



What's Google Really Up to?

Roger Clarke

Xamax Consultancy Pty Ltd, Canberra

Visiting Professor in eCommerce at Uni. of Hong Kong
and in Cyberspace Law & Policy, U.N.S.W. & Comp. Sci., A.N.U.

<http://www.anu.edu.au/people/Roger.Clarke/.../II/GoogleHK-0611 {.html, .ppt}>

eCom/iCom Uni. of Hong Kong – 2 November 2006

Copyright 2005-06



1

What's Google Really Up To? Agenda

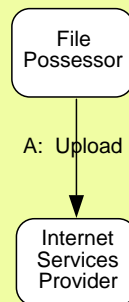
- **Google's Context**
 - The Web
 - Search-Engines
- **Google's Business(es)**
 - Content-Discovery Services
 - Content Services
 - Data about Users
- **Perspectives on Google**
 - Business Model
 - Copyright Law
 - Consumer Law and Practice
 - Privacy Law and Practice

Copyright 2005-06



2

The Web Process

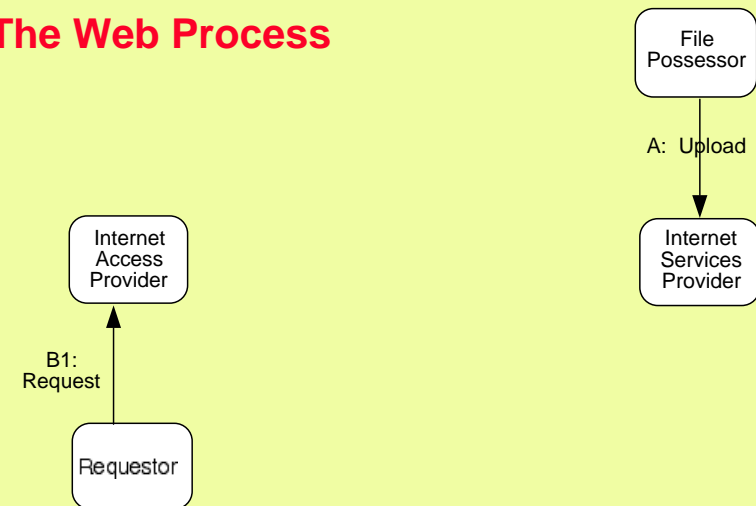


Copyright 2005-06



3

The Web Process

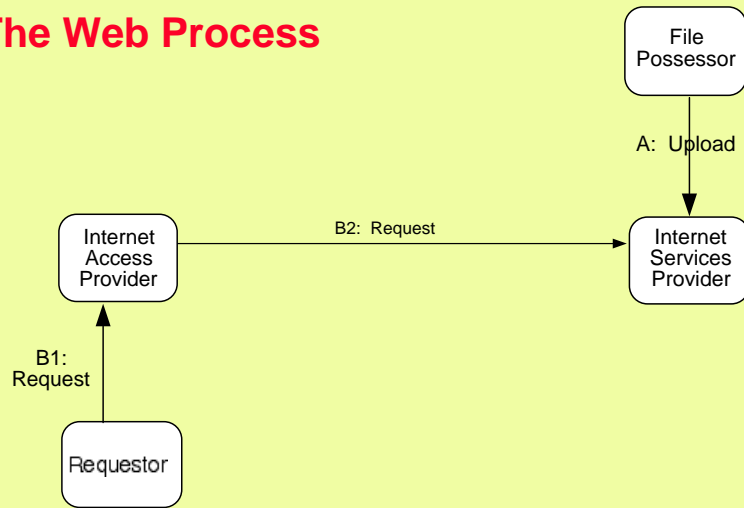


