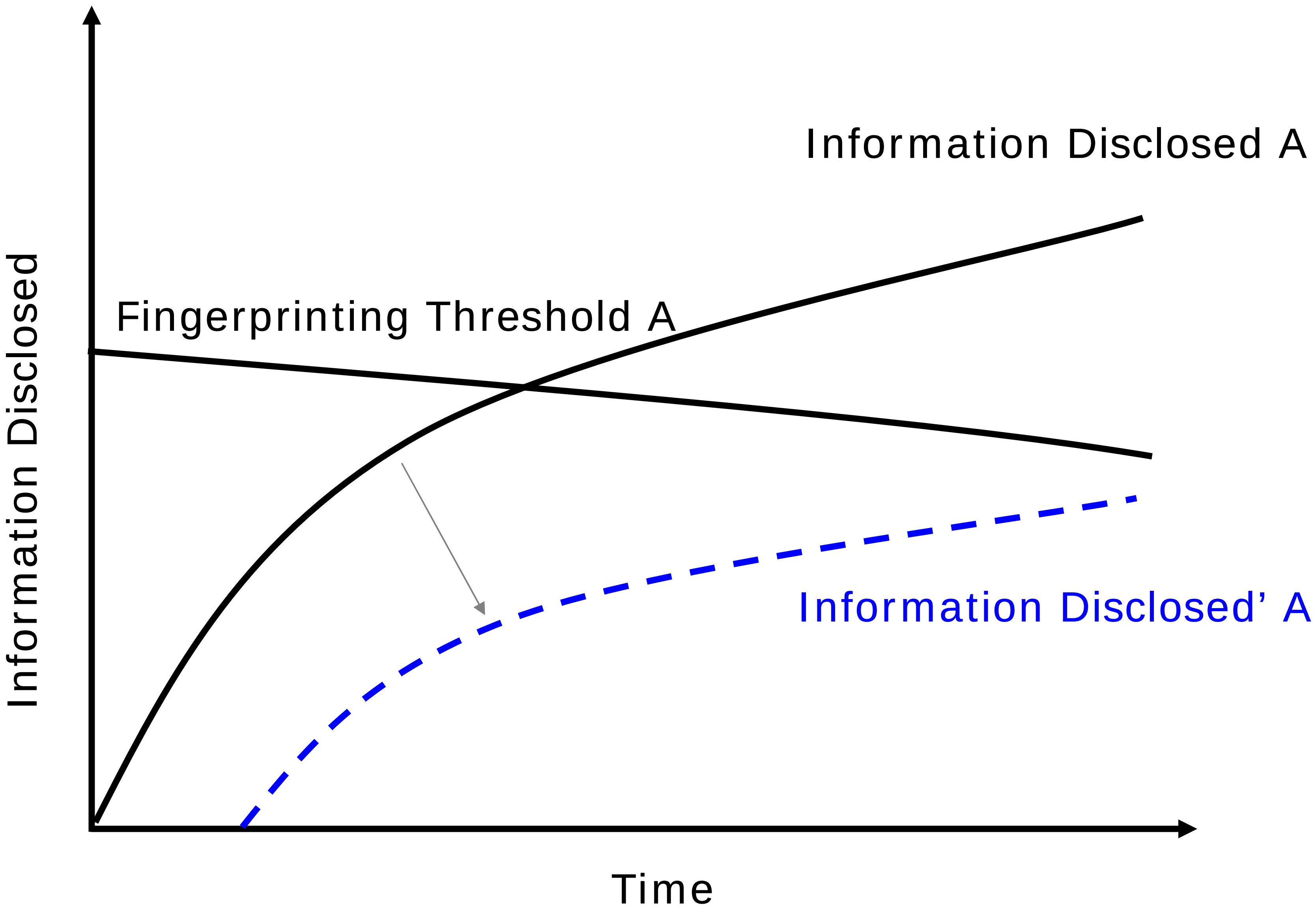
Company
Organization
Social Network

I have a Google cookie on my computer.

My Employees have Google cookies on their computers.

I have registered for a Google account.

Countermeasures


Information Disclosed A
Fingerprinting Threshold A
Information Disclosed’ A
Information Disclosed
Time

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

Countermeasures

- Multiple email accounts, host own email
- Chaffing
- DHCP/NAT
- Retain only the “necessary” amount of information
- Time limits
- Transparency
- Opt-in / out

I actively manage/remove cookies.

I regularly use web anonymization tools.

Examples

Co-worker example #1

- 696 entries
- 18 Feb 05 - 26 Jul 06
  - 1.33 unique searches / day
- Sanitization criteria
  - names of family, close friends
  - addresses
  - credit card numbers
  - SSN's

5%

sanitization

Co-worker example #2

- 1264 entries,
- 9 Nov 05 - 31 Jul 06
    - 4.8 unique searches / day
- criteria
    - names
    - locations close to home
    - phone numbers
    - friends blog

24%

sanitization

Co-worker example #3

- 3974 entries,
- 8 Dec 04 - 31 Jul 06
  - – 6.6 unique searches / day
- criteria
  - names
  - friends
  - phone numbers
  - co-workers

31%

sanitization

Demo

Discussion