Copyright 2005-06

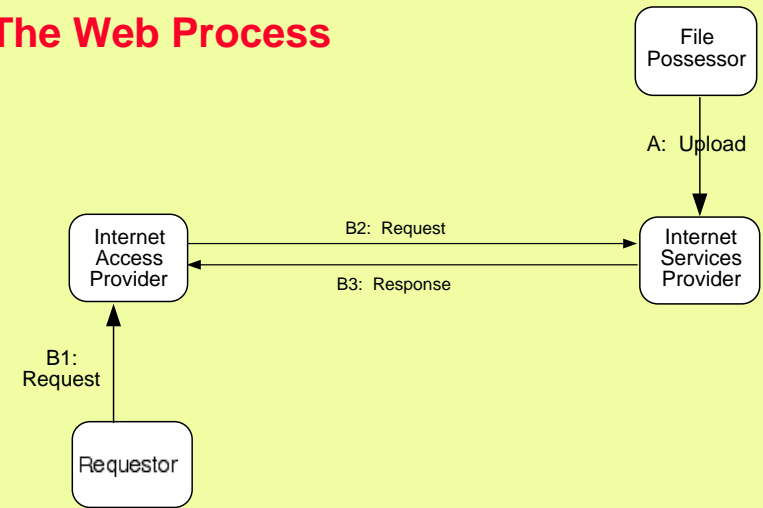


4

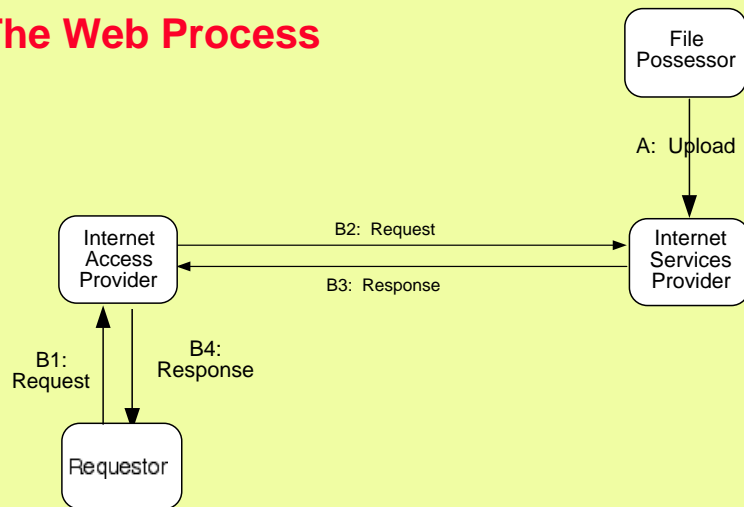
The Web Process



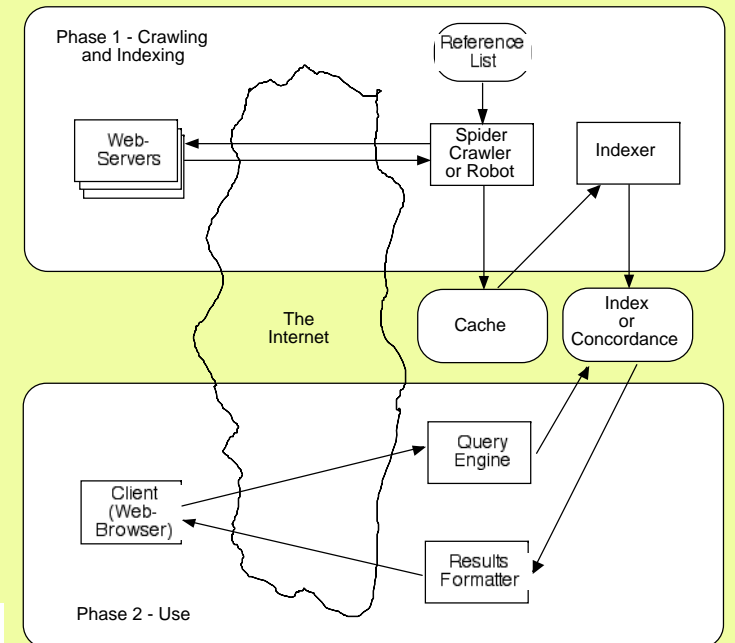
The Web Process



The Web Process



Search Engine Operation



Google's Business(es)

1. Content Discovery Services

- The Largest Coverage (size of the Reference List)
- The Smartest Precedence Algorithm (the sorting part of the Results Formatter)
- The Fastest, Simplest, Best? Search-Service (a UI for normal people, not specialists)
- Multiple Constrained Searches (images, blogs, Froogle)
- Multiple Extension (Froogle, Scholar)



Scholar)

Copyright
2005-06



froo-gle (fru'gal) n. Smart shopping through Google 9

Google's Business(es)

2. Content Services

- Google Earth
- Google Base
- Google Video / YouTube
- ...
- Google News
- Google Library / Print
- ...



Copyright
2005-06



10

Is Google a Publisher ?

- Q. How will **Google** control for irrelevant items like pornography or SPAM material?
- A. Items that don't meet Google policies may be taken down based on **our own checks** or in response to user complaints
- Members of **YouTube** rate videos they like, and **YouTube reviews highly-rated and recent videos for consideration in 'Today's Featured Videos'**.

In addition, YouTube takes recommendations from members at editor@youtube.com and **YouTube scans the site for videos of interest**

Copyright
2005-06



11

Google's Business(es)

3. Data about Users

"We are moving to a Google that knows more about you"
- **Google's CEO**

- **Round 1**
 - Search-Terms
 - IP-address(es)
 - Click-Trail
 - Click-Throughs

Copyright
2005-06



12

Google's Business(es)

3. Data about Users

“We are moving to a Google that knows more about you”
- Google's CEO

Round 1

- Search-Terms
- IP-address(es)
- Click-Trail
- Click-Throughs

Round 2

- Google Accounts
- Personalizing Services
- Email-Address as Username
- A Common Cookie

Copyright
2005-06



13



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

Copyright
2005-06



14

Gmail Privacy Page

[FAQs](#) | [Resources](#)

Frequently Asked Questions

1. What is Gmail and what privacy risks does it raise?

- [1.1 What is Google's Gmail?](#)
- [1.2 What is your position on Google's Gmail?](#)
- [1.3 What privacy risks are presented by Gmail?](#)
- [1.4 When did this issue arise, and what has happened since then?](#)
- [1.5 What other things has Google been doing that might affect my privacy?](#)

2. Technical details about Gmail

- [2.1 How does Google's "content extraction" work?](#)
- [2.2 What is "internal" and "external" e-mail information used in the analysis?](#)
- [2.3 Will Gmail build profiles of subscribers and/or non-subscribers?](#)
- [2.4 Why is Gmail different than spam filtering?](#)
- [2.5 It's a computer, not a person reading your e-mail. What's the big deal?](#)
- [2.6 What patents has Google filed for Gmail?](#)

3. Legal details

- [3.1 What are the Federal wiretapping laws, and does Gmail implicate them?](#)
- [3.2 What is California's wiretapping law, and why does Gmail implicate it?](#)
- [3.3 Is there a "service provider exception" under California wiretapping law?](#)
- [3.4 What legal objections have been raised in other countries?](#)

4. What can you do?

- [4.1 Don't sign up for Gmail](#)
- [4.2 Don't send e-mails to @gmail.com addresses](#)
- [4.3 Reduce the possibility of tracking you through your cookies](#)
- [4.4 What are YOU guys doing about it?](#)



v. 1 – October 2004

Search Within Your Own Computer

“A desktop search application that provides full text search over your email, files, music, photos, chats, Gmail, web pages that you've viewed, ...”

(cf. Apple's Sherlock 1998, later Spotlight, and many third-party products for Wintel)

It allows people to scan their computers for information in the same way that they use Google to search the web

Copyright
2005-06



<http://desktop.google.com/about.html>

15



v. 3 – 9 Feb 2006

Search Across Your Computers

BUT

“In order to share your indexed files between your computers, we securely transmit this content to Google Desktop servers located at Google”

cf. MS Passport data, centralised at Redmond WA

Copyright
2005-06



<http://desktop.google.com/features.html#searchremote>

16

Google's Business(es)

3. Data about Users

“We are moving to a Google that knows more about you”
- Google's CEO

Round 3

- Gmail
- Desktop
- Desktop v.3
- Orkut

Round 4

- Google as Wireless IAP
Gratis (i.e. ad-funded)
Today Mountain View
Shortly San Francisco
Tomorrow the World?

Copyright
2005-06



17

Google as Wireless Internet Access Provider

October 04, 05

Google to unwire San Francisco?

IF GOOGLE GETS INTO ACCESS, CAN IT TAKE ON THE OPERATORS?

By Stephen Lawson and Stacy Cowley, IDG News Service

Google's expansion into public Wi-Fi could be a boon to the embattled municipal broadband movement and presents intriguing new possibilities for the search giant, though putting the traditional carriers out of business isn't likely to be one of them.

Acceptance of Google's tender confirmed 5 April 2006

Copyright
2005-06



<http://www.techworld.com/mobility/features/index.cfm?featureid=1837>

18

What's Google Really Up To?

Some Perspectives

- Business Model
- Copyright Law
- Consumer Law and Practice
- Privacy Law and Practice

Copyright
2005-06



19

Business Model

Who pays, what for, to whom, and why?

- Nil Revenue from Consumers (so far)
- **Old Model:**
The Advertiser Pays for Ads Displayed
- **New Model:**
The Advertiser Does Not Pay for Ads Displayed
The Advertiser Does Pay for 'Click-Throughs'
And the Advertiser Pays for AdWords
More Advertisers Pay More, More Often
Market-share is being won from other Channels

Copyright
2005-06



20

AdWords

<http://adwords.google.com>

<http://www.tomw.net.au/technology/it/adwords.shtml>

- Advertisers bid and pay for keywords (metadata)
- These are used to target the placement of ads on the pages that Google delivers to searchers

AdSense

<http://adsense.google.com>

<http://www.tomw.net.au/technology/it/adsense.shtml>

- Page-Owners let Google place Ads on their Pages
- The Page-Owner gets some of Google's revenue

Copyright
2005-06



21

Sources of Google's Revenue Growth

- **Scale of Google's Operations**
 - more Google searchers
 - more Google pages published to them
 - more ads per page
- **Acquisition of New Operations**
 - popular web-sites, new facilities
- **'Syndication' to Other People's Operations**
 - other people's pages
 - costs a small share to the other people

Copyright
2005-06



22

Google's Revenue Generation (\$5bn in the 12 mths to June 2006) Key Features

- Positive Feedback / Network Economics
i.e. Transactions breed Movable Data
- Cross-Leveraging among Elements of
the Google Product Range ('Platform')
- Use of the Treasure-Chest for Wildcat
Acquisitions, in search of 'network effect'
- Ad Pricing is based on Auctions whose
processes are attuned to Google's interests

Copyright
2005-06

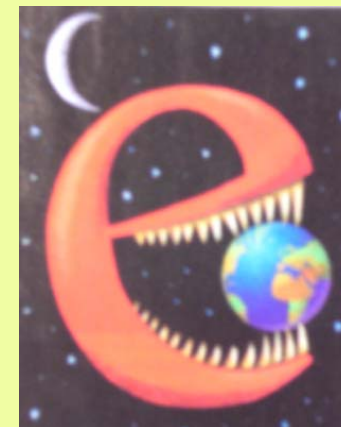


Hardy Q. 'The Google Industrial Complex' Forbes 16 Oct 2006
http://www.forbes.com/forbes/2006/1016/108_print.html

23

Google's Challenges to Competition Law ?

Search-Engine Market Dominance ?
Big Enough to Challenge Microsoft ?



Copyright
2005-06



24

Google's Challenges to Trademark Law

Hire of 'Adwords' enables advertisers to gain priority display-space when particular search-terms are nominated by users

Trademarked terms are hired out as Adwords

Some cases have already found against the practice, and clarification is awaited

Google's Challenges to Copyright Law

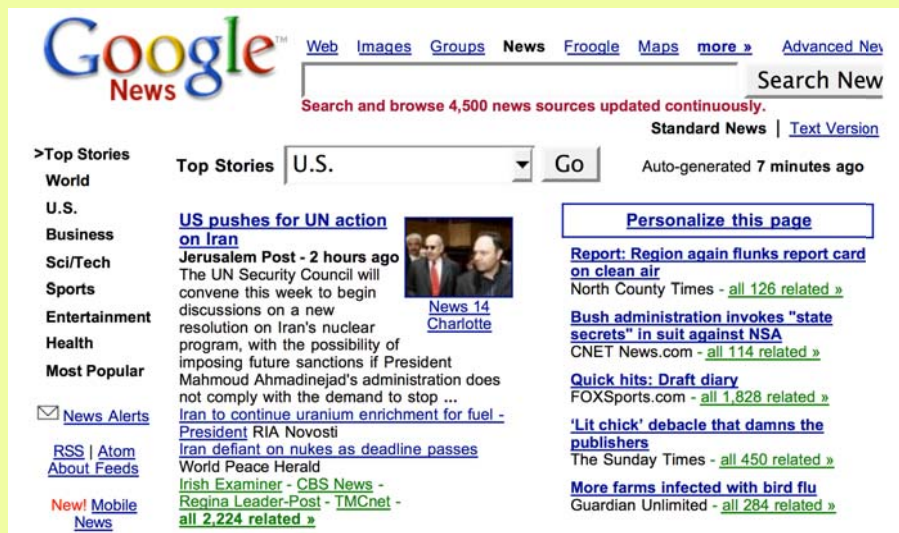
Normal Search-Engine Operations

'Web operations ==>> implied licence'

For web-spider extraction, indexing, caching ?

But for **widespread access to Google Cache** ??

Is Google's respect for exclusion clauses in robots.txt files sufficient defence?



Google's Challenges to Copyright Law Features of Google News

- News report consolidation service
- Offers links to news reports
- Customisation and Alert extensions
- Displays headline, source and hotlink, some text, and in some cases thumbnails, reproduced from Google's Cache
- **Agence France Press (AFP) have sued**
- AFP has gained support from the Paris-based World Association of Newspapers (WAN-Press)

Google's Challenges to Copyright Law Thumbnails

- Perfect 10, an adult-oriented Web site featuring "beautiful natural women" in the nude, has shown that Google image search probably infringes copyright law "by creating and displaying thumbnail copies of its photographs"
- **"existing judicial precedents do not allow [even enormous public benefit] to trump" copyright law**

U.S. District Judge A. Howard Matz ruled [17 Feb 06]

Copyright
2005-06



[http://news.com.com/...
Nude-photo+site+wins+injunction+against+Google/...
2100-1030_3-6041724.html](http://news.com.com/...Nude-photo+site+wins+injunction+against+Google/...2100-1030_3-6041724.html)

29

Google Print

5 Partner-Libraries
Bodleian, ..., Uni. of Michigan



Categories of Work

- Out-of-Copyright
- Orphan
- In-Copyright
 - Licensed
 - **Unlicensed**

Potentially Infringing Acts

- Scanning (image)
- Extracting text (OCR)
- Reproduction in the index
- Serving of some from Google Cache to users
- Soft-Copy to the Library who provided it

Copyright
2005-06



30

Google's Challenges to Copyright Law Content Publishing

Google News and Google Library

Two actions in the U.S. District Court, NY:

- Authors Guild v. Google (Sep 2005)
- McGraw Hill & Ors v. Google (Oct 2005)

**'West Coast Code v. East Coast Code'
is no longer just a catchphrase**

Copyright
2005-06



31

The Context Copyright Expansionism Rampant

- Large corporations dependent on monopolies
- Scared by the digital era, they have requested and been granted 'corporate welfare'
- That is constraining socially on consumers, but also economically on innovators
- The U.S. is acting bilaterally and multi-laterally to extend copyright expansionism
- Many Governments are (willingly?) blind to the interests of their own countries

Copyright
2005-06



32

Google's Challenges to Copyright Expansionism

- Opponents have included Barlow (1994), Samuelson (1996), Lessig (2000), Zittrain (2003) and recently Adelphi Charter (2005)
- **Opponents now also include Google**
- 'Old big business' sits fatly, exploits its monopolies and arranges extensions to them
- **'Newly Big Business' adapts quickly to new contexts, realises the potentials lurking inside them, and creates new monopolies that it can dominate from the outset**

Copyright
2005-06



33

Google's Challenges to Consumer Law

Consumer Benefits

- Enormous
- Gratis
- But there is consideration: acceptance of advertising, including intrusive attention-grabbing devices ('blink', popups)

Terms:

- non-negotiable
- non-transparent
- changeable at whim
- not version-managed

Recourse

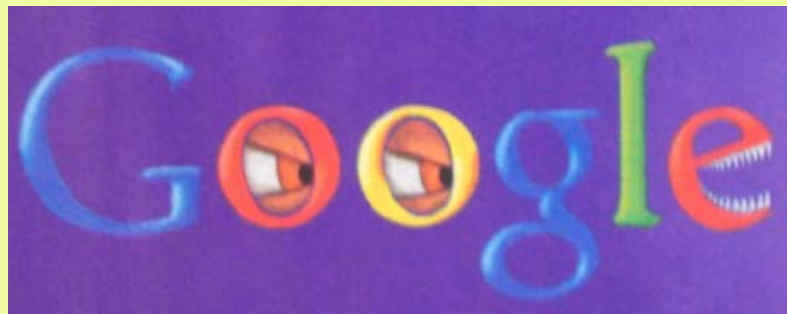
- all-but non-existent

No sign of recovery of lost consumer protections
e.g. WSIS 2005 is vacuous

Copyright
2005-06



34



Copyright
2005-06



35

Google's Challenges to Privacy Law

- Privacy Threats from Search-Engines Generally
- Privacy Threats from Google especially
- The Regulatory Environment(s)
- Google's Privacy Undertakings

Copyright
2005-06



36

'Research Your Next Appointment'

- Their Own Site(s)
- Event Programs
- Committee Minutes
- Court Reports
- Media Reports
 - as subject
 - as commentator
 - as bystander
 - as reporter/contributor /poet
- Letters to the Editor
- Postings
 - email-lists
 - fora
 - blogs
- Logs (e.g. in court)
 - IAPs
 - ISPs
 - own machine
- 'Little Black Books'
- Commercial Databases
- Dead Pages, from the Wayback Machine
- Specialist Sites, e.g. Zoominfo.com

Copyright
2005-06



37

Privacy Threats from Open Information

Discoverability

- Data
- Associations

Second-Round Effects

- More Data Retention
- More Data Capture

Data Quality Problems

- Out-of-Date
- Incomplete
- Acontextual
- Inaccurate
- Scurrilous
- Spurious

Copyright
2005-06



38

Email – Long-Term Risk Exposures

Both Parties's IAPs:

- IP-address(es) used, disclosing location, trail
- authorised / unauthorised disclosure, with/without notification; data retention

Mail-Recipient's ISP:

- access to, use of traffic
- access to, use of content
- authorised / unauthorised disclosure, with/without notification
- retention after download

ISP Mail-Hosting / Webmail

- long-term retention

Gmail Subscribers

- targeted ads based on text from senders, cf. consumer behaviour manipulation
- correlation with data from other services

Senders to Gmail Addresses

- examination of text
- long-term retention
- long-term unauthorised disclosure, no notification

Copyright
2005-06



39

What's Google Really Up To? Summary

Google's Context

- The Web
- Search-Engines

Google's Business(es)

- Content-Discovery Services
- Content Services
- Data about Users

Perspectives on Google

- Business Model
- Copyright Law
- Consumer Law and Practice
- Privacy Law and Practice

Copyright
2005-06



40

What's Google Really Up to?



Roger Clarke

Xamax Consultancy Pty Ltd, Canberra

Visiting Professor in eCommerce at Uni. of Hong Kong
and in Cyberspace Law & Policy, U.N.S.W. & Comp. Sci., A.N.U.

[http://www.anu.edu.au/people/Roger.Clarke/...](http://www.anu.edu.au/people/Roger.Clarke/.../II/GoogleHK-0611)
.../II/GoogleHK-0611 {.html, .ppt}

eCom/iCom Uni. of Hong Kong – 2 November 2006

Copyright
2005-06

